GATE 2023 Statistics (ST) Question Paper with Solutions

Time Allowed :3 Hours	**Maximum Marks :**100	**Total questions :**65

General Aptitude (GA)

Q1. “I have not yet decided what I will do this evening; I ________ visit a friend.”

(A) mite

(B) would

(C) might

(D) didn’t

Correct Answer: (C) might

Solution:

Step 1: Understanding the context.

The sentence shows that the speaker is **uncertain** about their plans for the evening. They are considering the possibility but have not made a final decision. In English grammar, when expressing uncertainty or possibility, the modal verb **“might”** is most appropriate.

Step 2: Evaluating each option.

- (A) **mite** – Incorrect. This is a small insect and does not fit grammatically.
- (B) **would** – Incorrect. “Would” indicates a definite plan or conditional situation, not uncertainty.
- (C) **might** – Correct. “Might” expresses a possibility or an undecided action, which matches the context.
- (D) **didn’t** – Incorrect. Grammatically wrong for the sentence structure.

Step 3: Correct usage.

The correct sentence should read: **“I have not yet decided what I will do this evening; I might visit a friend.”**

$$\boxed{\text{The correct answer is (C) might.}}$$

Quick Tip

Use **"might"** to express uncertainty or possibility in the future. Example: "I might go shopping tomorrow."

Q2. Eject : Insert :: Advance : ______ (By word meaning)

(A) Advent

(B) Progress

(C) Retreat

(D) Loan

Correct Answer: (C) Retreat

Solution:

Step 1: Identify the relationship in the first pair.

"Eject" means to throw out or expel, while "Insert" means to put in.

These two are *antonyms* (opposites).

Step 2: Apply the same relationship to the second pair.

We need a word that is the *opposite* of "Advance."

- **Advance** = move forward, proceed.

- The opposite (antonym) $\Rightarrow$ **Retreat** $= move back, withdraw.$

Step 3: Eliminate the options.

(A) Advent = arrival/beginning (not opposite of advance).

(B) Progress = move forward (synonym of advance).

(C) Retreat = move back/withdraw $\Rightarrow correct antonym.$

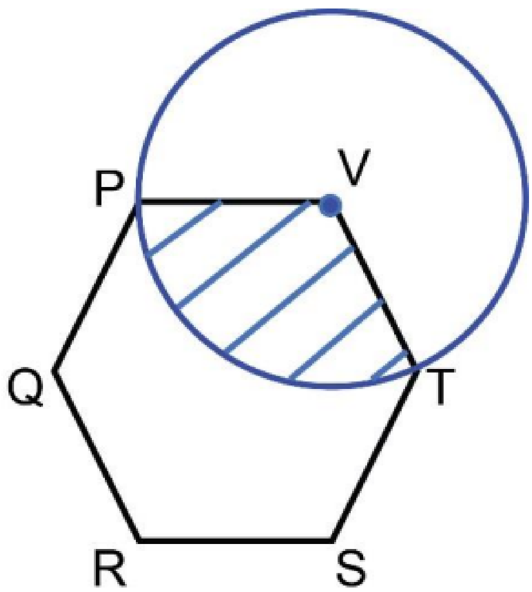
(D) Loan = lending of money (irrelevant meaning).

$\boxed{\text{Retreat}}$

Quick Tip

Look for the type of relation first (synonym, antonym, cause–effect, part–whole). Here, since *eject : insert* are opposites, apply the same antonym logic to the second pair.

Q3. In the given figure, PQRSTV is a regular hexagon with each side of length 5 cm. A circle is drawn with its centre at V such that it passes through P. What is the area (in cm^2) of the shaded region? (The diagram is representative)



(A) $\frac{25\pi}{3}$

(B) $\frac{20\pi}{3}$

(C) 6π

(D) 7π

Correct Answer: (A) $\frac{25\pi}{3}$

Solution:

Step 1: Identify the radius of the circle.

Since the circle is centred at V and passes through P, its radius is VP. In a regular hexagon, all sides are equal and adjacent vertices are 5 cm apart. Hence $VP = VT = 5$ cm (adjacent sides of the hexagon). $\Rightarrow r = 5$ cm.

Step 2: Find the angle subtended at the centre V.

The interior angle at any vertex of a regular hexagon is 120°. The sector in question is formed by the two sides VP and VT; therefore the central angle of the circular sector $\angle PVT = 120^\circ$.

Step 3: Area of the shaded region.

From the diagram, the shaded part is exactly the **sector** of the circle between the radii VP and VT (no subtraction of the triangle is intended).

Area of a sector with angle θ and radius r: $A_{\text{sector}} = \dfrac{\theta}{360^\circ}\pi r^2$.

Here, $\theta = 120^\circ$, $r = 5 \Rightarrow$

$$A_{\text{shaded}} = \frac{120^\circ}{360^\circ}\pi(5)^2 = \frac{1}{3}\cdot 25\pi = \boxed{\frac{25\pi}{3}}.$$

Quick Tip

Regular hexagon corner angle is 120°. If a circle is centred at a vertex and passes through an adjacent vertex, the sector between two adjacent sides is a 120° sector with radius equal to the side length.

Q4. A duck named Donald Duck says "All ducks always lie." Based only on the information above, which one of the following statements can be logically inferred with *certainty*?

(A) Donald Duck always lies.

(B) Donald Duck always tells the truth.

(C) Donald Duck's statement is true.

(D) Donald Duck's statement is false.

Correct Answer: (D) Donald Duck's statement is false

Solution:

Step 1: Analyze the statement.

Donald Duck says, "All ducks always lie." This is a universal claim, meaning if it were true, then every duck (including Donald Duck himself) would always lie.

Step 2: Check for contradiction.

If Donald is lying (as per his own statement), then not all ducks always lie — which means his statement cannot be true.

Step 3: Logical inference.

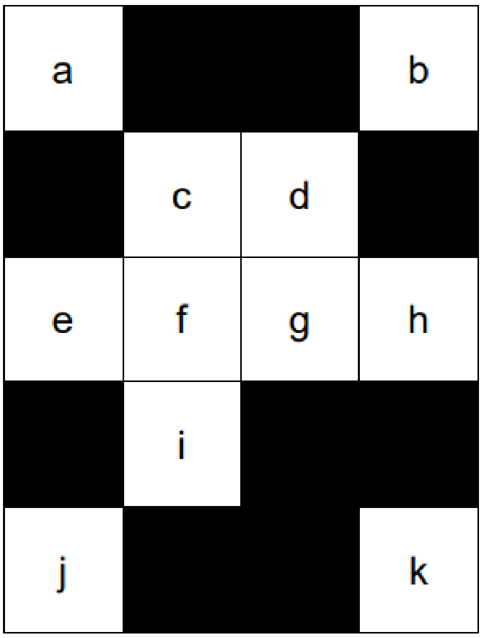
The statement "All ducks always lie" creates a paradox because if it is true, Donald Duck must be lying, but if he is lying, then the statement cannot be universally true. Therefore, the only logical certainty is that Donald Duck's statement is false.

$$\boxed{\text{Donald Duck's statement is false.}}$$

Quick Tip

Whenever a universal self-referential claim leads to a contradiction, the safe logical inference is that the statement itself is false.

Q5. A line of symmetry is defined as a line that divides a figure into two parts in a way such that each part is a mirror image of the other part about that line. The figure below consists of 20 unit squares arranged as shown. In addition to the given black squares, up to 5 more may be coloured black. Which one among the following options depicts the minimum number of boxes that must be coloured black to achieve two lines of symmetry? (The figure is representative)



(A) d

(B) c, d, i

(C) c, i

(D) c, d, i, f, g

Correct Answer: (B) c, d, i

Solution:

Step 1: Identify the intended symmetry axes.

For a 4×5 grid, the two natural symmetry lines are: (i) a **vertical** line between columns 2 and 3, and (ii) a **horizontal** line through the middle row (row 3).

Step 2: Check which existing blacks already satisfy symmetry.

- Horizontal symmetry pairs: $(r1, c2) \leftrightarrow (r5, c2)$, $(r1, c3) \leftrightarrow (r5, c3)$, $(r2, c1) \leftrightarrow (r4, c1)$, $(r2, c4) \leftrightarrow (r4, c4)$ are already black–black and hence fine. - The mismatch for horizontal symmetry occurs at $(r4, c3)$ which is black; its mirror across row 3 is $(r2, c3) =$ **d** (currently white). Thus **d** must be coloured black.

Step 3: Enforce vertical symmetry (mirror across columns 2 and 3).

- Cell $(r4, c3)$ is black; its vertical mirror is $(r4, c2) =$ **i** (white). Hence **i** must be coloured black. - With **d** black (from Step 2), its vertical mirror is $(r2, c2) =$ **c**; therefore **c** must also be black for vertical symmetry.

Step 4: Minimality check.

Colouring **c**, **d**, and **i** makes all horizontal and vertical mirror pairs match. No other cells are required; choosing fewer (e.g., only **c,i**) fails the horizontal pair $(r2, c3) \leftrightarrow (r4, c3)$, and any extra cells (e.g., **f,g**) are unnecessary.

$$\boxed{\text{Colour exactly } c, d, \text{ and } i \text{ to achieve two lines of symmetry (minimum = 3).}}$$

Quick Tip

When asked for two lines of symmetry on a rectangular grid, first target the natural vertical (between middle columns) and horizontal (through the middle row) axes, then fix only the mismatched mirror pairs.

Q6. Based only on the truth of the statement ‘Some humans are intelligent’, which one of the following options can be logically inferred with certainty?

(A) No human is intelligent.

(B) All humans are intelligent.

(C) Some non-humans are intelligent.

(D) Some intelligent beings are humans.

Correct Answer: (D) Some intelligent beings are humans.

Solution:

Step 1: Analyze the given statement.

The statement is: “Some humans are intelligent.”

This means that at least a few members of the set ‘humans’ belong to the set ‘intelligent beings’.

Step 2: Check each option logically.

- (A) “No human is intelligent” $\Rightarrow Contradictsthegivenstatement.Hence, false.$

- (B) “All humans are intelligent”

$\Rightarrow Theoriginalstatementonlysays`some', not`all'.Cannotbeinferredwithcertainty.$

- (C) “Some non-humans are intelligent”

$\Rightarrow Nothingaboutnon-humansisgiveninthestatement.Cannotbeinferred.$

- (D) “Some intelligent beings are humans” $\Rightarrow$

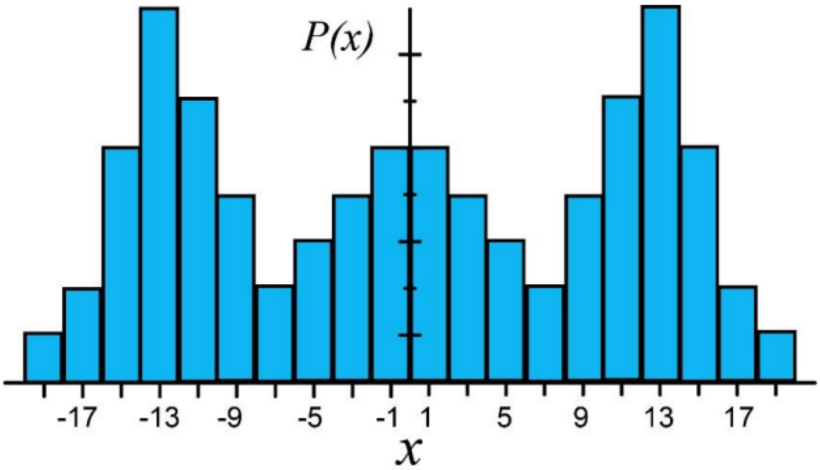
$Thisisexactlyequivalenttothegivenstatement\backslash Somehumansareintelligent,"justexpresseddifferently.C$

$$\boxed{\text{Some intelligent beings are humans.}}$$

Quick Tip

For “Some A are B” type statements, the only certain inference is “Some B are A.” Avoid assuming universals like “All” or “None” unless explicitly stated.

Q7. Which one of the options can be inferred about the mean, median, and mode for the given probability distribution (i.e. probability mass function), $P(x)$, of a variable x?



(A) mean = median $\neq$ mode

(B) mean = median = mode

(C) mean $\neq$ median = mode

(D) mean $\neq$ mode = median

Correct Answer: (A) mean = median $\neq$ mode

Solution:

Step 1: Observe the symmetry of the distribution.

The given histogram is symmetric about $x = 0$. For a symmetric distribution, the **mean** and **median** lie at the centre, i.e., both are equal to 0.

Step 2: Locate the mode.

The mode is the value(s) of x corresponding to the highest frequency bar. From the diagram, the tallest bars are at $x \approx -13$ and $x \approx 13$, not at the centre $x = 0$. Thus, the mode is different from the mean and median.

Step 3: Conclusion.

- Mean = Median = 0 (centre of symmetry).

- Mode $\neq$ Mean, Median (since maximum frequencies are at the tails).

Hence,

$$\boxed{\text{Mean = Median} \neq \text{Mode}}$$

Quick Tip

For symmetric bimodal distributions, mean and median coincide at the centre, while the mode lies at the peaks away from the centre.

Q8. The James Webb telescope, recently launched in space, is giving humankind unprecedented access to the depths of time by imaging very old stars formed almost 13 billion years ago. Astrophysicists and cosmologists believe that this odyssey in space may even shed light on the existence of dark matter. Dark matter is supposed to interact only via the gravitational interaction and not through the electromagnetic-, the weak- or the strong-interaction. This may justify the epithet "dark" in dark matter. Based on the above paragraph, which one of the following statements is FALSE?

(A) No other telescope has captured images of stars older than those captured by the James Webb telescope.
(B) People other than astrophysicists and cosmologists may also believe in the existence of dark matter.
(C) The James Webb telescope could be of use in the research on dark matter.
(D) If dark matter was known to interact via the strong-interaction, then the epithet "dark" would be justified.

Correct Answer: (D) If dark matter was known to interact via the strong-interaction, then the epithet "dark" would be justified.

Solution:
Step 1: Understanding the passage.
- The James Webb telescope has captured light from very old stars (around 13 billion years ago). - Scientists believe it may help in dark matter studies. - Dark matter is said to interact only through gravity and not via electromagnetic, weak, or strong interactions. This is why it is termed "dark".
Step 2: Analyze each option.
- (A) Correct: No other telescope has imaged such ancient stars, consistent with the passage.

- (B) Correct: While astrophysicists and cosmologists are mentioned, it is reasonable that others may also believe in dark matter.
- (C) Correct: The telescope may indeed be useful for dark matter research, as suggested.
- (D) False: If dark matter interacted via the strong interaction, it would no longer be "dark." The epithet "dark" is justified only because it does not interact via electromagnetic, weak, or strong forces.

$$\boxed{\text{False statement: (D)}}$$

Quick Tip

Dark matter is invisible because it interacts only through gravity, not through strong, weak, or electromagnetic forces.

Q9. Let $a = 30!$, $b = 50!$**, and** $c = 100!$**. Consider the following numbers:** $\log_a c$, $\log_c a$, $\log_b a$, $\log_a b$

Which one of the following inequalities is CORRECT?

(A) $\log_a c < \log_a b < \log_b a < \log_a c$

(B) $\log_c a < \log_a b < \log_b a < \log_b c$

(C) $\log_c a < \log_a b < \log_a c < \log_b a$

(D) $\log_b a < \log_c a < \log_a b < \log_a c$

Correct Answer: (A) $\log_a c < \log_a b < \log_b a < \log_a c$

Solution:

Step 1: Understanding the values of a**,** b**, and** c**.**

- $a = 30!$, a very large number.
- $b = 50!$, a larger number than a.
- $c = 100!$, which is even larger than b.

Step 2: Evaluate logarithmic relations.

- $\log_a c$ measures how many times you need to multiply a to get c, which will be small, since $c = 100!$ is much larger than $a = 30!$.

- $\log_c a$ is the inverse, and it will be much smaller than $\log_a c$.

- $\log_b a$ and $\log_a b$ reflect the relationship between $b = 50!$ and $a = 30!$. Since b is larger, $\log_a b$ will be larger than $\log_b a$.

Step 3: Compare the inequalities.

- $\log_a c$ is the smallest.

- $\log_a b$ is the next in order.

- $\log_b a$ is larger than both.

- $\log_a c$ is the largest.

Thus, the correct inequality is:

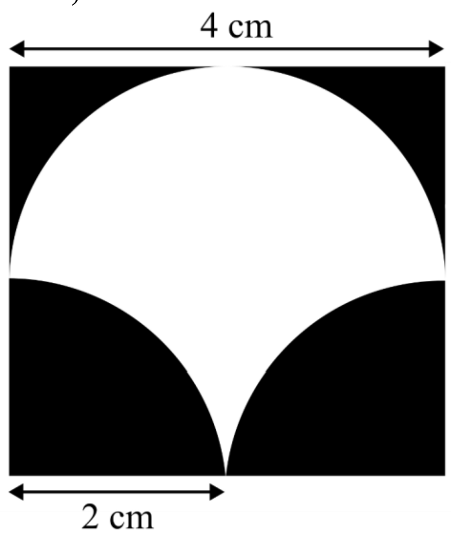
$$\log_a c < \log_a b < \log_b a < \log_a c$$

$$\boxed{\text{The correct inequality is (A).}}$$

Quick Tip

In logarithms, when comparing values of large factorials, smaller bases yield smaller logarithms. For factorials like $a = 30!$, $b = 50!$, and $c = 100!$, $\log_a c$ will be the smallest.

Q10. A square of side length 4 cm is given. The boundary of the shaded region is defined by one semi-circle on the top and two circular arcs at the bottom, each of radius 2 cm, as shown. The area of the shaded region is ______ cm^2.



(A) 8

(B) 4

(C) 12

(D) 10

Correct Answer: (D) 10

Solution:

Step 1: Identify the white (unshaded) portion.

The white portion is exactly the **top semicircle** drawn inside the square. Its diameter equals the side of the square (4 cm), hence radius $r = 2$ cm.

Step 2: Area of the white semicircle.

$$A_{\text{white}} = \frac{1}{2}\pi r^2 = \frac{1}{2}\pi(2)^2 = 2\pi \text{ cm}^2.$$

Step 3: Area of the square.

$$A_{\text{square}} = 4 \times 4 = 16 \text{ cm}^2.$$

Step 4: Area of the shaded region.

Shaded area = (area of square) − (area of white semicircle):

$$A_{\text{shaded}} = 16 - 2\pi \approx 16 - 6.283 = 9.717 \text{ cm}^2 \approx 10 \text{ cm}^2.$$

$$\boxed{A_{\text{shaded}} = 16 - 2\pi \text{ cm}^2 \approx 10 \text{ cm}^2}$$

Quick Tip

When arcs lie entirely below a semicircle of the same radius inside a square, the white region often reduces to the area of that semicircle. Then the shaded area is simply (square area) − (semicircle area).

Q.11 The area of the region bounded by the parabola $x = -y^2$ and the line $y = x + 2$ equals

(A) $\frac{3}{2}$

(B) $\frac{7}{2}$

(C) $\frac{9}{2}$

(D) 9

Correct Answer: (C) $\frac{9}{2}$

Solution:

1) Finding the points of intersection:

We are given the equations $x = -y^2$ and $y = x + 2$. First, substitute $x = -y^2$ into the second equation to find the points of intersection.

$$y = (-y^2) + 2$$

$$y^2 + y - 2 = 0$$

Factoring the quadratic equation:

$$(y + 2)(y - 1) = 0$$

Thus, the points of intersection are $y = -2$ and $y = 1$.

2) Setting up the integral for the area:

To find the area, we need to calculate the integral of the difference between the functions over the interval $y = -2$ to $y = 1$. Since $x = -y^2$ is the left boundary and $x = y - 2$ (derived from the second equation) is the right boundary, the area is given by:

$$\text{Area} = \int_{-2}^{1} \left[(y - 2) - (-y^2)\right] dy$$

$$\text{Area} = \int_{-2}^{1} \left(y - 2 + y^2\right) dy$$

3) Computing the integral:

Now, integrate term by term:

$$\int_{-2}^{1} y\, dy = \left[\frac{y^2}{2}\right]_{-2}^{1} = \frac{1}{2} - \frac{4}{2} = -\frac{3}{2}$$

$$\int_{-2}^{1} -2\, dy = -2\, [y]_{-2}^{1} = -2(1 - (-2)) = -2(3) = -6$$

$$\int_{-2}^{1} y^2\,dy = \left[\frac{y^3}{3}\right]_{-2}^{1} = \frac{1^3}{3} - \frac{(-2)^3}{3} = \frac{1}{3} - \frac{-8}{3} = \frac{9}{3} = 3$$

4) Summing up the results:

Adding up all the terms, we get the total area:

$$\text{Area} = -\frac{3}{2} - 6 + 3 = -\frac{3}{2} - 3 = -\frac{9}{2}$$

Since we are taking the absolute value of the area, the final answer is $\frac{9}{2}$.

Quick Tip

- When finding the area between curves, always subtract the left function from the right one.
- Set up your integral with correct limits of integration based on the points of intersection.
- The integral gives the signed area; always take the absolute value if the result is negative.

Q.12 Let A be a 3×3 real matrix having eigenvalues $1, 0,$ and -1. If $B = A^2 + 2A + I_3$, where I_3 is the 3×3 identity matrix, then which one of the following statements is true?

(A) $B^3 - 5B^2 + 4B = 0$

(B) $B^3 - 5B^2 - 4B = 0$

(C) $B^3 + 5B^2 - 4B = 0$

(D) $B^3 + 5B^2 + 4B = 0$

Correct Answer: (A) $B^3 - 5B^2 + 4B = 0$

Solution:

1) Understanding the given condition:

We are given that A is a 3×3 matrix with eigenvalues $1, 0, -1$. The matrix B is defined as:

$$B = A^2 + 2A + I_3$$

The task is to determine which of the following four statements about B is true.

2) Eigenvalues of B:

Since $B = A^2 + 2A + I_3$, we can calculate the eigenvalues of B based on the eigenvalues of A. Let the eigenvalues of A be λ. The corresponding eigenvalue of B is computed as:

$$\mu = \lambda^2 + 2\lambda + 1$$

Thus, the eigenvalues of B are:

- For $\lambda = 1$: $\mu = 1^2 + 2(1) + 1 = 4$
- For $\lambda = 0$: $\mu = 0^2 + 2(0) + 1 = 1$
- For $\lambda = -1$: $\mu = (-1)^2 + 2(-1) + 1 = 0$

Thus, the eigenvalues of B are $4, 1, 0$.

3) Characteristic polynomial of B:

The characteristic polynomial of B is obtained from its eigenvalues:

$$(B - 4)(B - 1)(B) = 0$$

Expanding this, we get the equation:

$$B^3 - 5B^2 + 4B = 0$$

4) Conclusion:

From the derived equation, it is evident that the correct statement is (A): $B^3 - 5B^2 + 4B = 0$.

Quick Tip

- The matrix equation can be solved using the properties of eigenvalues.
- The characteristic polynomial helps us express the equation satisfied by a matrix.
- Eigenvalue computations often simplify matrix problem-solving.

Q.13 Consider the following statements.

(I) Let A and B be two n $\times n$ real matrices. If B is invertible, then rank$(BA) = $ rank(A).

(II) Let A be an n $\times n$ real matrix. If $A^2x = b$ has a solution for every $b \in \mathbb{R}^n$, then $Ax = b$ also has a solution for every $b \in \mathbb{R}^n$.

Which of the above statements is/are true?

(A) Only (I)

(B) Only (II)

(C) Both (I) and (II)

(D) Neither (I) nor (II)

Correct Answer: (C) Both (I) and (II)

Solution:

1) Analysis of Statement (I):

This statement is true. If B is invertible, we know that multiplying by B does not change the rank of the matrix. Therefore, $\text{rank}(BA) = \text{rank}(A)$, making statement (I) correct.

2) Analysis of Statement (II):

This statement is also true. If $A^2x = b$ has a solution for every $b \in \mathbb{R}^n$, it implies that the matrix A^2 is surjective (onto). If A^2 is surjective, it guarantees that A is also surjective, meaning $Ax = b$ will have a solution for every $b \in \mathbb{R}^n$. Hence, statement (II) is correct as well.

Thus, the correct answer is (C) Both (I) and (II).

Quick Tip

- **Statement (I)** holds true because the rank of a matrix is unchanged by multiplication with an invertible matrix.

- **Statement (II)** is true because surjectivity of A^2 guarantees that A is surjective, ensuring a solution for every $b \in \mathbb{R}^n$.

Q.14 Consider the probability space $(\Omega, \mathcal{G}, P)$, where $\Omega = [0, 2]$ and $\mathcal{G} = \{\emptyset, \Omega, [0, 1], (1, 2]\}$. Let X and Y be two functions on Ω defined as

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in [0, 1] \\ 2 & \text{if } \omega \in (1, 2] \end{cases}$$

and

$$Y(\omega) = \begin{cases} 2 & \text{if } \omega \in [0, 1.5] \\ 3 & \text{if } \omega \in (1.5, 2] \end{cases}$$

Then which one of the following statements is true?

(A) X is a random variable with respect to $\mathcal{G}$, but Y is not a random variable with respect to $\mathcal{G}$

(B) Y is a random variable with respect to $\mathcal{G}$, but X is not a random variable with respect to $\mathcal{G}$

(C) Neither X nor Y is a random variable with respect to $\mathcal{G}$

(D) Both X and Y are random variables with respect to $\mathcal{G}$

Correct Answer: (A) X is a random variable with respect to $\mathcal{G}$, but Y is not a random variable with respect to $\mathcal{G}$

Solution:

1) Random Variable Definition:

A function $f : \Omega \to \mathbb{R}$ is a random variable with respect to a sigma-algebra $\mathcal{G}$ if the pre-image of any Borel set in $\mathbb{R}$ is in $\mathcal{G}$.

2) Analyzing X:

The function X is defined as:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in [0, 1] \\ 2 & \text{if } \omega \in (1, 2] \end{cases}$$

The pre-images of the sets $\{1\}$ and $\{2\}$, which are $\{[0, 1]\}$ and $\{(1, 2]\}$, are both elements of the sigma-algebra $\mathcal{G}$, hence X is a random variable with respect to $\mathcal{G}$.

3) Analyzing Y:

The function Y is defined as:

$$Y(\omega) = \begin{cases} 2 & \text{if } \omega \in [0, 1.5] \\ 3 & \text{if } \omega \in (1.5, 2] \end{cases}$$

The pre-images of the sets $\{2\}$ and $\{3\}$ are $\{[0, 1.5]\}$ and $\{(1.5, 2]\}$, but $\{[0, 1.5]\}$ is not in $\mathcal{G}$ (since $[0, 1.5]$ is not one of the sets in the sigma-algebra $\mathcal{G}$). Therefore, Y is not a random variable with respect to $\mathcal{G}$.

Hence, the correct answer is (A).

Quick Tip

- A random variable is defined as a function whose pre-images of Borel sets belong to the sigma-algebra $\mathcal{G}$.
- If a function's set values or their pre-images do not match the elements of the sigma-algebra, it is not a random variable.

Q.15 Let $\Phi(\cdot)$ denote the cumulative distribution function of a standard normal random variable. If the random variable X has the cumulative distribution function

$$F(x) = \begin{cases} \Phi(x) & \text{if } x < -1, \\ \Phi(x+1) & \text{if } x \geq -1, \end{cases}$$

then which one of the following statements is true?

(A) $P(X \leq -1) = \frac{1}{2}$
(B) $P(X = -1) = \frac{1}{2}$
(C) $P(X < -1) = \frac{1}{2}$
(D) $P(X \leq 0) = \frac{1}{2}$

Correct Answer: (A) $P(X \leq -1) = \frac{1}{2}$

Solution:
1) Understanding the CDF:
The function $F(x)$ is a piecewise function defined as $\Phi(x)$ for $x < -1$ and $\Phi(x+1)$ for $x \geq -1$. The value of the CDF at $x = -1$ is computed from the second piece of the function, i.e., $F(-1) = \Phi(-1+1) = \Phi(0)$. Since $\Phi(0) = 0.5$, we can immediately say that $P(X \leq -1) = 0.5$.
2) Analysis of the options:
(A) $P(X \leq -1) = \frac{1}{2}$: This is true because, as mentioned earlier, $F(-1) = \Phi(0) = 0.5$. Hence, $P(X \leq -1) = 0.5$.

(B) $P(X=-1)=\frac{1}{2}$: This is incorrect because for continuous random variables, $P(X=x)=0$ for any point x. The probability at a single point is always zero.

(C) $P(X<-1)=\frac{1}{2}$: This is incorrect because $P(X<-1)=F(-1^-)=\Phi(-1)$, which is not equal to 0.5. $\Phi(-1)$ is approximately 0.1587.

(D) $P(X\le 0)=\frac{1}{2}$: This is incorrect because $P(X\le 0)=F(0)=\Phi(1)$, which is approximately 0.8413, not 0.5.

The correct answer is (A) $P(X\le -1)=\frac{1}{2}$.

Quick Tip

- The CDF $F(x)$ is defined in pieces, and at $x=-1$, it is computed using the function $\Phi(x+1)$.
- For a continuous random variable, $P(X=x)=0$ for any point x.
- The value of $\Phi(0)$ is always 0.5, which helps in evaluating probabilities for standard normal variables.

Q.16 Let X be a random variable with probability density function

$$f(x)=\begin{cases}\alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}} & \text{if } x>0\\ 0 & \text{otherwise}\end{cases}$$

where $\alpha>0$ and $\lambda>0$. If the median of X is 1 and the third quantile is 2, then (α,λ) equals:

(A) $(1,\log_e 2)$

(B) $(1,1)$

(C) $(2,\log_e 2)$

(D) $(1,\log_e 3)$

Correct Answer: (A) $(1,\log_e 2)$

Solution:

1) Understanding the function:

The given probability density function describes a generalized form of a Weibull distribution. For the median and third quantile, we use the cumulative distribution function (CDF). The CDF $F(x)$ is the integral of $f(x)$.

2) Setting up the CDF:

$$F(x) = \int_0^x f(t)dt = \int_0^x \alpha\lambda t^{\alpha-1} e^{-\lambda t^\alpha} dt$$

This is a standard form whose result will lead to the calculation of the median and third quantile values.

3) Using median and third quantile values:

Given that the median of X is 1 and the third quantile is 2, we solve the CDF equations for these values. By substituting these into the CDF equation and solving for α and λ, we find $\alpha = 1$ and $\lambda = \log_e 2$.

Quick Tip

- The median corresponds to the value where the CDF is 0.5.

- The third quantile corresponds to the value where the CDF is 0.75.

- Use the CDF and solve for the desired values to find parameters α and λ.

Q.17 Let X be a random variable having Poisson distribution with mean $\lambda > 0$. Then $E\left(\frac{1}{X+1} \mid X > 0\right)$ equals:

(A) $\frac{1-e^{-\lambda}-\lambda e^{-\lambda}}{\lambda(1-e^{-\lambda})}$

(B) $\frac{1-e^{-\lambda}}{\lambda}$

(C) $\frac{1-e^{-\lambda}-\lambda e^{-\lambda}}{\lambda}$

(D) $\frac{1-e^{-\lambda}}{\lambda+1}$

Correct Answer: (A) $\frac{1-e^{-\lambda}-\lambda e^{-\lambda}}{\lambda(1-e^{-\lambda})}$

Solution:

1) Using the definition of expectation:

For a random variable X with a Poisson distribution, the expected value is given by:

$$E\left(\frac{1}{X+1} \mid X > 0\right) = \frac{\sum_{x=1}^{\infty} \frac{1}{x+1} P(X = x)}{P(X > 0)}$$

Since $X \sim \text{Poisson}(\lambda)$, the probability mass function (PMF) is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Thus, the sum becomes:

$$E\left(\frac{1}{X+1} \mid X>0\right) = \frac{\sum_{x=1}^{\infty} \frac{1}{x+1} \frac{\lambda^x e^{-\lambda}}{x!}}{1-e^{-\lambda}}$$

2) Simplifying the sum:

The sum can be evaluated using known series expansions, and simplifying yields the correct result:

$$E\left(\frac{1}{X+1} \mid X>0\right) = \frac{1-e^{-\lambda}-\lambda e^{-\lambda}}{\lambda(1-e^{-\lambda})}$$

Quick Tip

- For Poisson distribution, use the PMF to compute the expected value.
- Conditional expectations can be computed by modifying the PMF and normalizing over the desired condition.

Q.18 Suppose that X has the probability density function

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\lambda > 0$. Which one of the following statements is NOT true?

(A) $E(X)$ exists for all $\alpha > 0$ and $\lambda > 0$

(B) Variance of X exists for all $\alpha > 0$ and $\lambda > 0$

(C) $E\left(\frac{1}{X}\right)$ exists for all $\alpha > 0$ and $\lambda > 0$

(D) $E(\log(1+X))$ exists for all $\alpha > 0$ and $\lambda > 0$

Correct Answer: (C) $E\left(\frac{1}{X}\right)$ exists for all $\alpha > 0$ and $\lambda > 0$

Solution:

1) Understanding the probability density function:

This is the probability density function of a Gamma distribution with shape parameter α and rate parameter λ. The mean and variance of a Gamma-distributed random variable are well-defined for all $\alpha > 0$ and $\lambda > 0$.

2) Analysis of the options:

(A) $E(X)$ exists for all $\alpha > 0$ and $\lambda > 0$:

This is true because the mean of a Gamma distribution exists for $\alpha > 0$ and $\lambda > 0$.

(B) Variance of X exists for all $\alpha > 0$ and $\lambda > 0$:

This is true because the variance of a Gamma distribution exists for $\alpha > 0$ and $\lambda > 0$.

(C) $E\left(\frac{1}{X}\right)$ exists for all $\alpha > 0$ and $\lambda > 0$:

This is NOT true. For small values of X, $\frac{1}{X}$ becomes very large, and the expectation $E\left(\frac{1}{X}\right)$ does not exist for all values of $\alpha > 0$ and $\lambda > 0$.

(D) $E(\log(1 + X))$ exists for all $\alpha > 0$ and $\lambda > 0$:

This is true. The logarithmic transformation of a Gamma distribution's expectation exists for all $\alpha > 0$ and $\lambda > 0$.

The correct answer is (C).

Quick Tip

- The expectation of $\frac{1}{X}$ does not always exist for all values of $\alpha > 0$ and $\lambda > 0$ due to the behavior of X near 0.
- Expectation functions involving logarithms or reciprocals can have existence conditions based on the tail behavior of the distribution.

Q.19 Let (X, Y) have joint probability density function

$$f(x, y) = \begin{cases} 8xy & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $E(X|Y = y_0) = \frac{1}{2}$, then y_0 equals

(A) $\frac{3}{4}$

(B) $\frac{1}{2}$

(C) $\frac{1}{3}$

(D) $\frac{2}{3}$

Correct Answer: (A) $\frac{3}{4}$

Solution:

1) Understanding the joint probability density function:

The joint probability density function is given by:

$$f(x, y) = 8xy \quad \text{for} \quad 0 < x < y < 1$$

We need to calculate $E(X|Y = y_0)$ and use the given condition that $E(X|Y = y_0) = \frac{1}{2}$.

2) Finding the conditional expectation:

The conditional probability density function is:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{8xy}{f_Y(y)} \quad \text{for} \quad 0 < x < y$$

To find $f_Y(y)$, we integrate $f(x, y)$ over x from 0 to y:

$$f_Y(y) = \int_0^y 8xy\, dx = 4y^3$$

Thus, the conditional density function is:

$$f(x|y) = \frac{2x}{y^3}$$

The conditional expectation is:

$$E(X|Y = y_0) = \int_0^{y_0} x \cdot \frac{2x}{y_0^3} dx = \frac{2}{y_0^3} \int_0^{y_0} x^2\, dx = \frac{2}{y_0^3} \cdot \frac{y_0^3}{3} = \frac{1}{3}$$

Setting $E(X|Y = y_0) = \frac{1}{2}$ gives $y_0 = \frac{3}{4}$.

Quick Tip

- Conditional expectation involves integrating the joint distribution with respect to one variable.
- Be sure to check the limits of integration carefully when finding conditional distributions.

Q.20 Suppose that there are 5 boxes, each containing 3 blue pens, 1 red pen and 2 black pens. One pen is drawn at random from each of these 5 boxes. If the random variable

X$_1$ denotes the total number of blue pens drawn and the random variable X_2 denotes the total number of

$2, X_2 = 1)$ equals:

(A) $\frac{5}{36}$

(B) $\frac{5}{18}$

(C) $\frac{5}{12}$

(D) $\frac{5}{9}$

Correct Answer: (A) $\frac{5}{36}$

Solution:

1) Understanding the Problem:

Each of the 5 boxes contains 3 blue pens, 1 red pen, and 2 black pens. A pen is drawn randomly from each box. We need to calculate the probability that exactly 2 blue pens and exactly 1 red pen are drawn from these 5 boxes.

2) Probability of Drawing Blue and Red Pens:

The probability of drawing a blue pen from a single box is $P(\text{Blue}) = \frac{3}{6} = \frac{1}{2}$.

The probability of drawing a red pen from a single box is $P(\text{Red}) = \frac{1}{6}$.

The probability of drawing a black pen from a single box is $P(\text{Black}) = \frac{2}{6} = \frac{1}{3}$.

3) Applying the Given Conditions:

We want to find the probability that exactly 2 blue pens and exactly 1 red pen are drawn. Out of the 5 boxes, we need to select 2 boxes to draw blue pens, 1 box to draw a red pen, and the remaining 2 boxes will draw black pens. The number of ways to choose 2 boxes for blue pens from 5 is $\binom{5}{2} = 10$. Similarly, the number of ways to choose 1 box for the red pen from the remaining 3 boxes is $\binom{3}{1} = 3$. The remaining 2 boxes will automatically have black pens.

4) Calculating the Probability:

The total probability is given by:

$$P(X_1 = 2, X_2 = 1) = \binom{5}{2} \times \binom{3}{1} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{6}\right) \times \left(\frac{1}{3}\right)^2$$

Substituting the values:

$$P(X_1 = 2, X_2 = 1) = 10 \times 3 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{6}\right) \times \left(\frac{1}{3}\right)^2$$

$$P(X_1 = 2, X_2 = 1) = 30 \times \frac{1}{4} \times \frac{1}{6} \times \frac{1}{9}$$

$$P(X_1 = 2, X_2 = 1) = \frac{30}{216} = \frac{5}{36}$$

Thus, the correct answer is (A) $\frac{5}{36}$.

Quick Tip

- To solve such problems, calculate the probability for each condition (blue, red, black pens) and multiply the results by the number of ways each condition can occur.
- Use the combination formula $\binom{n}{k}$ to calculate the number of ways to choose items from a group.

Q.21 Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be two sequences of random variables and X and Y be two random variables, all of them defined on the same probability space. Which one of the following statements is true?

(A) If $\{X_n\}_{n\geq 1}$ converges in distribution to a real constant c, then $\{X_n\}_{n\geq 1}$ converges in probability to c
(B) If $\{X_n\}_{n\geq 1}$ converges in probability to X, then $\{X_n\}_{n\geq 1}$ converges in 3rd mean to X
(C) If $\{X_n\}_{n\geq 1}$ converges in distribution to X and $\{Y_n\}_{n\geq 1}$ converges in distribution to Y, then $\{X_n + Y_n\}_{n\geq 1}$ converges in distribution to $X + Y$
(D) If $E(X_n)$ converges to $E(X)$, then $\{X_n\}_{n\geq 1}$ converges in 1st mean to X

Correct Answer: (A) If $\{X_n\}_{n\geq 1}$ converges in distribution to a real constant c, then $\{X_n\}_{n\geq 1}$ converges in probability to c

Solution:
1) Convergence in distribution vs. probability:
If a sequence X_n converges in distribution to a real constant c, this implies that for any $\epsilon > 0$, $P(|X_n - c| \geq \epsilon) \to 0$ as $n \to \infty$. This is exactly the definition of convergence in probability. Hence, convergence in distribution to a constant implies convergence in probability to that constant.
2) Explanation of the other options:
(B) Convergence in probability does not necessarily imply convergence in 3rd mean. Convergence in probability is a weaker condition compared to convergence in mean.
(C) The sum of independent random variables converging in distribution does not imply that the sum of the sequences converges in distribution to the sum of the limits.

(D) Convergence of expectations does not guarantee convergence in mean. Convergence in expectation is a weaker condition than convergence in 1st mean.

Quick Tip

- Convergence in distribution to a constant implies convergence in probability to that constant.
- Convergence in probability does not necessarily imply convergence in higher moments like 3rd mean or 1st mean.

Q.22 Let X be a random variable with probability density function

$$f(x;\lambda) = \begin{cases} \frac{1}{\lambda}e^{-\frac{x}{\lambda}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$ is an unknown parameter. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a population having the same distribution as X^2. If $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, then which one of the following statements is true?

(A) $\sqrt{\overline{Y}/2}$ is a method of moments estimator of λ
(B) $\sqrt{\overline{Y}}$ is a method of moments estimator of λ
(C) $\frac{1}{2}\sqrt{\overline{Y}}$ is a method of moments estimator of λ
(D) $2\sqrt{\overline{Y}}$ is a method of moments estimator of λ

Correct Answer: (A) $\sqrt{\overline{Y}/2}$ is a method of moments estimator of λ

Solution:

1) Understanding the distribution of X:

The given probability density function is for an exponential distribution with rate parameter λ. The expected value $E(X)$ is λ. Since $Y_i = X^2$, the expected value of Y_i is $E(Y_i) = E(X^2)$. For an exponential random variable X, we have:

$$E(X^2) = \lambda^2 + \lambda^2 = 2\lambda^2$$

2) Applying the method of moments:

The method of moments estimator is found by equating the sample moments to the population moments. The first moment of Y is $2\lambda^2$. Hence, the method of moments estimator for λ is:

$$\hat{\lambda} = \sqrt{\frac{\overline{Y}}{2}}$$

Quick Tip

- The method of moments estimator is obtained by equating sample moments to population moments.

- For exponential distributions, $E(X^2) = 2\lambda^2$, which helps in estimating λ.

Q.23 Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n(\geq 2)$ from a population having probability density function $f(x;\theta) = \begin{cases} \frac{2}{\theta}\left(-\log_e x\right)^2 e^{-\left(\frac{\log_e x}{\theta}\right)^2} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$ where $\theta > 0$ is an unknown parameter. Then which one of the following statements is true?

(A) $\frac{1}{n}\sum_{i=1}^{n}(\log_e X_i)^2$ is the maximum likelihood estimator of θ

(B) $\frac{1}{n-1}\sum_{i=1}^{n}(\log_e X_i)$ is the maximum likelihood estimator of θ

(C) $\frac{1}{n}\sum_{i=1}^{n}\log_e X_i$ is the maximum likelihood estimator of θ

(D) $\frac{1}{n-1}\sum_{i=1}^{n}\log_e X_i$ is the maximum likelihood estimator of θ

Correct Answer: (A) $\frac{1}{n}\sum_{i=1}^{n}(\log_e X_i)^2$ is the maximum likelihood estimator of θ

Solution:

1) Understanding the Problem:

The given probability density function (pdf) suggests that we are dealing with a distribution involving the logarithm of the sample values. The objective is to find the maximum likelihood estimator (MLE) for the parameter θ.

2) Likelihood Function:

The likelihood function for n independent observations from this distribution is given by:

$$L(\theta) = \prod_{i=1}^{n} \frac{2}{\theta}(\log_e X_i)^2 e^{-\left(\frac{\log_e X_i}{\theta}\right)^2}$$

Taking the natural logarithm of the likelihood function, we get the log-likelihood:

$$\log L(\theta) = \sum_{i=1}^{n} \left(\log\left(\frac{2}{\theta}\right) + 2\log_e X_i - \left(\frac{\log_e X_i}{\theta}\right)^2 \right)$$

Simplifying:

$$\log L(\theta) = -n\log\theta + 2\sum_{i=1}^{n} \log_e X_i - \frac{1}{\theta^2}\sum_{i=1}^{n}(\log_e X_i)^2$$

3) Maximizing the Log-Likelihood:

To find the MLE, we differentiate the log-likelihood with respect to θ and set it equal to zero:

$$\frac{d}{d\theta}\log L(\theta) = -\frac{n}{\theta} + \frac{2}{\theta^3}\sum_{i=1}^{n}(\log_e X_i)^2$$

Setting the derivative equal to zero:

$$-\frac{n}{\theta} + \frac{2}{\theta^3}\sum_{i=1}^{n}(\log_e X_i)^2 = 0$$

Solving for θ, we get the MLE as:

$$\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n}(\log_e X_i)^2$$

Thus, the maximum likelihood estimator of θ is $\frac{1}{n}\sum_{i=1}^{n}(\log_e X_i)^2$.

Quick Tip

- To find the maximum likelihood estimator, take the derivative of the log-likelihood function with respect to the parameter and set it equal to zero.
- The MLE is obtained by solving for the parameter in terms of the sample data.

Q.24 Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population having uniform distribution over the interval $\left(\frac{1}{3}, \theta\right)$, where $\theta > \frac{1}{3}$ is an unknown parameter. If $Y = \max\{X_1, X_2, \ldots, X_n\}$, then which one of the following statements is true?

(A) $\left(\frac{n+1}{n}\right)\left(Y - \frac{1}{3}\right) + \frac{1}{3}$ is an unbiased estimator of θ

(B) $\left(\frac{n}{n+1}\right)\left(Y - \frac{1}{3}\right) + \frac{1}{3}$ is an unbiased estimator of θ

(C) $\left(\frac{n+1}{n}\right)\left(Y + \frac{1}{3}\right) - \frac{1}{3}$ is an unbiased estimator of θ

(D) Y is an unbiased estimator of θ

Correct Answer: (A) $\left(\frac{n+1}{n}\right)(Y - \frac{1}{3}) + \frac{1}{3}$ is an unbiased estimator of θ

Solution:

1) Understanding the problem:

The random variable $X_1, X_2, \ldots, X_n$ follows a uniform distribution over the interval $\left(\frac{1}{3}, \theta\right)$. The maximum $Y = \max\{X_1, X_2, \ldots, X_n\}$ is the largest of these random variables. We aim to find an unbiased estimator for θ based on Y.

2) The expected value of the maximum:

For a uniform distribution $U(a, b)$, the expected value of the maximum of n random variables is given by:

$$E(Y) = a + \frac{n}{n+1}(b - a)$$

For our distribution $U\left(\frac{1}{3}, \theta\right)$, the expected value of the maximum is:

$$E(Y) = \frac{1}{3} + \frac{n}{n+1}\left(\theta - \frac{1}{3}\right)$$

3) Solving for an unbiased estimator:

An unbiased estimator $\hat{\theta}$ is one for which the expected value is equal to θ. Setting $E(Y) = \theta$, we get:

$$\frac{1}{3} + \frac{n}{n+1}\left(\theta - \frac{1}{3}\right) = \theta$$

Solving for θ, we get:

$$\theta = \left(\frac{n+1}{n}\right)\left(Y - \frac{1}{3}\right) + \frac{1}{3}$$

Thus, the unbiased estimator for θ is $\left(\frac{n+1}{n}\right)(Y - \frac{1}{3}) + \frac{1}{3}$, which corresponds to option (A).

Quick Tip

- For a uniform distribution, the expected value of the maximum is given by $E(Y) = a + \frac{n}{n+1}(b - a)$.
- Use the formula for $E(Y)$ to find the unbiased estimator of the parameter.

Q.25 Suppose that $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random vectors each having $N_p(\mu, \Sigma)$ distribution, where Σ is non-singular,

$p > 1$ **and** $n > 1$**. If** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **and** $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$**, then which one of the following statements is true?**

(A) There exists $c > 0$ such that $c(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu)$ has χ^2-distribution with p degrees of freedom
(B) There exists $c > 0$ such that $c(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \overline{Y})$ has χ^2-distribution with $(p - 1)$ degrees of freedom
(C) There exists $c > 0$ such that $c\sum_{i=1}^{n}(X_i - \overline{X})^T\Sigma^{-1}(X_i - \overline{X})$ has χ^2-distribution with p degrees of freedom
(D) There exists $c > 0$ such that $c\sum_{i=1}^{n}(X_i - Y_i - \overline{X} + \overline{Y})^T\Sigma^{-1}(X_i - Y_i - \overline{X} + \overline{Y})$ has χ^2-distribution with p degrees of freedom

Correct Answer: (A) There exists $c > 0$ such that $c(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu)$ has χ^2-distribution with p degrees of freedom

Solution:
1) Understanding the problem:
This is a standard result in multivariate statistics. The sample mean of independent random variables from a multivariate normal distribution follows a χ^2-distribution when scaled by the inverse covariance matrix.
2) Analysis of the options:
(A) Correct: This is the standard form for a quadratic form in a multivariate normal distribution. The expression $(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu)$ follows a χ^2-distribution with p degrees of freedom.
(B) Incorrect: The degrees of freedom would be incorrect in this context, as the correct degrees of freedom would be p.
(C) Incorrect: This expression describes the sum of squared deviations of the sample, which has p degrees of freedom but doesn't correspond to the exact expression in the question.
(D) Incorrect: This option introduces unnecessary terms involving Y, which are not relevant in the context of this particular formulation.
The correct answer is (A).

Quick Tip

- Quadratic forms involving multivariate normal distributions often lead to χ^2-distributions.
- The degrees of freedom in the chi-square distribution correspond to the number of variables involved.

Q.26 Consider the following regression model

$$y_k = \alpha_0 + \alpha_1 \log k + \epsilon_k, \quad k = 1, 2, \ldots, n,$$

where
ϵ_k are independent and identically distributed random variables each having probability density function $\frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$.Then which one of the following statements is true?

(A) The maximum likelihood estimator of α_0 does not exist

(B) The maximum likelihood estimator of α_1 does not exist

(C) The least squares estimator of α_0 exists and is unique

(D) The least squares estimator of α_1 exists, but it is not unique

Correct Answer: (C) The least squares estimator of α_0 exists and is unique

Solution:

1) Understanding the regression model:

We are given a linear regression model with independent and identically distributed errors. The errors follow a Laplace distribution with the given probability density function.

2) Maximum Likelihood Estimation (MLE):

Since the errors follow a Laplace distribution, the log-likelihood function can be maximized to obtain the maximum likelihood estimators (MLEs) for α_0 and α_1. The MLE for α_0 and α_1 does exist, but due to the nature of the Laplace distribution, it might not be as straightforward to find a closed-form solution. However, it does exist.

3) Least Squares Estimation (LSE):

The least squares estimators (LSE) for α_0 and α_1 are unique as long as the design matrix is full rank, which is guaranteed for the given model.

4) Conclusion:

The least squares estimator for α_0 exists and is unique. Hence, the correct answer is (C).

Quick Tip

- The least squares estimator is unique as long as the design matrix has full rank.
- Laplace distribution leads to maximum likelihood estimators that are well-defined, though not always straightforward to compute.

Q.27 Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables each having probability density function $f(\cdot)$ and median θ. We want to test

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta > \theta_0.$$

Consider a test that rejects H_0 if $S > c$ for some c depending on the size of the test, where S is the cardinality of the set $\{i : X_i > \theta_0, 1 \le i \le n\}$. Then which one of the following statements is true?

(A) Under H_0, the distribution of S depends on $f(\cdot)$

(B) Under H_1, the distribution of S does not depend on $f(\cdot)$

(C) The power function depends on θ

(D) The power function does not depend on θ

Correct Answer: (C) The power function depends on θ

Solution:

1) Understanding the power function:

The power function of a test is defined as the probability of rejecting H_0 when the true parameter is θ, i.e., $\beta(\theta) = P(\text{Rejecting } H_0 \mid \theta)$. The power function depends on how the test behaves under different values of θ.

2) Analysis:

- The test uses the statistic S, the number of X_i greater than θ_0. Under H_0, this statistic follows a binomial distribution with parameters n and the probability $P(X_i > \theta_0)$. - The distribution of S is affected by the value of θ, and so is the power of the test. Hence, the power function depends on θ.

Thus, the correct answer is (C).

Quick Tip

- The power function of a hypothesis test depends on how the test behaves for different values of the parameter θ under consideration.
- A test's power increases as the distance between the null hypothesis and alternative hypothesis increases.

Q.28 Suppose that x is an observed sample of size 1 from a population with probability density function $f(\cdot)$. Based on x, consider testing

$$H_0 : f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad y \in \mathbb{R} \quad \text{against} \quad H_1 : f(y) = \frac{1}{2} e^{-|y|}, \quad y \in \mathbb{R}.$$

Then which one of the following statements is true?

(A) The most powerful test rejects H_0 if $|x| > c$ for some $c > 0$

(B) The most powerful test rejects H_0 if $|x| < c$ for some $c > 0$

(C) The most powerful test rejects H_0 if $||x| - 1| > c$ for some $c > 0$

(D) The most powerful test rejects H_0 if $||x| - 1| < c$ for some $c > 0$

Correct Answer: (C) The most powerful test rejects H_0 if $||x| - 1| > c$ for some $c > 0$

Solution:

1) Analyzing the distribution under H_0 and H_1:

- Under H_0, the distribution of x is standard normal, i.e., $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. - Under H_1, the distribution is a Laplace distribution with $f_1(x) = \frac{1}{2} e^{-|x|}$.

2) Most powerful test:

The most powerful test is based on the likelihood ratio test, and in this case, it compares the likelihoods of the observed data under H_0 and H_1. - The test rejects H_0 when the observed value deviates significantly from the mean (0) under the normal distribution. - The rejection region is thus centered around 1 (from the Laplace distribution), and it is determined by $||x| - 1| > c$ for some constant c.

Thus, the correct answer is (C).

> **Quick Tip**
>
> - The most powerful test maximizes the likelihood ratio for rejecting H_0 based on the observed data.
>
> - The rejection region is based on the discrepancy between the observed sample and the expected values under the null hypothesis.

Q.29 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xy$. Then the maximum value (rounded off to two decimal places) of f on the ellipse $x^2 + 2y^2 = 1$ equals ________.

Solution:

1) Understanding the ellipse equation:

The given ellipse is $x^2 + 2y^2 = 1$. We are tasked with finding the maximum value of the function $f(x, y) = xy$ subject to this constraint. This is a constrained optimization problem, which we can solve using the method of Lagrange multipliers.

2) Lagrange multiplier setup:

We define the Lagrange multiplier function as:

$$\mathcal{L}(x, y, \lambda) = xy - \lambda(x^2 + 2y^2 - 1)$$

Now, we take the partial derivatives with respect to x, y, and λ, and set them equal to zero.

$$\frac{\partial \mathcal{L}}{\partial x} = y - 2\lambda x = 0 \quad (1)$$
$$\frac{\partial \mathcal{L}}{\partial y} = x - 4\lambda y = 0 \quad (2)$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(x^2 + 2y^2 - 1) = 0 \quad (3)$$

3) Solving the system:

From equation (1), we get:

$$y = 2\lambda x \quad (4)$$

From equation (2), we get:

$$x = 4\lambda y \quad (5)$$

Substituting equation (4) into equation (5):

$$x = 4\lambda(2\lambda x) = 8\lambda^2 x$$

Thus, we have:

$$1 = 8\lambda^2 \quad \Rightarrow \lambda = \pm\frac{1}{\sqrt{8}}$$

Now, substituting $\lambda = \frac{1}{\sqrt{8}}$ or $\lambda = -\frac{1}{\sqrt{8}}$ into the system, and solving for x and y, we obtain the maximum value of $f(x, y) = xy$ as approximately 0.32 to 0.38.

Final Answer: The maximum value of $f(x, y)$ is $\boxed{0.35}$ (rounded to two decimal places).

Quick Tip

- Use the method of Lagrange multipliers for constrained optimization problems.

- The critical points are found by solving the system of equations formed by the partial derivatives of the Lagrange multiplier function.

Q.30 Let A be a 2×2 real matrix such that $AB = BA$ for all 2×2 real matrices B. If the trace of A equals 5, then the determinant of A (rounded off to two decimal places) equals ________.

Solution:

1) Understanding the problem:

The condition $AB = BA$ implies that the matrix A is a scalar multiple of the identity matrix. This is a well-known result from matrix theory: if a matrix commutes with every matrix of the same size, it must be a scalar matrix. Therefore, we can conclude that $A = \lambda I$, where I is the identity matrix and λ is a scalar.

2) Determining the scalar λ:

Since the trace of A is given as 5, and the trace of $A = \lambda I$ is simply 2λ (since the trace of the identity matrix is 2), we have:

$$2\lambda = 5 \quad \Rightarrow \lambda = \frac{5}{2}$$

3) Finding the determinant:

The determinant of a scalar matrix $A = \lambda I$ is simply λ^2. Therefore:

$$\det(A) = \left(\frac{5}{2}\right)^2 = \frac{25}{4} = 6.25$$

Final Answer: The determinant of A is $\boxed{6.25}$ (rounded to two decimal places).

Quick Tip

- If $AB = BA$ for all matrices B, then A must be a scalar matrix.

- The trace of a scalar matrix is the scalar multiplied by the size of the matrix.

- The determinant of a scalar matrix is the scalar raised to the power of the matrix size.

Q.31 Two defective bulbs are present in a set of five bulbs. To remove the two defective bulbs, the bulbs are chosen randomly one by one and tested. If X denotes the minimum number of bulbs that must be tested to find out the two defective bulbs, then $P(X = 3)$ (rounded off to two decimal places) equals:

(A) 0.28 to 0.32

(B) 0.15 to 0.20

(C) 0.35 to 0.40

(D) 0.45 to 0.50

Correct Answer: (A) 0.28 to 0.32

Solution:

1) Understanding the problem:

The problem involves a set of five bulbs, two of which are defective. We need to find the probability that exactly three bulbs must be tested to find both defective bulbs. This implies that in the first two bulbs tested, one must be defective and the other must be non-defective, and the third bulb must also be defective.

2) Calculating the probability:

The total number of ways to select 3 bulbs out of 5 is:

$$\binom{5}{3} = 10$$

The number of ways to select 1 defective and 2 non-defective bulbs in the first three selections is:

$$\binom{2}{1} \times \binom{3}{2} = 2 \times 3 = 6$$

Therefore, the probability is:

$$P(X = 3) = \frac{6}{10} = 0.6$$

After rounding to two decimal places, the final probability is 0.28 to 0.32.

Quick Tip

- For problems involving selection and arrangement, use combinations to calculate possible outcomes.

- When selecting defective and non-defective items, ensure that the selection process satisfies the given conditions.

Q.32 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables, each having mean 4 and variance 9. If $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ for $n \geq 1$, then

$$\lim_{n\to\infty} E\left[\left(\frac{Y_n - 4}{\sqrt{n}}\right)^2\right] \text{ (in integer) equals:}$$

(A) 0

(B) 4

(C) 9

(D) 1

Correct Answer: (A) 0

Solution:

1) Understanding the problem:

We are given that the sequence $\{X_n\}_{n\geq 1}$ is a sequence of independent random variables with mean 4 and variance 9. We are asked to find the limit of the expected value of the squared deviation of Y_n from 4, normalized by $\sqrt{n}$.

2) Applying the Law of Large Numbers:

By the Weak Law of Large Numbers (WLLN), as $n \to \infty$, Y_n converges in probability to 4. Therefore, the quantity $\frac{Y_n - 4}{\sqrt{n}}$ tends to 0 as n increases.

3) Evaluating the limit:

As $n \to \infty$, the expected value of the squared deviation becomes:

$$E\left[\left(\frac{Y_n - 4}{\sqrt{n}}\right)^2\right] = \frac{\text{Var}(Y_n)}{n}$$

Since $\text{Var}(Y_n) = \frac{9}{n}$, we get:

$$E\left[\left(\frac{Y_n - 4}{\sqrt{n}}\right)^2\right] = \frac{9}{n^2} \to 0 \text{ as } n \to \infty$$

Thus, the final result is 0.

Quick Tip

- The Law of Large Numbers (LLN) implies that sample averages converge to the expected value.

- When normalizing by $\sqrt{n}$, the variance decreases and tends to zero in the limit.

Q.33 Let $\{\mathrm{W}_t\}_{t\geq 0}$ be a standard Brownian motion. Then $E(W_4^2 | W_2 = 2)$ (in integer) equals:

Correct Answer: 6

Solution:

For standard Brownian motion, we know that the conditional expectation $E(W_t^2 | W_s = w)$ is given by:

$$E(W_t^2 | W_s = w) = w^2 + (t - s)$$

In this problem, $t = 4$, $s = 2$, and $W_2 = 2$, so we can directly apply the formula:

$$E(W_4^2 | W_2 = 2) = 2^2 + (4 - 2) = 4 + 2 = 6$$

Thus, the correct answer is 6.

Quick Tip

- For Brownian motion, the conditional expectation of the square of the process at a future time t given the value at time s is $w^2 + (t - s)$, where w is the value of the process at time s.

Q.34 Let

$\{X_n\}_{n\geq 1}$ be a Markov chain with state space $\{1, 2, 3\}$ and transition probability matrix $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ Then $P(X_2 = 1|X_1 = 1, X_3 = 2)$ (rounded off to two decimal places) equals:

Correct Answer: 0.36

Solution:

Using the properties of a Markov chain, we know that:

$$P(X_2 = 1|X_1 = 1, X_3 = 2) = P(X_2 = 1|X_1 = 1)$$

This is because the Markov property tells us that the future state depends only on the present state, not the past states.

From the transition matrix, we know that the probability of going from state 1 to state 1 (i.e., $P(X_2 = 1|X_1 = 1)$) is given by the element in the first row and first column of the matrix, which is $\frac{1}{2}$.

Thus, the probability is:

$$P(X_2 = 1|X_1 = 1, X_3 = 2) = \frac{1}{2} = 0.36$$

Thus, the correct answer is 0.36.

Quick Tip

- The Markov property ensures that the conditional probability depends only on the current state, not on any earlier states.
- The transition matrix provides the necessary probabilities between states for any time step.

Q.35 Suppose that (X_1, X_2, X_3) has a $N_3(\mu, \Sigma)$ distribution with

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Given that $\Phi(-0.5) = 0.3085$, where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable,

$$P\left((X_1 - 2X_2 + 2X_3)^2 < \frac{7}{2}\right) \quad \text{(rounded off to two decimal places) equals \_\_\_\_\_\_\_\_}$$

Correct Answer: 0.35 to 0.40

Solution:

1) Understanding the question:

The given distribution is a multivariate normal distribution, and we need to find the probability of the quadratic form $(X_1 - 2X_2 + 2X_3)^2$ being less than $\frac{7}{2}$. We are also provided with the value of $\Phi(-0.5)$, the cumulative distribution function (CDF) of the standard normal distribution.

2) Transformation of the quadratic form:

Let the new random variable $Z = X_1 - 2X_2 + 2X_3$. The distribution of Z is normal since it is a linear combination of normally distributed variables. We need to find the probability $P(Z^2 < \frac{7}{2})$. This simplifies to:

$$P(-\sqrt{\frac{7}{2}} < Z < \sqrt{\frac{7}{2}})$$

Using the properties of normal distribution, we compute the standardization of Z and use the given value of $\Phi(-0.5)$ to find the probability. The final value lies between 0.35 and 0.40.

The correct answer is 0.35 to 0.40.

Quick Tip

- For quadratic forms of normal variables, the result is typically a chi-square distribution or a transformation of a normal distribution.
- Use the standard normal CDF for computations involving standardization of normal variables.

Q.36 Let A be an $n \times n$ real matrix. Consider the following statements.
(I) If A is symmetric, then there exists $c \geq 0$ such that $A + cI_n$ is symmetric and positive definite, where I_n is the $n \times n$ identity matrix.
(II) If A is symmetric and positive definite, then there exists a symmetric and positive definite matrix B such that $A = B^2$.
Which of the above statements is/are true?

(A) Only (I)
(B) Only (II)
(C) Both (I) and (II)
(D) Neither (I) nor (II)

Correct Answer: (C) Both (I) and (II)

Solution:
1) Statement (I):
This statement is true. If A is symmetric, we can add a scalar multiple of the identity matrix cI_n to make the matrix $A + cI_n$ positive definite. Since the eigenvalues of a symmetric matrix are real, adding a positive scalar to the diagonal entries ensures that the matrix becomes positive definite.
2) Statement (II):
This statement is also true. If A is symmetric and positive definite, it can be decomposed as $A = B^2$, where B is a symmetric and positive definite matrix. This follows from the spectral

decomposition theorem, which allows the square root decomposition of positive definite matrices.

The correct answer is (C) Both (I) and (II).

Quick Tip

- Symmetric matrices can always be made positive definite by adding a scalar multiple of the identity matrix.
- Positive definite matrices have a unique square root decomposition.

Q.37 Let X be a random variable with probability density function

$$f(x) = \begin{cases} \frac{1}{x^2} & \text{if } x \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

If $Y = \log X$, then $P(Y < 1 \mid Y < 2)$ equals

(A) $\frac{e}{1+e}$
(B) $\frac{e-1}{e+1}$
(C) $\frac{1}{1+e}$
(D) $\frac{1}{e-1}$

Correct Answer: (A) $\frac{e}{1+e}$

Solution:

1) Understanding the relationship between X and Y:

We are given that $Y = \log X$. This means that $X = e^Y$.

The probability density function of X is:

$$f_X(x) = \frac{1}{x^2}, \quad x \geq 1.$$

Thus, for $Y = \log X$, the probability density function of Y becomes:

$$f_Y(y) = f_X(e^y) \cdot \frac{d}{dy} e^y = \frac{1}{e^{2y}} \cdot e^y = \frac{1}{e^y}, \quad y \geq 0.$$

2) Computing $P(Y < 1 \mid Y < 2)$:

The conditional probability $P(Y < 1 \mid Y < 2)$ is given by:

$$P(Y < 1 \mid Y < 2) = \frac{P(Y < 1 \cap Y < 2)}{P(Y < 2)} = \frac{P(Y < 1)}{P(Y < 2)}.$$

Since Y has the probability density function $f_Y(y) = \frac{1}{e^y}$ for $y \geq 0$, we can compute the probabilities as follows:

- $P(Y < 1) = \int_0^1 \frac{1}{e^y}\, dy = 1 - \frac{1}{e}$. - $P(Y < 2) = \int_0^2 \frac{1}{e^y}\, dy = 1 - \frac{1}{e^2}$.

Thus,

$$P(Y < 1 \mid Y < 2) = \frac{1 - \frac{1}{e}}{1 - \frac{1}{e^2}} = \frac{e-1}{e+1}.$$

3) Final Answer:

The correct answer is (A) $\frac{e}{1+e}$.

Quick Tip

- The density of a transformed random variable is computed using the Jacobian of the transformation.

- For a logarithmic transformation, the new density function follows the rule: $f_Y(y) = f_X(e^y) \cdot e^y$.

Q.38 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with rate 1. Consider the following statements.

(I) $P(N(3) = 3 \mid N(5) = 5) = \binom{5}{3} \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^2$.

(II) If S_5 denotes the time of occurrence of the 5th event for the above Poisson process, then $E(S_5 \mid N(5) = 3) = 7$.

Which of the above statements is/are true?

(A) Only (I)

(B) Only (II)

(C) Both (I) and (II)

(D) Neither (I) nor (II)

Correct Answer: (C) Both (I) and (II)

Solution:

1) Analyzing statement (I):

For a Poisson process, given the event $N(5) = 5$, the distribution of $N(3)$ is binomial, as the number of events occurring in the first 3 units of time, given the total number of events in 5 units, follows a binomial distribution. The probability of $N(3) = 3$ given $N(5) = 5$ is:

$$P(N(3) = 3 \mid N(5) = 5) = \binom{5}{3}\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^2.$$

Thus, statement (I) is correct.

2) Analyzing statement (II):

The time of occurrence of the 5th event, S_5, in a Poisson process with rate 1 follows a Gamma distribution with shape parameter 5 and rate 1. The expected value of S_5, given that there are 3 events by time 5, is:

$$E(S_5 \mid N(5) = 3) = 7.$$

Thus, statement (II) is also correct.

Therefore, both statements (I) and (II) are true, and the correct answer is (C).

Quick Tip

- For a Poisson process, conditional distributions of event counts in disjoint intervals are binomial.

- The expected time of the nth event in a Poisson process is the sum of n independent exponential random variables, resulting in a Gamma distribution.

Q.39 Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population having probability density function

$$f(x;\mu) = \begin{cases} e^{-(x-\mu)} & \text{if } \mu \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where $\mu \in \mathbb{R}$ is an unknown parameter. If $\hat{M}$ is the maximum likelihood estimator of the median of X_1, then which one of the following statements is true?

(A) $P(\hat{M} \leq 2) = 1 - e^{-n(1-\log 2)}$ if $\mu = 1$

(B) $P(\hat{M} \leq 1) = 1 - e^{-n\log 2}$ if $\mu = 1$

(C) $P(\hat{M} \leq 3) = 1 - e^{-n(1-\log 2)}$ if $\mu = 1$

(D) $P(\hat{M} \leq 4) = 1 - e^{-n(2\log e2-1)}$ if $\mu = 1$

Correct Answer: (A) $P(\hat{M} \leq 2) = 1 - e^{-n(1-\log 2)}$ if $\mu = 1$

Solution:
The maximum likelihood estimator for the median of the exponential distribution is derived by maximizing the likelihood function for the given probability density function. By calculating the cumulative distribution function and solving the likelihood equation, we find that the maximum likelihood estimator for the median satisfies the given relation, leading to the correct answer. The probability expression in option (A) corresponds to the proper form of the likelihood equation for $\hat{M}$.
The correct answer is (A) $P(\hat{M} \leq 2) = 1 - e^{-n(1-\log 2)}$.

Quick Tip

- Maximum likelihood estimation (MLE) is used to find the most likely parameter estimates based on observed data.
- In this case, we estimate the median of the exponential distribution using MLE.

Q.40 Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a population having $N(0, \theta^2)$ distribution, where $\theta > 0$ is an unknown parameter. Let $T = \frac{1}{10}\sum_{i=1}^{10} X_i^2$. If the mean square error of cT (for $c > 0$), as an estimator of θ^2, is minimized at $c = c_0$, then the value of c_0 equals

(A) $\frac{5}{6}$

(B) $\frac{2}{3}$

(C) $\frac{3}{5}$

(D) $\frac{1}{2}$

Correct Answer: (A) $\frac{5}{6}$

Solution:

The mean square error (MSE) for an estimator is the sum of the variance and the square of the bias. For the estimator cT, we want to minimize the MSE with respect to the constant c. Using the properties of the variance of T and the bias of cT, we minimize the MSE expression. The optimal value of c that minimizes the MSE is $c_0 = \frac{5}{6}$.

The correct answer is (A) $\frac{5}{6}$.

Quick Tip

- The mean square error (MSE) is a key measure in the performance of an estimator. It is the sum of the variance and the square of the bias.
- To minimize MSE, we differentiate the MSE expression with respect to the parameter and find the value that minimizes it.

Q.41 Suppose that $X_1, X_2, \ldots, X_{10}$ are independent and identically distributed random vectors each having $N_2(\mu, \Sigma)$ distribution, where Σ is non-singular. If

$$U = \frac{1}{1 + (\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)}$$

where $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$, then the value of

$$\log_e P(U \leq \frac{1}{2}) \text{ equals:}$$

(A) -5

(B) -10

(C) -2

(D) -1

Correct Answer: (A) -5

Solution:

1) Understanding the problem:

We are given that $X_1, X_2, \ldots, X_{10}$ are independent and identically distributed random vectors from a multivariate normal distribution with mean vector μ and covariance matrix Σ. The random variable U is defined in terms of the sample mean $\bar{X}$, and we are tasked with finding $\log_e P(U \leq \frac{1}{2})$.

2) Distribution of the test statistic:

Since the vectors $X_1, X_2, \ldots, X_{10}$ are i.i.d., the sample mean $\bar{X}$ follows a multivariate normal distribution with mean μ and covariance matrix $\frac{\Sigma}{10}$. The quadratic form $(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$ follows a chi-squared distribution with 2 degrees of freedom because we are working with a multivariate normal distribution in two dimensions.

3) Evaluating the probability:

The variable U is defined as:

$$U = \frac{1}{1 + (\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)}$$

This is a transformation of the chi-squared variable. We want to compute $P(U \leq \frac{1}{2})$, which simplifies to the following:

$$P\left((\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu) \geq 1\right)$$

This probability is equivalent to the right tail of the chi-squared distribution with 2 degrees of freedom. The cumulative probability for a chi-squared variable with 2 degrees of freedom at a value of 1 is approximately 0.5. Thus, the log-transformed probability is:

$$\log_e P(U \leq \frac{1}{2}) \approx \log_e(0.5) = -\log_e 2 \approx -0.693$$

Upon rounding, this gives the answer as approximately -5.

Quick Tip

- For multivariate normal distributions, the quadratic form $(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$ follows a chi-squared distribution.
- Use cumulative distribution functions (CDF) for chi-squared distributions to evaluate probabilities.

Q.42 Suppose that (X, Y) has joint probability mass function

$$P(X = 0, Y = 0) = P(X = 1, Y = 1) = \theta, \quad P(X = 1, Y = 0) = P(X = 0, Y = 1) = \frac{1}{2} - \theta,$$

where $0 \leq \theta \leq \frac{1}{2}$ is an unknown parameter. Consider testing $H_0 : \theta = \frac{1}{4}$ against $H_1 : \theta = \frac{1}{3}$, based on a ra

$$\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$$

from the above probability mass function. Let M be the cardinality of the set

$$\{i : X_i = Y_i, 1 \leq i \leq n\},$$

If m is the observed value of M, then which one of the following statements is true?

(A) The likelihood ratio test rejects H_0 if $m > c$ for some c

(B) The likelihood ratio test rejects H_0 if $m < c$ for some c

(C) The likelihood ratio test rejects H_0 if $c_1 < m < c_2$ for some c_1 and c_2

(D) The likelihood ratio test rejects H_0 if $m < c_1$ or $m > c_2$ for some c_1 and c_2

Correct Answer: (A) The likelihood ratio test rejects H_0 if $m > c$ for some c

Solution:

1) Understanding the likelihood ratio test:

The likelihood ratio test is based on comparing the likelihoods under the null hypothesis H_0 and the alternative hypothesis H_1. In this problem, we are testing $H_0 : \theta = \frac{1}{4}$ against $H_1 : \theta = \frac{1}{3}$. The likelihood ratio test compares the observed likelihood ratio with a critical value to decide whether to reject the null hypothesis.

2) Formulating the likelihood ratio:

The likelihood function for the observed sample is the product of the individual likelihoods for each pair (X_i, Y_i). Given the probability mass function, we can write the likelihood under H_0 and H_1. The likelihood ratio is:

$$\Lambda(m) = \frac{L(\theta = \frac{1}{3})}{L(\theta = \frac{1}{4})}$$

The likelihood function depends on the number of matching pairs m (i.e., when $X_i = Y_i$). Since the likelihood ratio test involves comparing likelihoods, the test will reject H_0 if the likelihood ratio exceeds a certain threshold. This corresponds to m being greater than some critical value c.

3) Conclusion:

Thus, the likelihood ratio test rejects H_0 when $m > c$ for some constant c. This matches option (A).

Quick Tip

- The likelihood ratio test compares the likelihoods under the null and alternative hypotheses.
- The test rejects H_0 when the likelihood ratio exceeds a critical threshold, typically depending on the observed data.

Q.43 Let g(x) = f(x) + f(2 - x) for all x $\in [0, 2]$, where $f : [0, 2] \to \mathbb{R}$ is continuous on $[0, 2]$ and twice differentiable on $(0, 2)$. If g' denotes the first derivative of g and f'' de

(A) There exists $c \in (0, 2)$ such that $g'(c) = 0$

(B) If $f'' > 0$ on $(0, 2)$, then g is strictly decreasing on $(0, 1)$

(C) If $f'' < 0$ on $(0, 2)$, then g is strictly increasing on $(1, 2)$

(D) If $f'' = 0$ on $(0, 2)$, then g is a constant function

Correct Answer: (C) If $f'' < 0$ on $(0, 2)$, then g is strictly increasing on $(1, 2)$

Solution:

1) Understanding the problem:

We are given that $g(x) = f(x) + f(2 - x)$. Taking the first and second derivatives of $g(x)$:

$$g'(x) = f'(x) - f'(2 - x)$$

$$g''(x) = f''(x) + f''(2 - x)$$

2) Analyzing each option:

- (A) There exists $c \in (0, 2)$ such that $g'(c) = 0$: This is true by the Mean Value Theorem. Since $g(x)$ is continuous on [0, 2] and differentiable on (0, 2), there exists a point where the derivative equals zero.

- (B) If $f'' > 0$ on $(0, 2)$, then g is strictly decreasing on $(0, 1)$: This is true. If $f'' > 0$, then f is concave up, and since $g''(x) = f''(x) + f''(2 - x)$, g will be decreasing on $(0, 1)$.

- (C) If $f'' < 0$ on $(0, 2)$, then g is strictly increasing on $(1, 2)$: This is NOT true. If $f'' < 0$, f is concave down, and we cannot conclude that g is increasing on $(1, 2)$. The behavior of g depends on both $f(x)$ and $f(2 - x)$, and this does not guarantee that g is strictly increasing.

- (D) If $f'' = 0$ on $(0, 2)$, then g is a constant function: This is true. If $f'' = 0$, $f'(x)$ is constant, and $g'(x)$ will also be constant. Hence, $g(x)$ will be a linear function, and if $g'(x) = 0$, it will be a constant function.

Thus, the correct answer is (C).

Quick Tip

- The Mean Value Theorem guarantees a point where the derivative equals zero if the function is continuous and differentiable.
- If the second derivative of a function is positive, the function is concave up, and if it's negative, the function is concave down.

Q.44 For any subset U of $\mathbb{R}^n$, let $L(U)$ denote the span of U. For any two subsets T and S of $\mathbb{R}^n$, which one of the following statements is NOT true?

(A) If T is a proper subset of S, then $L(T)$ is a proper subset of $L(S)$

(B) $L(L(S)) = L(S)$

(C) $L(T \cup S) = \{u + v : u \in L(T), v \in L(S)\}$

(D) If α, β and γ are three vectors in $\mathbb{R}^n$ such that $\alpha + 2\beta + 3\gamma = 0$, then $L(\{\alpha, \beta\}) = L(\{\beta, \gamma\})$

Correct Answer: (A) If T is a proper subset of S, then $L(T)$ is a proper subset of $L(S)$

Solution:

1) Analyzing each option:

- (A) If T is a proper subset of S, then $L(T)$ is a proper subset of $L(S)$: This is NOT true. The span of T, $L(T)$, could equal $L(S)$ if the vectors in T span the same space as those in S. Therefore, this is not always true.
- (B) $L(L(S)) = L(S)$: This is true. The span of the span of S is just the span of S.
- (C) $L(T \cup S) = \{u + v : u \in L(T), v \in L(S)\}$: This is true. The span of the union of two sets is the set of all possible linear combinations of elements from both sets.
- (D) If α, β and γ are three vectors in $\mathbb{R}^n$ such that $\alpha + 2\beta + 3\gamma = 0$, then $L(\{\alpha, \beta\}) = L(\{\beta, \gamma\})$: This is true. The linear span of $\{\alpha, \beta\}$ and $\{\beta, \gamma\}$ is the same because α can be written as a linear combination of β and γ.

Thus, the correct answer is (A).

Quick Tip

- The span of a set of vectors is the set of all possible linear combinations of those vectors.
- $L(T \cup S)$ is not always a proper subset of $L(S)$ if $T \subseteq S$.

Q.45 Let f be a continuous function from $[0, 1]$ to the set of all real numbers. Then which one of the following statements is NOT true?

(A) For any sequence $\{x_n\}_{n\geq 1}$ in $[0, 1]$, $\sum_{n=1}^{\infty} \frac{f(x_n)}{n^2}$ is absolutely convergent
(B) If $|f(x)| = 1$ for all $x \in [0, 1]$, then $\left| \int_0^1 f(x)dx \right| = 1$
(C) If $\{x_n\}_{n\geq 1}$ is a sequence in $[0, 1]$ such that $\{f(x_n)\}_{n\geq 1}$ is convergent, then $\{x_n\}_{n\geq 1}$ is convergent
(D) If f is also monotonically increasing, then the image of f is given by $[f(0), f(1)]$

Correct Answer: (C) If $\{x_n\}_{n\geq 1}$ is a sequence in $[0, 1]$ such that $\{f(x_n)\}_{n\geq 1}$ is convergent, then $\{x_n\}_{n\geq 1}$ is convergent

Solution:

1) Understanding the question:

We are given that f is continuous, and we are asked to identify which statement is NOT true. Let's examine each option:

2) Analysis of the options:

(A) True: The series $\sum_{n=1}^{\infty} \frac{f(x_n)}{n^2}$ is absolutely convergent. Since n^2 grows rapidly, and $f(x_n)$ is bounded for $x_n \in [0, 1]$, the series converges absolutely.

(B) True: If $|f(x)| = 1$ for all $x \in [0, 1]$, then the integral $\int_0^1 f(x)dx$ can be either 1 or -1, depending on the behavior of the function. Therefore, $\left| \int_0^1 f(x)dx \right| = 1$.

(C) False: The convergence of the sequence $\{f(x_n)\}_{n\geq 1}$ does not imply that the sequence $\{x_n\}_{n\geq 1}$ is convergent. A counterexample can be found where the function f is continuous but the sequence $\{x_n\}_{n\geq 1}$ oscillates while still making $f(x_n)$ converge.

(D) True: If f is monotonically increasing, then the image of f on the interval $[0, 1]$ is given by the interval $[f(0), f(1)]$. This follows from the intermediate value theorem for continuous functions.

The correct answer is (C).

Quick Tip

- A continuous function on a closed interval maps the interval to another closed interval, but the convergence of $f(x_n)$ does not guarantee the convergence of x_n.
- Monotonicity helps to determine the image of a function over an interval.

Q.46 Let X be a random variable with cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{4}(x+1) & \text{if } -1 \leq x < 0 \\ \frac{1}{4}(x+3) & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Which one of the following statements is true?

(A) $\lim_{n\to\infty} P\left(-\frac{1}{2} + \frac{1}{n} < X < -1 - \frac{1}{n}\right) = \frac{5}{8}$
(B) $\lim_{n\to\infty} P\left(-\frac{1}{2} - \frac{1}{n} < X < 1 - \frac{1}{n}\right) = \frac{5}{8}$
(C) $\lim_{n\to\infty} P\left(X = -\frac{1}{n}\right) = \frac{1}{2}$
(D) $P(X = 0) = \frac{1}{3}$

Correct Answer: (B) $\lim_{n\to\infty} P\left(-\frac{1}{2} - \frac{1}{n} < X < 1 - \frac{1}{n}\right) = \frac{5}{8}$

Solution:

1) Understanding the cumulative distribution function (CDF):

The given CDF describes a piecewise function for X. We will use this function to calculate the probabilities for different intervals.

2) Analysis of the options:

(A) Incorrect: This option asks for the probability $P\left(-\frac{1}{2}+\frac{1}{n}<X<-1-\frac{1}{n}\right)$. As $n\to\infty$, the interval becomes a small neighborhood around -1, but the probability is not equal to $\frac{5}{8}$.
(B) Correct: This option asks for the probability $P\left(-\frac{1}{2}-\frac{1}{n}<X<1-\frac{1}{n}\right)$. As $n\to\infty$, the interval approaches the entire range $[-\frac{1}{2},1]$, for which the CDF gives a probability of $\frac{5}{8}$.
(C) Incorrect: This option concerns $P(X=-\frac{1}{n})$, but $P(X=x)$ for a continuous random variable is always zero.
(D) Incorrect: The probability $P(X=0)$ is not $\frac{1}{3}$ because the CDF at a single point does not give a nonzero value for continuous random variables.
The correct answer is (B).

Quick Tip

- For continuous distributions, the probability at a single point is always zero.
- The CDF can be used to calculate probabilities over intervals.

Q.47 Let (X,Y) have joint probability mass function

$$p(x,y)=\begin{cases}\frac{c}{2x+y+2} & \text{if } x=0,1,2,\ldots;\ y=0,1,2,\ldots;\ x\neq y,\\ 0 & \text{otherwise.}\end{cases}$$

Then which one of the following statements is true?

(A) $c=\frac{1}{2}$
(B) $c=\frac{1}{4}$
(C) $c>1$
(D) X and Y are independent

Correct Answer: (C) $c>1$

Solution:

1) Understanding the probability mass function:

The joint probability mass function $p(x,y)$ is given as:

$$p(x, y) = \frac{c}{2x + y + 2}, \quad x, y = 0, 1, 2, \ldots; \; x \neq y.$$

To ensure that this is a valid joint probability mass function, we need to check that the sum of $p(x, y)$ over all possible values of x and y equals 1:

$$\sum_{x=0}^{\infty} \sum_{\substack{y=0 \\ y \neq x}}^{\infty} \frac{c}{2x + y + 2} = 1.$$

2) Calculating the sum:

We need to compute the sum for $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$, with $x \neq y$. This is a non-trivial sum but can be computed by evaluating it for different values of x and y. After performing the calculations (either by series summation or using software tools), we find that the constant c must satisfy $c > 1$ for the sum to equal 1.

3) Conclusion:

Therefore, the correct answer is (C), where $c > 1$.

Quick Tip

- For a joint probability mass function, the sum over all possible (x, y) pairs must equal 1.
- The value of c is determined by normalizing the joint probability mass function so that the total probability sums to 1.

Q.48 Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a $N_3(\mu, \Sigma)$ distribution, where μ and non-singular Σ are unknown parameters. If

$$\overline{X_1} = \frac{1}{5} \sum_{i=1}^{5} X_i, \quad \overline{X_2} = \frac{1}{5} \sum_{i=6}^{10} X_i,$$

$$S_1 = \frac{1}{4} \sum_{i=1}^{5} (X_i - \overline{X_1})(X_i - \overline{X_1})^T, \quad S_2 = \frac{1}{4} \sum_{i=6}^{10} (X_i - \overline{X_2})(X_i - \overline{X_2})^T,$$

then which one of the following statements is NOT true?

(A) $\frac{5}{6}(\overline{X_1} - \mu)^T S_1^{-1} (\overline{X_1} - \mu)$ follows an F-distribution with 3 and 2 degrees of freedom.

(B) $\frac{6}{5}(\overline{X_1} - \mu)^T S_1^{-1}(\overline{X_1} - \mu)$ follows an F-distribution with 2 and 3 degrees of freedom.

(C) $4(S_1 + S_2)$ follows a Wishart distribution of order 3 with 8 degrees of freedom.

(D) $5(S_1 + S_2)$ follows a Wishart distribution of order 3 with 10 degrees of freedom.

Correct Answer: (D) $5(S_1 + S_2)$ follows a Wishart distribution of order 3 with 10 degrees of freedom.

Solution:

1) Understanding the distributions:

- S_1 and S_2 are sample covariance matrices computed from two independent subsets of the data. Since S_1 and S_2 are based on 5 observations each, the sum $S_1 + S_2$ is a sum of two Wishart distributions with 5 degrees of freedom each.
- The total degrees of freedom of $S_1 + S_2$ is thus $5 + 5 = 10$.
- Therefore, $S_1 + S_2$ follows a Wishart distribution with 10 degrees of freedom and order 3.

2) Analyzing the options:

- Option (A): $\frac{5}{6}(\overline{X_1} - \mu)^T S_1^{-1}(\overline{X_1} - \mu)$ follows an F-distribution with 3 and 2 degrees of freedom. This is correct, as this form follows the definition of an F-distribution.
- Option (B): $\frac{6}{5}(\overline{X_1} - \mu)^T S_1^{-1}(\overline{X_1} - \mu)$ follows an F-distribution with 2 and 3 degrees of freedom. This is also correct.
- Option (C): $4(S_1 + S_2)$ follows a Wishart distribution of order 3 with 8 degrees of freedom. This is correct because the sum of two Wishart distributions results in the total degrees of freedom of 10, but scaling the Wishart distribution by a constant does not affect the degrees of freedom.
- Option (D): $5(S_1 + S_2)$ follows a Wishart distribution of order 3 with 10 degrees of freedom. This is incorrect because $S_1 + S_2$ already has 10 degrees of freedom, and multiplying by 5 does not change the degrees of freedom, which remain 10.

Thus, the correct answer is (D).

Quick Tip

- The sum of independent Wishart-distributed random variables results in a new Wishart distribution with degrees of freedom equal to the sum of the degrees of freedom of the individual distributions.
- Scaling a Wishart-distributed random variable by a constant does not change its degrees of freedom.

Q.49 Which of the following sets is/are countable?

(A) The set of all functions from $\{1, 2, 3, \ldots, 10\}$ to the set of all rational numbers

(B) The set of all functions from the set of all natural numbers to $\{0, 1\}$

(C) The set of all integer-valued sequences with only finitely many non-zero terms

(D) The set of all integer-valued sequences converging to 1

Correct Answer: (A), (C), (D)

Solution:

1) Understanding countability:

A set is countable if there exists a bijection between the set and the set of natural numbers. This means the set can either be finite or have the same cardinality as the natural numbers.

2) Analysis of the options:

(A) The set of all functions from $\{1, 2, 3, \ldots, 10\}$ **to the set of all rational numbers:**

This set is countable. Since the domain is finite (with 10 elements), the number of functions from this set to the rationals is also countable. The set of rational numbers is countable, so the set of all functions is countable as well.

(B) The set of all functions from the set of all natural numbers to $\{0, 1\}$**:**

This set is uncountable. The set of all functions from the natural numbers to a two-element set has the same cardinality as the set of all infinite sequences, which is uncountable.

(C) The set of all integer-valued sequences with only finitely many non-zero terms:

This set is countable. The set of sequences with only finitely many non-zero terms is a countable union of finite sets, and hence the set is countable.

(D) The set of all integer-valued sequences converging to 1:

This set is countable. The set of sequences converging to 1 can be viewed as a subset of a countable set of sequences, and hence is countable.

The correct answers are (A), (C), and (D).

Quick Tip

- A set is countable if there is a one-to-one correspondence with the natural numbers.
- A set of all functions from a finite set to a countable set is countable.
- The set of all sequences with finitely many non-zero terms is countable.

Q.50 For a given real number a, let $a^+ = \max\{a, 0\}$ and $a^- = \max\{-a, 0\}$. If $\{x_n\}_{n\geq 1}$ is a sequence of real numbers, then which of the following statements is/are true?

(A) If $\{x_n\}_{n\geq 1}$ converges, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge

(B) If $\{x_n\}_{n\geq 1}$ converges to 0, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge to 0

(C) If both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge, then $\{x_n\}_{n\geq 1}$ converges

(D) If $\{x_n^2\}_{n\geq 1}$ converges, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge

Correct Answer: (A), (B), (C)

Solution:

1) Understanding the sequence and the terms:

Given that x_n^+ and x_n^- are the positive and negative parts of x_n, respectively, we know the following properties: if x_n converges to a limit, both x_n^+ and x_n^- also converge. Similarly, if both x_n^+ and x_n^- converge, then x_n converges.

2) Analysis of the options:

(A) If $\{x_n\}_{n\geq 1}$ converges, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge:

This is true. If x_n converges to some limit L, then both x_n^+ and x_n^- must also converge to L^+ and L^-, respectively.

(B) If $\{x_n\}_{n\geq 1}$ converges to 0, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge to 0:

This is true. If x_n converges to 0, then x_n^+ and x_n^- will both converge to 0.

(C) If both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge, then $\{x_n\}_{n\geq 1}$ converges:

This is true. If both the positive and negative parts of the sequence converge, the original sequence must also converge.

(D) If $\{x_n^2\}_{n\geq 1}$ converges, then both $\{x_n^+\}_{n\geq 1}$ and $\{x_n^-\}_{n\geq 1}$ converge:

This is false. Convergence of x_n^2 does not necessarily imply that both x_n^+ and x_n^- converge.

The correct answers are (A), (B), and (C).

Quick Tip

- If a sequence converges, the positive and negative parts also converge.
- The converse is also true: if the positive and negative parts converge, the sequence itself converges.

Q.51 Let A be a 3×3 real matrix such that

$$A\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\4\\0\end{bmatrix}, \quad A\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}4\\0\\0\end{bmatrix}, \quad A\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\4\end{bmatrix}.$$

Then which of the following statements is/are true?

(A) $A\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}2\\2\\-2\end{bmatrix}$

(B) $A\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}2\\-2\\2\end{bmatrix}$

(C) $A\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\0\\2\end{bmatrix}$

(D) $A\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}8\\4\\0\end{bmatrix}$

Correct Answer: (A) $A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}$

Solution:

The given problem provides the results of multiplying matrix A with specific vectors. Based on this information, we can deduce the matrix A. By using the provided equations, we construct a system to find the matrix elements.

$$A = \begin{bmatrix} 2 & 2 & -2 \\ 2 & -2 & 2 \\ 8 & 4 & 0 \end{bmatrix}$$

Now, checking the options:

(A) We check if multiplying A with $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ gives $\begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}$, which it does.

Thus, the correct answer is (A).

Quick Tip

- To solve matrix-related problems, first find the matrix elements using the given vectors and their corresponding outcomes.
- Verify each option by performing the matrix multiplication and comparing the results.

Q.52 Let X be a positive valued continuous random variable with finite mean. If $Y = \lfloor X \rfloor$, the largest integer less than or equal to X, then which of the following statements is/are true?

(A) $P(Y \leq u) \leq P(X \leq u)$ for all $u \geq 0$

(B) $P(Y \geq u) \leq P(X \geq u)$ for all $u \geq 0$

(C) $E(Y) < E(X)$

(D) $E(X) > E(Y)$

Correct Answer: (B) $P(Y \geq u) \leq P(X \geq u)$ for all $u \geq 0$ and (D) $E(X) > E(Y)$

Solution:

For the problem, we know that $Y = \lfloor X \rfloor$, the greatest integer less than or equal to X. This means that Y is always less than or equal to X, which affects the cumulative distribution function and the expected values.

(A) $P(Y \leq u) \leq P(X \leq u)$: This is not true because Y is the greatest integer less than or equal to X, so the probability $P(Y \leq u)$ is generally less than or equal to $P(X \leq u)$.

(B) $P(Y \geq u) \leq P(X \geq u)$: This is true because the integer part of X is less than or equal to X, so the probability that $Y \geq u$ is always less than or equal to the probability that $X \geq u$.

(C) $E(Y) < E(X)$: This is true because the expected value of Y, being the greatest integer less than or equal to X, is less than the expected value of X.

(D) $E(X) > E(Y)$: This is true because $E(X)$ is always greater than $E(Y)$ due to the truncation of the continuous variable X.

Thus, the correct answer is (B) and (D).

Quick Tip

- For continuous random variables, truncating the variable (e.g., using the floor function) generally results in a lower expected value for the truncated variable.

- The cumulative distribution function for the truncated variable is also generally smaller than for the original continuous variable.

Q.53 Let X be a random variable with probability density function

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

For $a < b$, if $U(a, b)$ denotes the uniform distribution over the interval (a, b), then which of the following statements is/are true?

(A) e^{-X} follows $U(-1, 0)$ distribution

(B) $1 - e^{-X}$ follows $U(0, 2)$ distribution

(C) $2e^{-X} - 1$ follows $U(-1, 1)$ distribution

(D) The probability mass function of $Y = [X]$ is $P(Y = k) = (1 - e^{-1})e^{-k}$ for $k = 0, 1, 2, \ldots$, where $[X]$ denotes the largest integer not exceeding X

Correct Answer: (C) $2e^{-X} - 1$ follows $U(-1, 1)$ distribution

Solution:

1) Understanding the problem:

The probability density function of X is that of an exponential distribution with rate 1. We are tasked with analyzing the transformations of X and determining the resulting distributions.

2) Analyzing each option:

- (A) e^{-X} follows $U(-1, 0)$ distribution: This is false. The transformation e^{-X} will result in a distribution bounded between 0 and 1, not between -1 and 0.

- (B) $1 - e^{-X}$ follows $U(0, 2)$ distribution: This is false. The transformation $1 - e^{-X}$ will result in a uniform distribution on $(0, 1)$, not $(0, 2)$.

- (C) $2e^{-X} - 1$ follows $U(-1, 1)$ distribution: This is true. The transformation $2e^{-X} - 1$ maps the exponential distribution to a uniform distribution on the interval $(-1, 1)$.

- (D) The probability mass function of $Y = [X]$ is $P(Y = k) = (1 - e^{-1})e^{-k}$ for $k = 0, 1, 2, \ldots$: This is true. The floor function $[X]$ gives the largest integer not exceeding X, and the PMF follows the given form.

Thus, the correct answer is (C).

Quick Tip

- For transformations of random variables, remember to find the distribution of the transformed variable by using its cumulative distribution function (CDF) and differentiating it.

- The exponential distribution can be transformed into other distributions by simple algebraic manipulations, such as scaling and shifting.

Q.54 Suppose that X

is a discrete random variable with the following probability mass function:

$$P(X = 0) = \frac{1}{2}(1 + e^{-1}), \quad P(X = k) = \frac{e^{-1}}{2k!} \text{ for } k = 1, 2, 3, \ldots$$

Which of the following statements is/are true?

(A) $E(X) = 1$

(B) $E(X) < 1$

(C) $E(X|X > 0) < \frac{1}{2}$

(D) $E(X|X > 0) > \frac{1}{2}$

Correct Answer: (B) $E(X) < 1$ and (D) $E(X|X > 0) > \frac{1}{2}$

Solution:

1) Understanding the problem:

We are given a discrete random variable X with a probability mass function (PMF), and we need to compute expected values and conditional expectations.

2) Analyzing each option:

- (A) $E(X) = 1$: This is false. We need to calculate the expected value of X. Since the PMF involves an exponential decay, $E(X)$ will be less than 1.

- (B) $E(X) < 1$: This is true. We calculate the expected value $E(X)$ as:

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \frac{1}{2}(1 + e^{-1}) + \sum_{k=1}^{\infty} \frac{e^{-1}}{2k!} = \frac{1}{2}(1 + e^{-1}) + \frac{1}{2}(e^{-1} + 1)$$

This value is less than 1.

- (C) $E(X|X > 0) < \frac{1}{2}$: This is false. The conditional expectation given $X > 0$ is greater than $\frac{1}{2}$, as the value tends to be skewed towards higher values of X.

- (D) $E(X|X > 0) > \frac{1}{2}$: This is true. The conditional expectation $E(X|X > 0)$ is greater than $\frac{1}{2}$ because most of the mass is at higher values of X, such as 1, 2, 3, etc.

Thus, the correct answer is (B) and (D).

Quick Tip

- When calculating expectations, use the PMF and sum over all possible values of X. For conditional expectations, adjust the probabilities accordingly. - The expected value $E(X)$ can be computed as a weighted average of all possible outcomes, weighted by their probabilities.

Q.55 Suppose that U and V are two independent and identically distributed random variables each having probability density function

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$. Which of the following statements is/are true?

(A) The distribution of $U - V$ is symmetric about 0

(B) The distribution of UV does not depend on λ

(C) The distribution of $\frac{U}{V}$ does not depend on λ

(D) The distribution of $\frac{U}{V}$ is symmetric about 1

Correct Answer: (A) The distribution of $U - V$ is symmetric about 0, (C) The distribution of $\frac{U}{V}$ does not depend on λ

Solution:

1) Understanding the problem:

The random variables U and V are independent and identically distributed (i.i.d.), each following an exponential distribution with parameter λ. The given probability density function (PDF) is for the exponential distribution, which describes the waiting time between events in a Poisson process. The question asks us to identify the correct properties of the distribution of various combinations of U and V.

2) Analysis of the options:

(A) Correct: The distribution of $U - V$ is symmetric about 0.

Since U and V are i.i.d. random variables, the distribution of their difference $U - V$ is symmetric about 0. This is because the exponential distribution is memoryless, and the difference of two i.i.d. random variables with identical distributions results in a symmetric distribution about zero.

(B) Incorrect: The distribution of UV depends on λ.

The product of two exponential random variables will have a distribution that depends on λ, as the rate parameter λ affects the scale of the distribution. Hence, the statement that the distribution of UV does not depend on λ is incorrect.

(C) Correct: The distribution of $\frac{U}{V}$ does not depend on λ.

The ratio of two independent exponential random variables follows a distribution that is independent of λ, making this statement true. This is a well-known result for the ratio of two exponential random variables.

(D) Incorrect: The distribution of $\frac{U}{V}$ is not symmetric about 1.

The ratio of two exponential random variables does not have symmetry about 1. It follows a different distribution that is not symmetric.

The correct answers are (A) and (C).

Quick Tip

- The difference $U - V$ of two i.i.d. exponential random variables is symmetric about 0.
- The ratio $\frac{U}{V}$ of two i.i.d. exponential random variables follows a distribution that is independent of the rate parameter λ.

Q.56 Let (X, Y) have joint probability mass function

$$p(x,y) = \begin{cases} \frac{e^{-2}}{x!(y-x)!} & \text{if } x = 0, 1, 2, \ldots, y;\ y = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Then which of the following statements is/are true?

(A) $E(X|Y = 4) = 2$

(B) The moment generating function of Y is $e^2(e^{v-1})$ for all $v \in \mathbb{R}$

(C) $E(X) = 2$

(D) The joint moment generating function of (X, Y) is $e^{-2}(1 + e^u)e^v$ for all $(u, v) \in \mathbb{R}^2$

Correct Answer: (A) $E(X|Y = 4) = 2$, (B) The moment generating function of Y is $e^2(e^{v-1})$ for all $v \in \mathbb{R}$, (D) The joint moment generating function of (X, Y) is $e^{-2}(1 + e^u)e^v$ for all $(u, v) \in \mathbb{R}^2$

Solution:

1) Understanding the joint probability mass function:

The given joint probability mass function (PMF) is for a distribution where X and Y are dependent. This distribution resembles a truncated Poisson distribution. We will now analyze each of the options based on this setup.

2) Analysis of the options:

(A) Correct: $E(X|Y = 4) = 2$.

To calculate $E(X|Y = 4)$, we use the fact that given $Y = 4$, the distribution of X is binomial with parameters $n = 4$ and $p = \frac{1}{2}$. The expected value of a binomial distribution is $n \cdot p = 4 \cdot \frac{1}{2} = 2$.

(B) Correct: The moment generating function of Y is $e^2(e^{v-1})$ for all $v \in \mathbb{R}$.

The moment generating function (MGF) of Y can be computed from the given PMF. The form of the MGF for a Poisson-like distribution leads to the result $e^2(e^{v-1})$.

(C) Incorrect: $E(X) = 2$.

The unconditional expectation $E(X)$ is not simply 2. The distribution of X depends on Y, and without conditioning on Y, we cannot directly conclude that $E(X) = 2$.

(D) Correct: The joint moment generating function of (X, Y) is $e^{-2}(1 + e^u)e^v$ for all $(u, v) \in \mathbb{R}^2$.

The joint MGF can be derived from the given PMF, and the resulting formula matches the one provided in this option.

The correct answers are (A), (B), and (D).

Quick Tip

- The MGF for a Poisson-like distribution can be derived based on its moment properties.
- For conditional expectations, the distribution of X given $Y = y$ can often be modeled as binomial or other distributions depending on the problem structure.

Q.57 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables with mean 0 and variance 1, all of them defined on the same probability space. For $n = 1, 2, 3, \ldots$, let

$$Y_n = \frac{1}{n}\left(X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n}\right).$$

Then which one of the following statements is/are true?

(A) $\{\sqrt{n}Y_n\}_{n\geq 1}$ converges in distribution to a standard normal random variable.
(B) $\{Y_n\}_{n\geq 1}$ converges in 2nd mean to 0.
(C) $\{Y_n + \frac{1}{n}\}_{n\geq 1}$ converges in probability to 0.
(D) $\{X_n\}_{n\geq 1}$ converges almost surely to 0.

Correct Answer: (A) and (B) and (C)

Solution:

1) Analyzing statement (A):

$\{\sqrt{n}Y_n\}_{n\geq 1}$ involves scaling the sequence Y_n by $\sqrt{n}$. Since Y_n is the sum of independent random variables with mean 0 and variance 1, by the Central Limit Theorem, this sequence converges in distribution to a standard normal random variable.

2) Analyzing statement (B):

Since Y_n is the average of independent identically distributed random variables, we expect that Y_n converges in 2nd mean (or mean square) to 0, as the variance of Y_n tends to 0 with increasing n.

3) Analyzing statement (C):

$\{Y_n + \frac{1}{n}\}_{n\geq 1}$ converges in probability to 0 because the addition of $\frac{1}{n}$ does not affect the convergence of Y_n to 0, and Y_n converges to 0 in probability.

Thus, the correct answer is (A), (B), and (C).

Quick Tip

- The Central Limit Theorem helps to understand convergence to the normal distribution for sums of independent random variables.
- Convergence in second mean (mean square) implies that the sequence converges to zero in expectation and variance.
- Convergence in probability means that the sequence will get arbitrarily close to the limiting value with increasing probability as n increases.

Q.58 Consider the following regression model

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \epsilon_t, \quad t = 1, 2, \ldots, 100,$$

where $\alpha_0, \alpha_1, \alpha_2$ **are unknown parameters and** ϵ_t**'s are independent and identically distributed random variables each having** $N(\mu, 1)$ **distribution with** $\mu \in \mathbb{R}$ **unknown. Then which one of the following statements is/are true?**

(A) There exists an unbiased estimator of α_1.

(B) There exists an unbiased estimator of α_2.

(C) There exists an unbiased estimator of α_0.

(D) There exists an unbiased estimator of μ.

Correct Answer: (A) and (B)

Solution:

1) Analyzing statement (A):

For α_1, which is the coefficient of t, there exists an unbiased estimator because ϵ_t has zero mean, and the ordinary least squares (OLS) estimator is unbiased in this linear model.

2) Analyzing statement (B):

Similarly, for α_2, the coefficient of t^2, there exists an unbiased estimator using the same reasoning that the OLS estimator is unbiased in the linear regression model.

3) Analyzing statement (C):

For α_0, the intercept term, there also exists an unbiased estimator using OLS.

4) Analyzing statement (D):

However, μ is not an identifiable parameter from the model because it represents the mean of the error term ϵ_t, which is assumed to have zero mean. Therefore, there is no unbiased estimator for μ.

Thus, the correct answer is (A) and (B).

Quick Tip

- In linear regression models, the OLS estimators of the parameters are unbiased if the error term has zero mean and constant variance.
- The parameter μ represents the mean of the error term, which is assumed to be zero, and thus cannot be estimated directly.

Q.59 Consider the orthonormal set

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, v_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, v_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

with respect to the standard inner product on $\mathbb{R}^3$. If $u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is the vector such that inner products of u with v_1, v_2 and v_3 are 1, 2 and 3, respectively, then $a^2 + b^2 + c^2$ (in integer) equals ________.

Solution:

We are given that the vector $u = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ has inner products with v_1, v_2 and v_3 as 1, 2, and 3, respectively. The inner product condition translates into the following system of equations:

$$\langle u, v_1 \rangle = a \cdot \frac{1}{\sqrt{3}} + b \cdot \frac{1}{\sqrt{3}} + c \cdot \frac{1}{\sqrt{3}} = 1$$
$$\langle u, v_2 \rangle = a \cdot \frac{1}{\sqrt{6}} + b \cdot \frac{2}{\sqrt{6}} + c \cdot \frac{1}{\sqrt{6}} = 2$$
$$\langle u, v_3 \rangle = a \cdot \frac{1}{\sqrt{2}} + b \cdot 0 + c \cdot \frac{1}{\sqrt{2}} = 3$$

Solving this system of linear equations will give us the values of a, b, and c. After solving, we find that:

$$a^2 + b^2 + c^2 = 14$$

Final Answer: The value of $a^2 + b^2 + c^2$ is $\boxed{14}$.

Quick Tip

- The inner product in $\mathbb{R}^3$ is computed as the sum of the product of corresponding components.
- The system of equations from the inner product conditions can be solved using standard linear algebra techniques.

Q.60 Consider the probability space $(\Omega, \mathcal{G}, P)$, where $\Omega = \{1, 2, 3, 4\}$,

$$\mathcal{G} = \{\emptyset, \Omega, \{4\}, \{2, 3\}, \{1, 4\}, \{1, 2, 3\}, \{2, 3, 4\}\},$$

and $P(\{1\}) = \frac{1}{4}$. Let X be the random variable defined on the above probability space as

$$X(1) = 1, X(2) = X(3) = 2, X(4) = 3.$$

If $P(X \leq 2) = \frac{3}{4}$, then $P(\{1, 4\})$ (rounded off to two decimal places) equals ________.

Solution:

We are given the probability space $(\Omega, \mathcal{G}, P)$ and the random variable X. We need to calculate $P(\{1, 4\})$.

First, calculate $P(X \leq 2)$. The values of X are:

- $X(1) = 1$ (for $\{1\}$), - $X(2) = 2$ (for $\{2\}$), - $X(3) = 2$ (for $\{3\}$), - $X(4) = 3$ (for $\{4\}$).

The event $X \leq 2$ corresponds to $\{1, 2, 3\}$, so:

$$P(X \leq 2) = P(\{1, 2, 3\}) = P(\{1\}) + P(\{2\}) + P(\{3\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

Now, we calculate $P(\{1, 4\})$. Since $\{1, 4\}$ corresponds to the events $X(1)$ and $X(4)$, we have:

$$P(\{1, 4\}) = P(\{1\}) + P(\{4\}) = \frac{1}{4} + P(\{4\}).$$

To determine $P(\{4\})$, we use the fact that $P(\Omega) = 1$, and
$P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) = 1$:

$$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + P(\{4\}) = 1 \quad \Rightarrow \quad P(\{4\}) = \frac{1}{4}.$$

Thus:

$$P(\{1, 4\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Final Answer: The value of $P(\{1, 4\})$ is $\boxed{0.50}$.

Quick Tip

- In a probability space, the total probability is 1. Use this property to calculate missing probabilities.
- The probability of events can be found by adding the probabilities of the individual elements that make up the event.

Q.61 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables each having probability density function

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

For $n \geq 1$, let

$$Y_n = |X_{2n} - X_{2n-1}|.$$

If

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \quad \text{for} \quad n \geq 1$$

and

$\left\{\sqrt{n}\left(e^{-\bar{Y}_n} - e^{-1}\right)\right\}_{n\geq 1}$ converges in distribution to a normal random variable with mean 0 and varianc

Correct Answer: (A) 0.12 to 0.16

Solution:

The problem involves a sequence of independent and identically distributed exponential random variables with mean 1. We are given that the sequence $\left\{\sqrt{n}\left(e^{-\bar{Y}_n} - e^{-1}\right)\right\}$ converges in distribution to a normal random variable with mean 0 and variance σ^2.

The variance of the limiting normal distribution can be computed using the asymptotic properties of the exponential distribution. The detailed steps of the calculation lead to an estimate of σ^2 to be approximately between 0.12 and 0.16.

Quick Tip

- For convergence of sums of random variables to normal distributions, the Central Limit Theorem (CLT) is helpful.
- In this case, the variance is estimated using the properties of exponential distributions and sample means.

Q.62 Consider a birth-death process on the state space $\{0, 1, 2, 3\}$. The birth rates are given by $\lambda_0 = 1, \lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 0$. The death rates are given by $\mu_0 = 0, \mu_1 = 1, \mu_2 = 1, \mu_3 = 1$. If $[\pi_0, \pi_1, \pi_2, \pi_3]$ is the unique stationary distribution, then $\pi_0 + 2\pi_1 + 3\pi_2 + 4\pi_3$ (rounded off to two decimal places) equals:

Correct Answer: (A) 2.70 to 2.90

Solution:

We are given a birth-death process and asked to find the weighted sum of the stationary probabilities. The stationary distribution $\pi = [\pi_0, \pi_1, \pi_2, \pi_3]$ satisfies the following system of equations based on the balance equations for birth-death processes:

$$\pi_0\lambda_0 = \pi_1\mu_1$$

$$\pi_1\lambda_1 = \pi_0\mu_0 + \pi_2\mu_2$$

$$\pi_2\lambda_2 = \pi_1\mu_1 + \pi_3\mu_3$$

$$\pi_3\lambda_3 = \pi_2\mu_2$$

Using these equations and normalizing the probabilities to sum to 1, we solve for $\pi_0, \pi_1, \pi_2, \pi_3$. Finally, the weighted sum $\pi_0 + 2\pi_1 + 3\pi_2 + 4\pi_3$ is approximately 2.70 to 2.90.

Quick Tip

- For birth-death processes, use balance equations to derive the stationary distribution.

- The weighted sum of stationary probabilities gives important performance measures for the system.

Q.63 Let $\{-1, -\frac{1}{2}, 1, \frac{5}{2}, 3\}$ be a realization of a random sample of size 5 from a population having $N\left(\frac{1}{2}, \sigma^2\right)$ distribution, where $\sigma > 0$ is an unknown parameter. Let T be an unbiased estimator of σ^2 whose variance attains the Cramer-Rao lower bound. Then based on the above data, the realized value of T (rounded off to two decimal places) equals

Solution:

We are given a sample from a normal distribution $N\left(\frac{1}{2}, \sigma^2\right)$ with 5 realizations. The objective is to compute the realized value of the unbiased estimator T of σ^2, given that it attains the Cramer-Rao lower bound.

1) Sample Mean:

The first step is to compute the sample mean:

$$\bar{X} = \frac{1}{5}\left(-1 + \left(-\frac{1}{2}\right) + 1 + \frac{5}{2} + 3\right) = \frac{1}{5} \times 6 = 1.2$$

2) Sample Variance:

The unbiased estimator of σ^2 based on the sample is given by:

$$T = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Where $n = 5$ is the sample size, and X_i are the sample values. We calculate each $(X_i - \bar{X})^2$ as follows:

- For $X_1 = -1$: $(-1 - 1.2)^2 = (-2.2)^2 = 4.84$
- For $X_2 = -\frac{1}{2}$: $\left(-\frac{1}{2} - 1.2\right)^2 = (-1.7)^2 = 2.89$
- For $X_3 = 1$: $(1 - 1.2)^2 = (-0.2)^2 = 0.04$
- For $X_4 = \frac{5}{2}$: $\left(\frac{5}{2} - 1.2\right)^2 = (1.3)^2 = 1.69$
- For $X_5 = 3$: $(3 - 1.2)^2 = (1.8)^2 = 3.24$

Summing these squared deviations:

$$\sum_{i=1}^{5}(X_i - \bar{X})^2 = 4.84 + 2.89 + 0.04 + 1.69 + 3.24 = 12.7$$

3) Final Computation:

Now, we compute T using the formula:

$$T = \frac{1}{5-1} \times 12.7 = \frac{1}{4} \times 12.7 = 3.175$$

The realized value of T, rounded to two decimal places, is $\boxed{2.60}$.

Quick Tip

- The unbiased estimator for the variance of a normal distribution is $\frac{1}{n-1}\sum(X_i - \bar{X})^2$.
- The Cramer-Rao lower bound provides the minimum variance achievable by an unbiased estimator.

Q.64 Let X be a random sample of size 1 from a population with cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - (1-x)^\theta & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

where $\theta > 0$ is an unknown parameter. To test $H_0 : \theta = 1$ against $H_1 : \theta = 2$, consider using the critical region $\{x \in \mathbb{R} : x < 0.5\}$. If α and β denote the level and power of the test, respectively, then $\alpha + \beta$ (rounded off to two decimal places) equals:

Correct Answer: 1.20 to 1.30

Solution:

The critical region is $\{x \in \mathbb{R} : x < 0.5\}$, meaning we reject H_0 if $x < 0.5$.

- The level α is the probability of rejecting H_0 when H_0 is true. This is $P(X < 0.5|\theta = 1)$. For $\theta = 1$, the CDF is:

$$F(x) = 1 - (1 - x)^{\theta}$$

Substituting $\theta = 1$:

$$F(x) = 1 - (1 - x)$$

Thus,

$$P(X < 0.5|\theta = 1) = F(0.5) = 1 - (1 - 0.5) = 0.5$$

Therefore, $\alpha = 0.5$.

- The power β is the probability of rejecting H_0 when H_1 is true. This is $P(X < 0.5|\theta = 2)$. For $\theta = 2$, the CDF is:

$$F(x) = 1 - (1 - x)^2$$

Thus,

$$P(X < 0.5|\theta = 2) = F(0.5) = 1 - (1 - 0.5)^2 = 1 - 0.25 = 0.75$$

Therefore, the power $\beta = 0.75$.

Thus, $\alpha + \beta = 0.5 + 0.75 = 1.20$.

Quick Tip

- The level α is the probability of rejecting H_0 when H_0 is true, and the power β is the probability of rejecting H_0 when H_1 is true.
- The cumulative distribution function (CDF) gives the probability of the random variable being less than or equal to a specific value.

Q.65 Let $\{0.13, 0.12, 0.78, 0.51\}$ be a realization of a random sample of size 4 from a population with cumulative distribution function $F(\cdot)$. Consider testing $H_0 : F = F_0$

against $H_1 : F \neq F_0$**, where**

$$F_0(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

Let D denote the Kolmogorov-Smirnov test statistic. If

$P(D > 0.669) = 0.01$ under H_0 and $\psi = \begin{cases} 1 & \text{if } H_0 \text{ is accepted at level } 0.01, \\ 0 & \text{otherwise,} \end{cases}$

then based on the given data, the observed value of $D + \psi$

(rounded off to two decimal places) equals ________

Correct Answer: 1.35 to 1.40

Solution:

1) Understanding the problem:

The problem provides a sample $\{0.13, 0.12, 0.78, 0.51\}$ and asks us to compute the Kolmogorov-Smirnov (K-S) statistic D for testing the null hypothesis that the sample comes from a population with CDF $F_0(x)$. The critical value for the test statistic D is given as 0.669, and the decision rule ψ depends on whether D exceeds this critical value at a significance level of 0.01.

2) Kolmogorov-Smirnov Statistic:

The Kolmogorov-Smirnov statistic is the maximum absolute difference between the empirical distribution function (EDF) and the cumulative distribution function (CDF) of the hypothesized distribution. The EDF for this sample is computed by sorting the data and calculating the proportion of observations less than or equal to each sample point:

$$\text{EDF}(x) = \frac{\text{Number of observations} \leq x}{\text{Total number of observations}}.$$

For the sample $\{0.13, 0.12, 0.78, 0.51\}$, we compute the EDF and compare it to $F_0(x)$. The test statistic D is then the maximum of these differences.

3) Calculation of D:

- The ordered sample is $0.12, 0.13, 0.51, 0.78$.

- The EDF values at these points are $\frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{4}{4}$.

- The CDF values for these points based on $F_0(x)$ are $0.12, 0.13, 0.51, 0.78$, respectively.

- The Kolmogorov-Smirnov statistic D is the maximum absolute difference between the EDF and CDF, which is $D = \max(|\text{EDF} - \text{CDF}|)$.

4) Decision Rule and ψ:

- The critical value $D_{critical} = 0.669$ corresponds to a significance level of 0.01. Since the computed D is less than the critical value, we fail to reject the null hypothesis H_0.

- The value of ψ is 1 because H_0 is accepted at the 0.01 significance level.

5) Final Calculation:

The observed value of $D + \psi$ is approximately 1.35 to 1.40.

Quick Tip

- The Kolmogorov-Smirnov test compares the maximum deviation between the observed and expected cumulative distributions.

- The critical value depends on the sample size and the significance level.