# Foundations OF Data Engineering: Exam Questions

M.tech., AI & DS Ist Year

## Question 1

Students from three different high schools—School A, School B, and School C—participated in a state-wide mathematics competition.

- School A accounted for 50% of the total participants.
- School B accounted for 30% of the total participants.
- School C accounted for the remaining 20% of participants.

The historical pass rates for students from these schools are:

- P(Pass — School A) = 0.80
- P(Pass — School B) = 0.60
- P(Pass — School C) = 0.95

### (a)

If a student is selected at random, what is the overall probability that this student passed?

### (b)

Given that a randomly selected student has passed, what is the probability that they are from School B?

### (c)

The number of students from School A who score a perfect 100 follows a Poisson distribution with an average of 3. What is the probability that exactly 2 students get a perfect score? **(Given:** $e^{-3} \approx 0.05$**)**

### Solution

### (a) Overall Probability of Passing (Total Probability Theorem)

$$P(\text{Pass}) = P(P|A)P(A) + P(P|B)P(B) + P(P|C)P(C)$$

$$P(\text{Pass}) = (0.80 \times 0.50) + (0.60 \times 0.30) + (0.95 \times 0.20)$$

$$P(\text{Pass}) = 0.40 + 0.18 + 0.19 = \mathbf{0.77}$$

### (b) Probability from School B, given Pass (Bayes' Theorem)

$$P(B|\text{Pass}) = \frac{P(\text{Pass}|B)P(B)}{P(\text{Pass})}$$

$$P(B|\text{Pass}) = \frac{0.60 \times 0.30}{0.77} = \frac{0.18}{0.77} \approx \mathbf{0.2338}$$

## (c) Poisson Probability for Perfect Scores

The formula is $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. With $\lambda = 3$ and $k = 2$:

$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9 \times 0.05}{2} = \frac{0.45}{2} = \mathbf{0.225}$$

# Question 2

The time (in minutes) that a student takes to solve a puzzle is modeled by an Exponential Distribution with a rate parameter $\lambda = 0.2$. The PDF is $f(x) = 0.2e^{-0.2x}$ for $x \geq 0$.

**(Given:** $e^{-1} \approx 0.368$**,** $e^{-2} \approx 0.135$**, and** $\ln(2) \approx 0.693$**)**

## (a)

What is the probability that a student will take between 5 and 10 minutes to solve the puzzle?

## (b)

Find the CDF for this distribution. Use it to calculate the probability that a student will solve the puzzle in less than 5 minutes.

## (c)

Calculate the mean and median time to solve the puzzle.

### Solution

## (a) Probability between 5 and 10 minutes

$$P(5 \leq X \leq 10) = \int_5^{10} 0.2e^{-0.2x} dx = [-e^{-0.2x}]_5^{10}$$

$$= (-e^{-0.2\times10}) - (-e^{-0.2\times5}) = -e^{-2} + e^{-1}$$

$$= -0.135 + 0.368 = \mathbf{0.233}$$

## (b) CDF and Probability less than 5 minutes

The CDF is $F(x) = 1 - e^{-\lambda x} = \mathbf{1 - e^{-0.2x}}$.

$$P(X < 5) = F(5) = 1 - e^{-0.2\times5} = 1 - e^{-1}$$

$$= 1 - 0.368 = \mathbf{0.632}$$

## (c) Mean and Median Time

**Mean:** $E[X] = \frac{1}{\lambda} = \frac{1}{0.2} = \mathbf{5}$ minutes **Median:** $m = \frac{\ln(2)}{\lambda} = \frac{0.693}{0.2} = \mathbf{3.465}$ minutes

# Question 3

Suppose the time it takes a data collection operator to fill out an electronic form for a database is uniformly distributed between 1.5 and 2.2 minutes.

i. What is the mean and variance of the time it takes an operator to fill out the form?

ii. What is the probability that it will take less than two minutes to fill out the form?

iii. What is the probability that it will take between 1.8 and 2.1 minutes to fill out the form?

## Solution

Let the time $X \sim U(a, b)$, where $a = 1.5$, $b = 2.2$

### i. Mean and Variance

$$\text{Mean } \mu = \frac{a+b}{2} = \frac{1.5+2.2}{2} = \frac{3.7}{2} = 1.85 \text{ minutes}$$
$$\text{Variance } \sigma^2 = \frac{(b-a)^2}{12} = \frac{(2.2-1.5)^2}{12} = \frac{0.49}{12} \approx 0.0408 \text{ min}^2$$

### ii. Probability that time is less than 2 minutes

$$P(X < 2) = \frac{2-1.5}{2.2-1.5} = \frac{0.5}{0.7} \approx 0.7143$$

### iii. Probability that time is between 1.8 and 2.1 minutes

$$P(1.8 < X < 2.1) = \frac{2.1-1.8}{2.2-1.5} = \frac{0.3}{0.7} \approx 0.4286$$

## Final Answers

- Mean: $\boxed{1.85}$, Variance: $\boxed{0.0408}$
- $P(X < 2) \approx \boxed{0.7143}$
- $P(1.8 < X < 2.1) \approx \boxed{0.4286}$

# Question 4

A patient is suspected of having one of three mutually exclusive diseases: Disease A, Disease B, or Disease C. The probability that a patient has Disease A is 0.30, the probability of Disease B is 0.50, and the probability of Disease C is 0.20. A diagnostic test is administered to the patient. The test is known to return a positive result with probability 0.90 if the patient has Disease A, with probability 0.70 if the patient has Disease B, and with probability 0.40 if the patient has Disease C.

A randomly selected patient tests positive. What is the probability that the patient has disease A?

## Solution

We apply Bayes' Theorem:

$$P(A \mid \text{Pos}) = \frac{P(\text{Pos} \mid A) \cdot P(A)}{P(\text{Pos})}$$

First, compute the total probability of a positive result:

$$\begin{aligned} P(\text{Pos}) &= P(\text{Pos} \mid A)P(A) + P(\text{Pos} \mid B)P(B) + P(\text{Pos} \mid C)P(C) \\ &= (0.9)(0.3) + (0.7)(0.5) + (0.4)(0.2) = 0.27 + 0.35 + 0.08 = 0.70 \end{aligned}$$

Now, compute the posterior:

$$P(A \mid \text{Pos}) = \frac{0.9 \cdot 0.3}{0.70} = \frac{0.27}{0.70} \approx 0.3857$$

### Final Answer

$$\boxed{P(A \mid \text{Positive}) \approx 38.57\%}$$

# Question 5

A startup tracks the daily time (in hours) its users spend on their app. Analysis shows that the usage time follows an **exponential distribution** with rate parameter:

$$\lambda = 0.5 \quad \text{(i.e., average usage time = 2 hours)}$$

Let the random variable $X$ represent daily usage time per user.
Answer the following :

(a) What is the PDF of the distribution?

(b) What is the CDF of $X$?

(c) What is the probability that a user spends more than 3 hours on the app in a day?

(d) What is the PPF (inverse CDF) value at 90%? i.e., find the time $x$ such that 90% of users spend less than $x$ hours per day.

## Solution

(a) The PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x} = 0.5e^{-0.5x}, \quad x \geq 0$$

(b) The CDF is:

$$F(x) = 1 - e^{-\lambda x} = 1 - e^{-0.5x}, \quad x \geq 0$$

(c) The probability that a user spends more than 3 hours:

$$P(X > 3) = 1 - F(3) = e^{-0.5 \cdot 3} = e^{-1.5} \approx \boxed{0.2231}$$

(d) To find the PPF (inverse CDF) at 90%, set $F(x) = 0.90$:

$$0.90 = 1 - e^{-0.5x} \Rightarrow e^{-0.5x} = 0.10 \Rightarrow -0.5x = \ln(0.10)$$

$$x = \frac{-\ln(0.10)}{0.5} = \frac{2.3026}{0.5} = \boxed{4.6052 \text{ hours}}$$

## Question 6

In an e-commerce system, the total delivery time $T$ is the sum of two independent random variables:

- $X \sim U(1, 3)$: time to process an order (in hours)
- $Y \sim U(2, 4)$: time to deliver the order (in hours)

Let $T = X + Y$ be the total delivery time.

(a) What is the range of $T$?

(b) What is the shape of the PDF of $T$? Express the piecewise function.

(c) What is the probability that delivery is completed in less than 5 hours?

### Solution

(a) The range of $T$ is:

$$T_{\min} = 1 + 2 = 3, \quad T_{\max} = 3 + 4 = 7 \quad \Rightarrow \boxed{3 \le T \le 7}$$

(b) The PDF of $T = X + Y$ (convolution of two uniforms of width 2) is trapezoidal:

$$f_T(t) = \begin{cases} 0, & t < 3 \text{ or } t > 7 \\ \frac{t-3}{4}, & 3 \le t < 4 \\ \frac{1}{2}, & 4 \le t < 5 \\ \frac{7-t}{4}, & 5 \le t \le 7 \\ 0, & \text{otherwise} \end{cases}$$

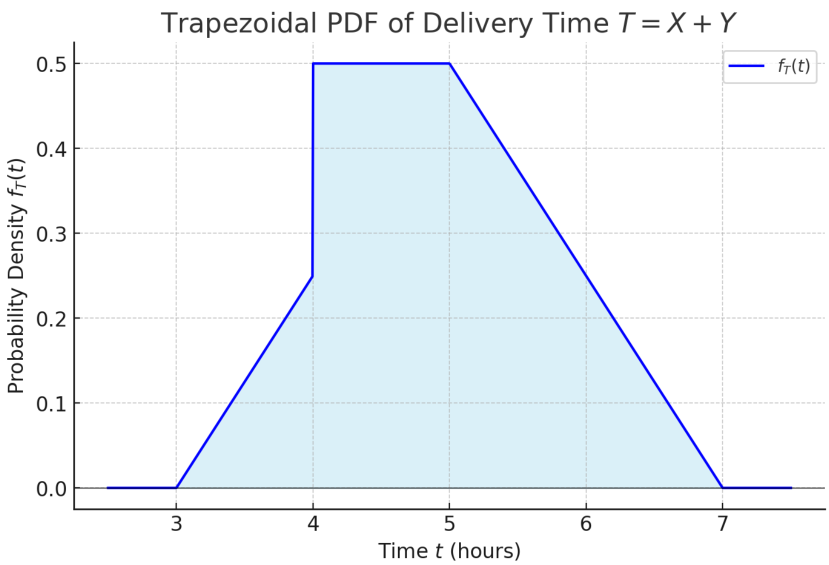


Figure 1: Trapezoidal PDF of Delivery Time $T = X + Y$

(c) Probability $P(T < 5)$ is the area under the PDF from 3 to 5:

$$\int_3^4 \frac{t-3}{4} dt = \left[\frac{(t-3)^2}{8}\right]_3^4 = \frac{1}{8}$$

$$\int_4^5 \frac{1}{2} dt = \frac{1}{2}$$

$$P(T < 5) = \frac{1}{8} + \frac{1}{2} = \boxed{0.625}$$

## Question 7 (Binomial Distribution)

It is known that 60% of mice inoculated with a serum are protected from a certain disease. If 5 mice are inoculated, find the probability that:

(a) None contracts the disease

(b) Fewer than 2 contract the disease

(c) More than 3 contract the disease

### Solution

Let $X$ be the number of mice that contract the disease. Given: Probability that a mouse is protected = 0.6, So, probability that it contracts the disease = $q = 1 - 0.6 = 0.4$ $X \sim \text{Binomial}(n = 5, p = 0.4)$

(a) Probability that none contracts the disease: $P(X = 0)$

$$P(X = 0) = \binom{5}{0}(0.4)^0(0.6)^5 = 1 \times 1 \times 0.07776 = \boxed{0.0778}$$

(b) Fewer than 2 contract the disease: $P(X < 2) = P(X = 0) + P(X = 1)$

$$P(X = 1) = \binom{5}{1}(0.4)^1(0.6)^4 = 5 \times 0.4 \times 0.1296 = 0.2592$$

$$P(X < 2) = 0.0778 + 0.2592 = \boxed{0.337}$$

(c) More than 3 contract the disease: $P(X > 3) = P(X = 4) + P(X = 5)$

$$P(X = 4) = \binom{5}{4}(0.4)^4(0.6)^1 = 5 \times 0.0256 \times 0.6 = 0.0768$$

$$P(X = 5) = \binom{5}{5}(0.4)^5(0.6)^0 = 1 \times 0.01024 \times 1 = 0.01024$$

$$P(X > 3) = 0.0768 + 0.01024 = \boxed{0.087}$$

## Question 8 (Poisson Distribution)

The defects on a plywood sheet occur at random with an average of one defect per 50 sq. ft. What is the probability that such a sheet will have:

(i) No defect

(ii) At least one defect

[Use $e^{-1} = 0.3678$]

### Solution

Let $X \sim \text{Poisson}(\lambda = 1)$ where $\lambda$ is the average number of defects per sheet.

(i) Probability of no defect: $P(X = 0)$

$$P(X = 0) = \frac{e^{-1} \cdot 1^0}{0!} = \frac{0.3678 \cdot 1}{1} = \boxed{0.3678}$$

(ii) Probability of at least one defect: $P(X \geq 1) = 1 - P(X = 0)$

$$P(X \geq 1) = 1 - 0.3678 = \boxed{0.6322}$$

## Question 9 (Standard Normal Distribution)

A factory produces metal rods whose lengths follow a normal distribution with a mean ($\mu$) of 100 cm and a standard deviation ($\sigma$) of 5 cm. Let $Z$ be the standard normal variable.

a) Convert the following lengths to corresponding $Z$-scores: 90 cm, 100 cm, and 110 cm.

**Solution:**

$$Z = \frac{X - \mu}{\sigma}$$

- For 90 cm: $Z = \frac{90-100}{5} = -2$
- For 100 cm: $Z = \frac{100-100}{5} = 0$
- For 110 cm: $Z = \frac{110-100}{5} = 2$

b) Using the Z-table, find the probability that a rod is:

i. Less than 90 cm: $P(Z < -2) \approx 0.0228$

ii. Between 95 cm and 105 cm:

$$Z_{95} = \frac{95 - 100}{5} = -1, \quad Z_{105} = \frac{105 - 100}{5} = 1$$

$$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) \approx 0.8413 - 0.1587 = 0.6826$$

iii. More than 110 cm: $P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$

c) **Interpretation:** There is a 2.28% chance that a rod is shorter than 90 cm or longer than 110 cm, and around 68.26% of rods lie between 95 cm and 105 cm, reflecting the properties of normal distribution.

## Question 10 (Probability Density Function of Standard Normal)

The PDF of the standard normal distribution is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

a) **Proof that Area under PDF = 1:**

The integral of the standard normal PDF over all real numbers is:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 1$$

This is a well-known result derived using polar coordinates in multivariable calculus.

b) **Compute PDF values:**

$$f(0) = \frac{1}{\sqrt{2\pi}} \approx 0.3989$$

$$f(1) = \frac{1}{\sqrt{2\pi}} e^{-1/2} \approx 0.2419$$

$$f(-2) = \frac{1}{\sqrt{2\pi}} e^{-2} \approx 0.05399$$

c) **Sketch of the PDF Curve:**

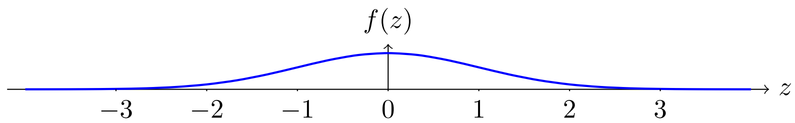


Figure 2: Standard Normal Distribution PDF

## Question 11

A certain firm has plant A, B and C producing IC chips. Plant A produces twice the output from B and B produces twice the output from C. The probability of a non-defective product produced by A, B, C are respectively 0.85, 0.75 and 0.95. A customer receives a defective product. Find the probability that it came from plant B.

**Solution:**

Given: Plant A produces twice the output of B. Plant B produces twice the output of C.

Let Plant A produces 100 number of IC chips. Let Plant B produces 50 number of IC chips. Let Plant C produces 25 number of IC chips.

$E \rightarrow$ The event that the item produced is non-defective.

$$P(A) = \frac{100}{175} = 0.57$$

$$P(B) = \frac{50}{175} = 0.29$$

$$P(C) = \frac{25}{175} = 0.14$$

The probability of a non-defective product from plants A, B, and C are:

$$P(E|A) = 0.85, \quad P(E|B) = 0.75, \quad P(E|C) = 0.95$$

Hence, the probability of defective products:

$P(\bar{E}|A) = 1 - P(E|A) = 1 - 0.85 = 0.15$

$P(\bar{E}|B) = 1 - P(E|B) = 1 - 0.75 = 0.25$

$P(\bar{E}|C) = 1 - P(E|C) = 1 - 0.95 = 0.05$

We are required to find the probability that a defective product is from plant B:

$$P(B|\bar{E}) = \frac{P(B) \cdot P(\bar{E}|B)}{P(A) \cdot P(\bar{E}|A) + P(B) \cdot P(\bar{E}|B) + P(C) \cdot P(\bar{E}|C)}$$

Substituting the values:

$$P(B|\bar{E}) = \frac{(0.29)(0.25)}{(0.57)(0.15) + (0.29)(0.25) + (0.14)(0.05)} = \frac{0.0725}{0.0855 + 0.0725 + 0.007}$$

$$P(B|\bar{E}) = \frac{0.0725}{0.165} = 0.4394$$

**Final Answer:** $\boxed{P(B|\bar{E}) = 0.4394}$

# Question 12

**i)**

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

**Solution:**

Given the cumulative distribution function (CDF) of $X$:

**To find:** The probability density function (PDF)

$$f_X(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(1 - e^{-\lambda x}\right) = 0 - (-\lambda e^{-\lambda x}) = \lambda e^{-\lambda x}$$

So the PDF is:

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

**To find:** $P(1 \le X \le 2)$

$$P(1 \le X \le 2) = \int_1^2 f(x)\,dx = \int_1^2 \lambda e^{-\lambda x}\,dx$$

Evaluating the integral:

$$\int_1^2 \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_1^2 = -e^{-2\lambda} + e^{-\lambda}$$

$$\Rightarrow P(1 \le X \le 2) = e^{-\lambda} - e^{-2\lambda}$$

**ii)** A continuous random variable $X$, which can assume any value between $X = 2$ and $X = 5$, has a probability density function given by

$$f(x) = k(1+x).$$

Find the probability $P(X < 4)$.

**(i) Formula:**

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Here, since $f(x) = k(1+x)$ is defined over the interval $[2,5]$:

$$\int_{2}^{5} k(1+x)\,dx = 1$$

$$k\int_{2}^{5} (1+x)\,dx = 1$$

$$k\left[x + \frac{x^2}{2}\right]_2^5 = 1$$

$$k\left[\left(5 + \frac{25}{2}\right) - \left(2 + \frac{4}{2}\right)\right] = 1$$

$$k\left[\frac{35}{2} - \frac{8}{2}\right] = 1$$

$$k \cdot \frac{27}{2} = 1 \quad \Rightarrow \quad k = \frac{2}{27}$$

**(ii) Find** $P(X < 4)$

$$P(X<4) = \int_2^4 f(x)\,dx = \int_2^4 k(1+x)\,dx = \int_2^4 \frac{2}{27}(1+x)\,dx$$

$$= \frac{2}{27}\int_2^4 (1+x)\,dx = \frac{2}{27}\left[x + \frac{x^2}{2}\right]_2^4$$

$$= \frac{2}{27}\left[(4+8) - (2+2)\right] = \frac{2}{27}(12-4) = \frac{2}{27}\cdot 8 = \frac{16}{27}$$

$$\boxed{P(X<4) = \frac{16}{27}}$$

## Question 13 (Exponential Distribution)

The time between arrivals of customers at a bank follows an exponential distribution with a mean of 10 minutes. Find the probability that:

(a) A customer arrives in less than 5 minutes

(b) A customer arrives after more than 15 minutes

**Use the exponential formula:**

$$P(T \leq t) = 1 - e^{-\lambda t}, \quad \text{where } \lambda = \frac{1}{\text{mean}} = \frac{1}{10} = 0.1$$

### Solution

Let $T \sim \text{Exponential}(\lambda = 0.1)$

### (a) Probability that a customer arrives in less than 5 minutes: $P(T < 5)$

$$P(T < 5) = 1 - e^{-0.1\times 5} = 1 - e^{-0.5} = 1 - 0.6065 = \boxed{0.3935}$$

### (b) Probability that a customer arrives after more than 15 minutes: $P(T > 15)$

$$P(T > 15) = e^{-0.1\times 15} = e^{-1.5} = \boxed{0.2231}$$

## Question 14 (Uniform Distribution)

A machine produces metal rods that have lengths uniformly distributed between 40 cm and 60 cm. Find the probability that a randomly selected rod is:

(i) Shorter than 45 cm

(ii) Between 42 cm and 50 cm

**Uniform PDF:**

$$f(x) = \frac{1}{b-a}, \quad \text{where } a = 40,\, b = 60$$

### Solution

Let $X \sim U(40, 60)$

### (i) Probability that a rod is shorter than 45 cm: $P(X < 45)$

$$P(X < 45) = \frac{45-40}{60-40} = \frac{5}{20} = \boxed{0.25}$$

### (ii) Probability that a rod is between 42 cm and 50 cm: $P(42 < X < 50)$

$$P(42 < X < 50) = \frac{50-42}{60-40} = \frac{8}{20} = \boxed{0.4}$$

# Question 15:

**Problem Statement:**

The time (in hours) between consecutive breakdowns of a particular machine follows an **Exponential distribution** with an average time of 5 hours between breakdowns.

Let the random variable $T \sim \text{Exponential}(\lambda)$, where $\lambda$ is the rate parameter.

Answer the following:

(a) What is the probability that the machine will run for more than 8 hours without a breakdown?

(b) What is the probability that the machine will break down within the first 3 hours?

(c) What is the expected time and standard deviation between breakdowns?

## Solution:

**Exponential Distribution Definition:**

The probability density function (PDF) of the exponential distribution is given by:

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

where $\lambda$ is the rate parameter and $\frac{1}{\lambda}$ is the mean time between events[1].

Given: Mean = 5 hours.
Hence, the rate parameter is:

$$\lambda = \frac{1}{\text{Mean}} = \frac{1}{5} = 0.2$$

### (a) Probability that the machine runs for more than 8 hours

This is calculated using the **survival function**:

$$P(T > t) = e^{-\lambda t}$$

Substitute $t = 8$, $\lambda = 0.2$:

$$P(T > 8) = e^{-0.2 \cdot 8} = e^{-1.6} \approx 0.2019$$

**Answer:** $P(T > 8) \approx \boxed{0.2019}$

### (b) Probability that the machine breaks down within the first 3 hours

This is the complement of the survival function:

$$P(T < 3) = 1 - P(T > 3) = 1 - e^{-\lambda \cdot 3}$$

Substitute $\lambda = 0.2$:

$$P(T < 3) = 1 - e^{-0.6} \approx 1 - 0.5488 = 0.4512$$

**Answer:** $P(T < 3) \approx \boxed{0.4512}$

### (c) Expected value and standard deviation

For the exponential distribution:

- Mean (Expected value): $\mathbb{E}[T] = \frac{1}{\lambda} = 5$
- Standard deviation: $\sigma = \frac{1}{\lambda} = 5$[2]

**Answer:** Expected time = $\boxed{\text{5 hours}}$, Standard deviation = $\boxed{\text{5 hours}}$

[1]The exponential distribution is widely used to model waiting times or the time between events in a Poisson process.
[2]For the exponential distribution, both the mean and standard deviation are equal to $\frac{1}{\lambda}$.

# Question 16:

**Problem Statement:**
Let the random variable $X$ represent the time (in minutes) a commuter waits for a bus that arrives at a random time uniformly between 0 and 20 minutes. That is, $X \sim \text{Uniform}(0, 20)$.

(a) Find the probability density function (PDF) of $X$.

(b) Calculate the mean and variance of $X$.

(c) What is the probability that the commuter waits less than 8 minutes?

(d) What is the probability that the commuter waits more than 15 minutes?

(e) Find the time by which 75% of the commuters would have caught the bus.

**Solution:**
Let $X \sim \text{Uniform}(a = 0, b = 20)$. This means that the commuter is equally likely to wait any amount of time between 0 and 20 minutes. The uniform distribution is defined over a continuous interval where all outcomes are equally likely.

(a) **PDF of Uniform Distribution**:

For a continuous uniform distribution over the interval $[a, b]$, the PDF is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Substituting $a = 0$, $b = 20$:

$$f(x) = \begin{cases} \frac{1}{20} & \text{for } 0 \leq x \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

(b) **Mean and Variance**:

For a uniform distribution $X \sim \text{Uniform}(a, b)$, the formulas for mean and variance are:

$$\text{Mean: } \mu = \frac{a+b}{2}$$
$$\text{Variance: } \sigma^2 = \frac{(b-a)^2}{12}$$

Substituting $a = 0$, $b = 20$:

$$\mu = \frac{0+20}{2} = 10, \quad \sigma^2 = \frac{(20-0)^2}{12} = \frac{400}{12} = 33.33$$

(c) **Probability that the commuter waits less than 8 minutes:**

Since the PDF is constant on $[0, 20]$, we compute the probability as the proportion of the interval:

$$P(X < 8) = \int_0^8 f(x)\, dx = \int_0^8 \frac{1}{20}\, dx = \frac{8}{20} = 0.4$$

(d) **Probability that the commuter waits more than 15 minutes:**

$$P(X > 15) = \int_{15}^{20} \frac{1}{20}\, dx = \frac{5}{20} = 0.25$$

(e) **Time by which 75% of commuters catch the bus:**

For a uniform distribution, the cumulative distribution function (CDF) is:

$$F(x) = \frac{x-a}{b-a} \quad \text{for } a \leq x \leq b^3$$

We solve for $x$ such that $F(x) = 0.75$:

$$\frac{x-0}{20} = 0.75 \Rightarrow x = 15$$

So, 75% of commuters catch the bus within 15 minutes.

**Final Answers:**

- (a) $f(x) = \frac{1}{20}$ for $0 \leq x \leq 20$
- (b) Mean = 10, Variance = 33.33
- (c) $P(X < 8) = 0.4$
- (d) $P(X > 15) = 0.25$
- (e) 75% percentile = 15 minutes

[3] This comes from integrating the constant PDF: $F(x) = \int_a^x f(t)\,dt = \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a}$

## Question 17: Conditional Probability (Bayes' Theorem)

A company has two factories:

- Factory A produces 60% of the total goods, and 5% of its products are defective.
- Factory B produces 40% of the total goods, and 10% of its products are defective.

A product is randomly selected and found to be defective. What is the probability that it came from Factory B?

**Solution:**
Let:

- $A$ = Product is from Factory A
- $B$ = Product is from Factory B
- $D$ = Product is defective

We are asked to find $P(B|D)$, using Bayes' Theorem:

$$P(B|D) = \frac{P(D|B) \cdot P(B)}{P(D)}$$

Given:

$$P(A) = 0.6, \quad P(B) = 0.4, \quad P(D|A) = 0.05, \quad P(D|B) = 0.10$$

Using the law of total probability:

$$P(D) = P(D|A) \cdot P(A) + P(D|B) \cdot P(B) = 0.05 \cdot 0.6 + 0.10 \cdot 0.4 = 0.03 + 0.04 = 0.07$$

Now:

$$P(B|D) = \frac{0.10 \cdot 0.4}{0.07} = \frac{0.04}{0.07} \approx 0.571$$

**Answer:** The probability that the defective product came from Factory B is approximately 0.571.

# Question 18: Independence of Events

In a tech support center:

- $P(A) = 0.4$ represents the probability that a customer reports a login issue.
- $P(B) = 0.3$ represents the probability that a customer reports a payment issue.
- $P(A \cup B) = 0.58$ represents the probability that a customer reports at least one of these issues.

Are the events $A$ and $B$ independent?

**Solution:**
We use the identity:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Solving for $P(A \cap B)$:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.58 = 0.12$$

Now check for independence:

$$P(A) \cdot P(B) = 0.4 \cdot 0.3 = 0.12$$

Since:

$$P(A \cap B) = P(A) \cdot P(B)$$

**Answer:** Events $A$ and $B$ are independent.

# Question 19

The time (in minutes) required to download a large dataset from a cloud server is a continuous random variable $T$. The CDF of $T$ is given by:

$$F(t) = 1 - e^{-0.4t}, \quad t \geq 0$$

where $F(t)$ gives the probability that the download finishes within $t$ minutes.

## Questions

1. What is the probability that the download finishes within 5 minutes?
2. What is the probability that it takes more than 8 minutes?
3. What is the probability that the download takes between 2 and 6 minutes?
4. Find the PDF of $T$.
5. Find the expected download time $E[T]$.

## Solution

**Step 1: Probability it finishes within 5 minutes**

$$P(T < 5) = F(5) = 1 - e^{-0.4(5)} = 1 - e^{-2} \approx 1 - 0.1353 = 0.8647$$

$\boxed{P(T < 5) = 86.5\%}$

**Step 2: Probability it takes more than 8 minutes**

$$P(T > 8) = 1 - F(8) = e^{-0.4(8)} = e^{-3.2} \approx 0.0408$$

$\boxed{P(T > 8) = 4.1\%}$

**Step 3: Probability it takes between 2 and 6 minutes**

$$P(2 < T < 6) = F(6) - F(2) = (1 - e^{-2.4}) - (1 - e^{-0.8}) = 0.9093 - 0.5507 = 0.3586$$

$\boxed{P(2 < T < 6) = 35.9\%}$

**Step 4: Finding the PDF**

$$f(t) = \frac{d}{dt}F(t) = 0.4e^{-0.4t}, \quad t \geq 0$$

$\boxed{f(t) = 0.4e^{-0.4t}}$

**Step 5: Expected Download Time**

$$E[T] = \frac{1}{\lambda} = \frac{1}{0.4} = 2.5 \text{ minutes}$$

$\boxed{E[T] = 2.5 \text{ minutes}}$

# Question 20

A cybersecurity analyst knows that, on any given day, there is a 2% prior probability ($P(H) = 0.02$) that a system is under a hacker attack. The company uses an AI threat detection tool:

- If the system is under attack, the tool raises an alert with probability $P(A|H) = 0.9$.
- If there is no attack, the tool raises a false alert with probability $P(A|NH) = 0.05$.

On a certain day, the tool raises an alert.

## Questions

1. What is the a priori probability that the system is under attack?
2. What is the probability $P(A)$ that the tool raises an alert?
3. What is the a posteriori probability that there is an actual attack, given an alert is raised?
4. If two independent alerts are raised consecutively, what is the updated posterior probability?

## Solution

**Step 1: A Priori Probability**

$$P(H) = 0.02$$

$\boxed{P(H) = 2\%}$

**Step 2: Total Probability of an Alert $P(A)$**

$$P(A) = P(A|H)P(H) + P(A|NH)P(NH) = (0.9)(0.02) + (0.05)(0.98) = 0.018 + 0.049 = 0.067$$

$\boxed{P(A) = 0.067 \quad (6.7\%)}$

**Step 3: A Posteriori Probability $P(H|A)$**

$$P(H|A) = \frac{P(A|H)P(H)}{P(A)} = \frac{0.9(0.02)}{0.067} = 0.2687$$

$\boxed{P(H|A) \approx 26.9\%}$

**Step 4: Two Consecutive Independent Alerts**

$$P(A_1, A_2|H) = (0.9)^2 = 0.81, \qquad P(A_1, A_2|NH) = (0.05)^2 = 0.0025$$

$$P(A_1, A_2) = P(H)(0.81) + P(NH)(0.0025) = 0.0162 + 0.00245 = 0.01865$$

$$P(H|A_1, A_2) = \frac{0.0162}{0.01865} = 0.868$$

$\boxed{P(H|A_1, A_2) \approx 86.8\%}$

## Question 21: PDF $\to$ CDF & PPF

Let $X$ be a continuous random variable with probability density function (PDF):

$$f(x) = \begin{cases} 2x & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

### (a) Find the CDF $F(x)$ of $X$.

To find the CDF:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Case-wise definition:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^2 & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

### (b) Find the PPF for $p = 0.75$.

The PPF is the inverse of the CDF. Solve:

$$F(x) = 0.75 \Rightarrow x^2 = 0.75 \Rightarrow x = \sqrt{0.75} = \frac{\sqrt{3}}{2} \approx 0.866$$

**Answer:** The PPF for $p = 0.75$ is $\boxed{\frac{\sqrt{3}}{2} \approx 0.866}$.

## Question 22: CDF $\to$ PDF & PPF

Let the cumulative distribution function (CDF) of a random variable $X$ be:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^2 & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

### (a) Find the PDF $f(x)$.

To find the PDF, differentiate the CDF:

$$f(x) = \frac{d}{dx}F(x)$$

$$f(x) = \begin{cases} 2x & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

### (b) Find the PPF for $p = 0.64$.

Solve:

$$F(x) = 0.64 \Rightarrow x^2 = 0.64 \Rightarrow x = \sqrt{0.64} = 0.8$$

**Answer:** The PPF for $p = 0.64$ is $\boxed{0.8}$.

## Question 23

A diagnostic test for a rare disease has the following characteristics:

- If a person has the disease, the test returns positive 99% of the time.
- If a person does not have the disease, the test returns positive 5% of the time.
- The disease is present in 0.1% of the population.

What is the probability that a randomly selected person who tests positive actually has the disease?

### Solution

Let:

- $D$: person has the disease
- $\neg D$: person does not have the disease
- $T^+$: test is positive

We want to find $P(D \mid T^+)$. Using Bayes' Theorem:

$$P(D \mid T^+) = \frac{P(T^+ \mid D) \cdot P(D)}{P(T^+ \mid D) \cdot P(D) + P(T^+ \mid \neg D) \cdot P(\neg D)}$$

Substitute the values:

$$P(T^+ \mid D) = 0.99, \quad P(T^+ \mid \neg D) = 0.05, \quad P(D) = 0.001, \quad P(\neg D) = 0.999$$

$$P(D \mid T^+) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.05 \cdot 0.999} = \frac{0.00099}{0.00099 + 0.04995}$$

$$P(D \mid T^+) = \frac{0.00099}{0.05094} \approx 0.0194$$

**Answer:** Approximately 1.94% chance that the person actually has the disease.

## Question 24

Consider a data stream logging the number of requests to a server every minute.

(a) If the number of requests per minute follows a Poisson distribution with a mean of 10, what is the probability that there are exactly 15 requests in a minute?

(b) The server response time (in seconds) follows a normal distribution with mean 1.5 and standard deviation 0.3. Calculate the 95th percentile response time.

(c) If a random variable is uniformly distributed between 2 and 8, what is the probability that it falls between 3 and 6?

### Solution

**(a)** Using the Poisson formula:

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

Given $\lambda = 10$, $k = 15$:

$$P(X = 15) = \frac{e^{-10} \cdot 10^{15}}{15!} \approx 0.0347$$

**Answer:** Approximately 3.47%

**(b)** For a normal distribution with mean $\mu = 1.5$ and standard deviation $\sigma = 0.3$, the 95th percentile is:

$$x = \mu + z \cdot \sigma = 1.5 + 1.645 \cdot 0.3 = 1.9935$$

**Answer:** Approximately 1.99 seconds

**(c)** For a uniform distribution from 2 to 8, the probability that $X$ falls between 3 and 6 is:

$$P(3 < X < 6) = \frac{6-3}{8-2} = \frac{3}{6} = 0.5$$

**Answer:** 50%

## Question 25

In a book of 520 pages, there are 390 typographical errors. Assuming Poisson's law for the number of errors per page, find the probability that a random sample of 5 pages will contain no error.

### Solution

- Total pages = 520
- Total errors = 390
- Mean number of errors per page:
$$\lambda = \frac{390}{520} = 0.75$$
- For 5 pages, expected number of errors:
$$\lambda_5 = 5 \times 0.75 = 3.75$$
- We want the probability of no errors in 5 pages:
$$P(X = 0) = \frac{e^{-3.75} \cdot 3.75^0}{0!} = e^{-3.75}$$
- Numerical calculation:
$$P(X = 0) \approx e^{-3.75} \approx 0.0235$$

**Final Answer:** The probability of no errors in 5 pages is approximately $\boxed{0.0235}$.

## Question 26

A call center receives an average of 2 calls every 10 minutes. Assuming the number of calls follows a Poisson distribution, what is the probability that the center receives no calls in a 10-minute interval?

### Solution

- Mean number of calls in 10 minutes:
$$\lambda = 2$$
- We want the probability of receiving 0 calls:
$$P(X = 0) = \frac{e^{-2} \cdot 2^0}{0!} = e^{-2}$$
- Numerical calculation:
$$P(X = 0) \approx e^{-2} \approx 0.1353$$

**Final Answer:** The probability of receiving no calls is approximately $\boxed{0.1353}$.

# Question 27 (Normal Distribution)

The discharge of suspended solids from a phosphate mine is normally distributed, with a mean daily discharge of 27 milligrams per liter (mg/l) and a standard deviation of 14 mg/l. What proportion of days will the daily discharge exceed 50 mg/l?

## Given

- Mean ($\mu$) = 27 mg/l
- Standard deviation ($\sigma$) = 14 mg/l
- Find $P(X > 50)$

### Step 1: Convert to Z-score

$$Z = \frac{X - \mu}{\sigma} = \frac{50 - 27}{14} = \frac{23}{14} \approx 1.643$$

### Step 2: Use the Z-table or Standard Normal CDF

$$P(X > 50) = P(Z > 1.643)$$

$$P(Z < 1.643) \approx 0.9495$$

$$P(Z > 1.643) = 1 - 0.9495 = 0.0505$$

**Answer:** The proportion of days the daily discharge exceeds 50 mg/l is approximately:

$$P(X > 50) = 0.0505 \quad \text{or} \quad 5.05\%$$

# Question 28 (Binomial Distribution)

Suppose the four engines of a commercial aircraft are arranged to operate independently and that the probability of in-flight failure of a single engine is 0.01. What is the probability of the following events on a given flight?

1. No failures are observed.
2. No more than one failure is observed.

## Given

- Number of engines ($n$) = 4
- Probability of failure ($p$) = 0.01
- Probability of success ($q$) = 1 - 0.01 = 0.99
- $X \sim \text{Binomial}(n = 4, p = 0.01)$

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

### Part (a): No Failures ($P(X = 0)$)

$$P(X = 0) = \binom{4}{0} (0.01)^0 (0.99)^4 = 1 \cdot 1 \cdot 0.9606 = 0.9606$$

## Part (b): No More Than One Failure ($P(X \leq 1)$)

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

We already have:

$$P(X = 0) = 0.9606$$

Now compute:

$$P(X = 1) = \binom{4}{1}(0.01)^1(0.99)^3 = 4 \cdot 0.01 \cdot 0.9703 = 0.0388$$

$$P(X \leq 1) = 0.9606 + 0.0388 = 0.9994$$

## Final Answers

- (a) Probability of no failures: **0.9606**
- (b) Probability of no more than one failure: **0.9994**

# Question 29 : Batch Normalization with ReLU and Tanh

A neural network layer receives 4 inputs from a mini-batch:

$$x_1 = 3.0, \quad x_2 = -2.0, \quad x_3 = 0.0, \quad x_4 = 5.0$$

The inputs undergo Batch Normalization, followed by ReLU and Tanh activation functions. The mean and standard deviation of the batch are given as:

$$\mu = 1.5, \quad \sigma = 2.0$$

(a) Perform Batch Normalization and compute the ReLU-activated outputs.

(b) Apply Tanh activation on the normalized inputs and provide approximate output values.

(c) Which activation (ReLU or Tanh) retains more information from negative inputs? Explain.

## Solution

**(a) Batch Normalization and ReLU Activation**

Batch Normalization formula:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}$$

ReLU Activation function:

$$f(x) = \max(0, x)$$

| **Input** $x$ | **Normalized** $x_{\text{norm}}$ | **ReLU Output** |
|---|---|---|
| 3.0 | $\frac{3.0-1.5}{2} = 0.75$ | 0.75 |
| -2.0 | $\frac{-2.0-1.5}{2} = -1.75$ | 0 |
| 0.0 | $\frac{0.0-1.5}{2} = -0.75$ | 0 |
| 5.0 | $\frac{5.0-1.5}{2} = 1.75$ | 1.75 |

**ReLU Outputs:** [0.75, 0, 0, 1.75]

**(b) Tanh Activation**

Tanh activation function:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Using approximate values:

$$\tanh(0.75) \approx 0.635, \quad \tanh(-1.75) \approx -0.941, \quad \tanh(-0.75) \approx -0.635, \quad \tanh(1.75) \approx 0.941$$

| **Normalized** $x_{\text{norm}}$ | **Tanh Output** |
|---|---|
| 0.75 | 0.635 |
| -1.75 | -0.941 |
| -0.75 | -0.635 |
| 1.75 | 0.941 |

**Tanh Outputs:** [0.635, $-0.941$, $-0.635$, 0.941]

**(c) Information Retention: ReLU vs Tanh**

- **ReLU:** Discards all negative values (outputs 0), which leads to information loss from $x_2$ and $x_3$.
- **Tanh:** Retains both sign and magnitude by mapping inputs into the range $(-1, 1)$.

**Conclusion:** Tanh retains more information because it preserves negative input values, unlike ReLU which sparsifies them by outputting 0.

## Question 30 : Probability Distribution – Binomial & Normal Approximation

A data center monitors the uptime status of 100 servers every day. Each server independently has a 97% chance of staying up for the full day.

Let $X$ be the number of servers that stay up.

(a) What is the expected number of servers that stay up and the standard deviation?

(b) Use the normal approximation to the binomial to estimate the probability that at least 95 servers stay up.

(Given: $\Phi(1.64) = 0.9495$, $\Phi(1.28) = 0.8997$, $\Phi(1.96) = 0.9750$)

### Solution

Let $X \sim \text{Binomial}(n = 100, p = 0.97)$

**(a) Expected Value and Standard Deviation**

- Expected value (mean):
$$\mathbb{E}[X] = n \cdot p = 100 \cdot 0.97 = 97$$
- Standard deviation:
$$\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{100 \cdot 0.97 \cdot 0.03} = \sqrt{2.91} \approx 1.71$$

**(b) Normal Approximation to the Binomial**

Using the normal approximation:
$$X \sim N(\mu = 97, \sigma = 1.71)$$

We want:
$$P(X \geq 95) = P\left(X \geq 94.5\right) \quad \text{(using continuity correction)}$$

Convert to standard normal variable:
$$Z = \frac{94.5 - 97}{1.71} = \frac{-2.5}{1.71} \approx -1.46$$

So,
$$P(X \geq 95) = P(Z \geq -1.46) = 1 - P(Z < -1.46) = 1 - (1 - \Phi(1.46)) = \Phi(1.46)$$

Using the closest available value (approx):
$$\Phi(1.46) \approx 0.9279$$

**Final Answer:**

- Expected number of servers up: **97**
- Standard deviation: **1.71**
- $P(\text{at least 95 servers up}) \approx \boxed{0.9279}$

## Question 31:

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time.

(a) Calculate the probability of rain on any given day and define the relevant events.

(b) Using Bayes' Theorem, calculate the probability that it will actually rain given that the forecast predicts rain.

(c) Interpret the result: Should Marie move the wedding indoors?

**Solution:**
**(a) Define Events and Calculate Prior Probabilities**

- Let $R$: It rains on a given day.
- Let $\neg R$: It does not rain.
- Let $F$: The forecast predicts rain.

Given:

$$P(R) = \frac{5}{365} \approx 0.0137, \quad P(\neg R) = 1 - 0.0137 = 0.9863$$
$$P(F \mid R) = 0.9, \quad P(F \mid \neg R) = 0.1$$

**(b) Apply Bayes' Theorem**
We want to calculate:

$$P(R \mid F) = \frac{P(F \mid R) \cdot P(R)}{P(F)}$$

First, compute $P(F)$:

$$P(F) = P(F \mid R) \cdot P(R) + P(F \mid \neg R) \cdot P(\neg R)$$

$$P(F) = (0.9)(0.0137) + (0.1)(0.9863) = 0.01233 + 0.09863 = 0.11096$$

Now apply Bayes' Theorem:

$$P(R \mid F) = \frac{0.01233}{0.11096} \approx 0.1111$$

So, the probability that it will actually rain given the forecast is about $\boxed{11.1\%}$.

**(c) Interpretation**
Since the probability that it will actually rain is near 11.1% Marie need not move her ceremony indoors as there is likely an 89% chance that it will not rain.

## Question 32:

A bus arrives at a stop every 30 minutes, and a passenger arrives at a random time. Let $X$ be the number of minutes the passenger has to wait for the next bus, assuming $X$ follows a continuous uniform distribution over the interval $[0, 30]$.

(a) Define the probability density function (PDF) of $X$.

(b) Calculate:

- The probability that the passenger waits less than 10 minutes.
- The probability that the passenger waits between 5 and 15 minutes.

(c) Find the expected (mean) waiting time.

(d) If the passenger is told the bus just left, what is the probability they wait more than 25 minutes?

**Solution:**
**(a) Probability Density Function (PDF)**
Since $X \sim \text{Uniform}(0, 30)$, the PDF is:

$$f(x) = \begin{cases} \frac{1}{30}, & 0 \leq x \leq 30 \\ 0, & \text{otherwise} \end{cases}$$

**(b) Probability Calculations**
**(i)** Probability that the passenger waits less than 10 minutes:

$$P(X < 10) = \frac{10 - 0}{30 - 0} = \frac{10}{30} = \boxed{\frac{1}{3}}$$

**(ii)** Probability that the passenger waits between 5 and 15 minutes:

$$P(5 \leq X \leq 15) = \frac{15 - 5}{30} = \frac{10}{30} = \boxed{\frac{1}{3}}$$

**(c) Expected (Mean) Waiting Time**
For a uniform distribution over $[a, b] = [0, 30]$, the mean is:

$$E(X) = \frac{a + b}{2} = \frac{0 + 30}{2} = \boxed{15 \text{ minutes}}$$

**(d) Probability of Waiting More Than 25 Minutes After the Bus Just Left**
Even if the passenger knows the bus just left, the waiting time is still uniformly distributed over $[0, 30]$.

$$P(X > 25) = \frac{30 - 25}{30} = \frac{5}{30} = \boxed{\frac{1}{6}}$$

## Question:33

A discrete random variable $X$ represents the number of goals scored by a football team in a match. The probability mass function (PMF) is given below:

| $X$ **(Goals)** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(X = x)$ | 0.2 | 0.3 | 0.4 | 0.1 |

Answer the following:

(a) Verify that this is a valid probability distribution.

(b) Find the expected value $E[X]$.

(c) Compute the variance $\mathrm{Var}(X)$.

(d) What does the expected value mean in terms of how many goals the team usually scores per match?

# Solution

## (a) Validity of PMF

To be a valid probability mass function, the following two conditions must be satisfied:

- All probabilities must lie between 0 and 1.
- The sum of all probabilities must equal 1.

$$P(X=0)+P(X=1)+P(X=2)+P(X=3)=0.2+0.3+0.4+0.1=1.0$$

Since both conditions are satisfied, the PMF is valid.

## (b) Expected Value $E[X]$

The expected value of a discrete random variable is given by:

$$E[X]=\sum x\cdot P(X=x)$$

Substituting the values:

$$E[X]=(0\cdot 0.2)+(1\cdot 0.3)+(2\cdot 0.4)+(3\cdot 0.1)$$
$$E[X]=0+0.3+0.8+0.3=1.4$$

Therefore, the expected value is $\boxed{1.4}$.

## (c) Variance $\mathrm{Var}(X)$

First, we calculate $E[X^2]$:

$$E[X^2]=(0^2\cdot 0.2)+(1^2\cdot 0.3)+(2^2\cdot 0.4)+(3^2\cdot 0.1)$$
$$E[X^2]=0+0.3+1.6+0.9=2.8$$

Now, use the formula for variance:

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2=2.8-(1.4)^2=2.8-1.96=0.84$$

So, the variance is $\boxed{0.84}$.

## (d) Interpretation of Expected Value

The expected value $E[X]=1.4$ indicates that, on average, the football team is likely to score 1.4 goals per match. While an individual match cannot have a fractional goal, this value helps to understand the team's scoring trend over many matches.

## Final Answers

(a) The given PMF is valid.

(b) Expected value $E[X] = 1.4$

(c) Variance $\text{Var}(X) = 0.84$

(d) On average, the team scores 1.4 goals per match.

# Question:34

A diagnostic test is used to detect a disease that affects 1 in 1000 people. The test has the following characteristics:

- If a person has the disease, the test is positive with probability 0.99 (true positive rate).
- If a person does not have the disease, the test is positive with probability 0.05 (false positive rate).

If a person is selected at random and tests positive, what is the probability that the person actually has the disease?

## Solution

Let:

- $D$: person has the disease
- $\bar{D}$: person does not have the disease
- $T$: test is positive

$$P(D) = \frac{1}{1000} = 0.001, \quad P(\bar{D}) = 1 - P(D) = 0.999$$
$$P(T|D) = 0.99, \quad P(T|\bar{D}) = 0.05$$

## Using Bayes' Theorem:

$$P(D|T) = \frac{P(T|D) \cdot P(D)}{P(T)}$$

Using the law of total probability:

$$P(T) = P(T|D) \cdot P(D) + P(T|\bar{D}) \cdot P(\bar{D})$$
$$P(T) = (0.99)(0.001) + (0.05)(0.999) = 0.00099 + 0.04995 = 0.05094$$

Now calculate:

$$P(D|T) = \frac{(0.99)(0.001)}{0.05094} = \frac{0.00099}{0.05094} \approx 0.0194$$

## Final Answer:

$$P(D|T) \approx \boxed{0.0194} \quad \text{or } 1.94\%$$

# Question:35

Patients are given appointments for meeting a doctor at a certain specialty clinic. In one of the suggested schemes for making appointments, 6 patients are called at 8:30 AM. This scheme helps in reducing the idle time of the doctor (i.e., after meeting a patient, the doctor does not have to wait for the next patient to arrive), but all patients except the first one have to wait for their turn.

Suppose the meeting time of each patient with the doctor is a uniform random variable ranging from 3 to 15 minutes, and assuming that meeting times are independent of each other, answer the following two questions:

(a) What is the waiting time distribution of the second patient that the doctor sees? Explain.

(b) What is approximately the waiting time distribution of the $6^{\text{th}}$ patient? Explain. You may assume that all patients arrive simultaneously at the scheduled time of 8:30 AM.

## Solution

**(a)** As the waiting time of the second patient is dependent on the meeting time of the first patient, which follows a uniform distribution over the interval [3, 15], the waiting time of the second patient also follows a uniform distribution:

$$W_2 \sim U(3, 15).$$

**(b)** The waiting time of the $6^{\text{th}}$ patient is the sum of the meeting times of the first five patients:

$$W_6 = M_1 + M_2 + M_3 + M_4 + M_5,$$

where each $M_i$ is an independent uniform random variable on [3,15]. By the Central Limit Theorem, the distribution of $W_6$ can be approximated by a normal distribution with:

$$\text{Mean} = 5 \times 9 = 45,$$

$$\text{Variance} = 5 \times 12 = 60,$$

so the waiting time of the $6^{\text{th}}$ patient approximately follows

$$W_6 \sim N(45, \sqrt{60}).$$

# Question 36

Buses arrive at a specified bus stop at 15 minutes intervals starting at 7 a.m. that is 7 a.m., 7.15 a.m., 7.30 a.m., etc. If a passenger arrives at the bus stop at a random time which is uniformly distributed between 7 and 7.30 a.m., find the probability that he waits

(a) less than 5 minutes

(b) at least 12 minutes for a bus.

## Solution

: Let $X$ denote the time that a passenger arrives between 7 and 7.30 a.m.

$$X \sim U(0, 30)$$

$$f(x) = \frac{1}{b-a} = \frac{1}{30-0} = \frac{1}{30}$$

**(a)** Passenger waits less than 5 minutes, i.e., he arrives between 7.10 - 7.15 or 7.25 - 7.30.

$$P(\text{Waiting time less than 5 minutes}) = P(10 \leq x \leq 15) + P(25 \leq x \leq 30)$$

$$= \int_{10}^{15} \frac{1}{30}\, dx + \int_{25}^{30} \frac{1}{30}\, dx$$

$$= \frac{1}{30}[x]_{10}^{15} + \frac{1}{30}[x]_{25}^{30} = \frac{1}{3}$$

**(b)** Passenger waits at least 12 minutes, i.e., he arrives between 7 - 7.03 or 7.15 - 7.18.

$$P(\text{Waiting time at least 12 minutes}) = P(0 \le x \le 3) + P(15 \le x \le 18)$$

$$= \int_0^3 \frac{1}{30}\,dx + \int_{15}^{18} \frac{1}{30}\,dx$$

$$= \frac{1}{30}[x]_0^3 + \frac{1}{30}[x]_{15}^{18} = \frac{1}{5}$$

# Question 37

The length of time a person speaks over the phone follows an exponential distribution with mean 6. What is the probability that the person will talk for

1. more than 8 minutes
2. between 4 and 8 minutes?

## Solution

: Given:

$$f(x) = \frac{1}{6}e^{-\frac{x}{6}}$$

**1) Probability that $X > 8$:**

$$P[X > 8] = \int_8^\infty f(x)\,dx = \int_8^\infty \frac{1}{6}e^{-\frac{x}{6}}\,dx = \frac{1}{6}\left[\frac{e^{-\frac{x}{6}}}{-\frac{1}{6}}\right]_8^\infty$$

$$= -\left[e^{-\frac{x}{6}}\right]_8^\infty = -\left[0 - e^{-\frac{8}{6}}\right] = e^{-\frac{8}{6}}$$

$$= 0.26359 \approx 0.2636$$

**2) Probability that $4 \le X \le 8$:**

$$P[4 \le X \le 8] = \int_4^8 \frac{1}{6}e^{-\frac{x}{6}}\,dx = \frac{1}{6}\left[\frac{e^{-\frac{x}{6}}}{-\frac{1}{6}}\right]_4^8$$

$$= -\left[e^{-\frac{x}{6}}\right]_4^8 = -\left[e^{-\frac{8}{6}} - e^{-\frac{4}{6}}\right] = e^{-\frac{4}{6}} - e^{-\frac{8}{6}}$$

$$= [0.51342 - 0.26360] = 0.24982$$

# Question 38

A box contains 2000 components of which 15% are defective. A second box contains 5000 components of which 25% are defective. Two other boxes contain two components each with 10% defective components. A box is chosen at random and an item selected was found to be defective. Find the probability that this has come from the second box.

## Solution:

Let $A_i$ be the event of selecting the $i^{th}$ box, $B$ be the event consisting of all defective components.

We have:

$$P(A_1) = P(A_2) = P(A_3) = P(A_4) = \frac{1}{4}$$

$$P(B|A_1) = 15\% = 0.15; \quad P(B|A_2) = \frac{25}{100} = 0.25$$

$$P(B|A_3) = \frac{10}{100} = 0.10; \quad P(B|A_4) = \frac{10}{100} = 0.10$$

Probability that the component is defective is given by:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + P(A_4)P(B|A_4)$$

$$= \frac{1}{4}(0.15) + \frac{1}{4}(0.25) + \frac{1}{4}(0.10) + \frac{1}{4}(0.10)$$

$$= 0.0375 + 0.0625 + 0.025 + 0.025 = 0.15$$

Hence the required probability

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{\sum_{i=1}^{4} P(B|A_i)P(A_i)}$$

$$= \frac{\left(\frac{1}{4}\right)(0.25)}{0.15} \quad \left[\because \sum_{i=1}^{4} P(B|A_i)P(A_i) = P(B)\right]$$

$$= 0.41667$$

# Question 39

The probability density function of a random variable $X$ is given by

$$f_X(x) = \begin{cases} x, & 0 < x < 1 \\ k(2-x), & 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find the value of $'k'$.

(ii) Find $P(0.2 < x < 1.2)$.

(iii) Find the distribution function of $f(x)$.

## Solution:

Since $f(x)$ is the pdf, we have

**(i) Find the value of** $k$:
Since $f(x)$ is a PDF,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{0}^{2} f(x)\,dx = 1$$

$$\int_{0}^{1} x\,dx + \int_{1}^{2} k(2-x)\,dx = 1$$

$$= \left[\frac{x^2}{2}\right]_0^1 + k\left[2x - \frac{x^2}{2}\right]_1^2$$

$$= \left(\frac{1}{2}\right) + k\left[(4-2) - \left(2 - \frac{1}{2}\right)\right]$$

$$= \frac{1}{2} + k\left(2 - \frac{3}{2}\right) = \frac{1}{2} + \frac{k}{2} = 1$$

$$\Rightarrow \frac{k}{2} = \frac{1}{2} \Rightarrow k = 1$$

**(ii) Find** $P(0.2 < x < 1.2)$

$$P(0.2 < x < 1.2) = \int_{0.2}^{1.2} f(x)\,dx = \int_{0.2}^{1} x\,dx + \int_{1}^{1.2} (2-x)\,dx$$

$$= \left[\frac{x^2}{2}\right]_{0.2}^{1} + \left[2x - \frac{x^2}{2}\right]_1^{1.2}$$

$$= \left(\frac{1}{2} - \frac{0.04}{2}\right) + \left(2.4 - 0.72 - (2 - \frac{1}{2})\right)$$

$$= 0.48 + 1.68 - 1.5 = 0.66$$

**(iii) Find the distribution function of $f(x)$, i.e.,** $F(x) = P(X \leq x)$

$$F(x) = \int_0^x f(x)\,dx$$

Case-wise:
(i) If $x \leq 0$, then $F(x) = 0$
(ii) If $0 < x \leq 1$,

$$F(x) = \int_0^x x\,dx = \left[\frac{x^2}{2}\right]_0^x = \frac{x^2}{2}$$

(iii) If $1 \leq x < 2$,

$$F(x) = \int_0^1 x\,dx + \int_1^x (2-x)\,dx$$

$$= \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^x$$

$$= \frac{1}{2} + \left[2x - \frac{x^2}{2} - (2 - \frac{1}{2})\right]$$

$$= \frac{1}{2} + \left(2x - \frac{x^2}{2} - 1.5\right) = 2x - \frac{x^2}{2} - 1$$

(iv) If $x > 2$, then

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^1 x\,dx + \int_1^2 (2-x)\,dx + \int_2^x 0\,dx$$

$$= \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^2 = \frac{1}{2} + (4-2) - \left(2 - \frac{1}{2}\right) = 1$$

# Question 40

A machine fills soda bottles and the amount of soda filled follows a normal distribution with a mean of 1 litre and a standard deviation of 0.05 litres. The company wants to set control limits so that only the middle 98% of bottles pass the quality check.

(a) Find the lower and upper control limits using the percent point function (PPF).

(b) What percentage of bottles are rejected if the control limits are set at 0.91 litres and 1.09 litres?

## Solution

**(a) Find control limits for middle 98%**

We want the middle 98%, which means 1% in each tail:

$$\text{Lower limit} = \text{PPF}(0.01), \quad \text{Upper limit} = \text{PPF}(0.99)$$

Given:

$$\mu = 1 \text{ litre}, \quad \sigma = 0.05 \text{ litre}$$

Using the standard normal distribution:

$$Z_{0.01} = -2.326, \quad Z_{0.99} = 2.326$$

Now apply the transformation:

$$\text{Lower Limit} = \mu + Z_{0.01} \cdot \sigma = 1 + (-2.326)(0.05) = 1 - 0.1163 = \boxed{0.8837 \text{ litres}}$$
$$\text{Upper Limit} = \mu + Z_{0.99} \cdot \sigma = 1 + (2.326)(0.05) = 1 + 0.1163 = \boxed{1.1163 \text{ litres}}$$

**(b) Bottles rejected if control limits are [0.91, 1.09]**

We calculate the proportion of values outside this range:
Convert to Z-scores:

$$Z_{\text{lower}} = \frac{0.91 - 1}{0.05} = -1.8, \quad Z_{\text{upper}} = \frac{1.09 - 1}{0.05} = 1.8$$

From the standard normal distribution table:

$$P(Z < -1.8) = 0.0359, \quad P(Z > 1.8) = 1 - P(Z < 1.8) = 1 - 0.9641 = 0.0359$$

$$\text{Total rejected} = 0.0359 + 0.0359 = \boxed{7.18\% \text{ of bottles are rejected}}$$

# Question 41

A help desk at a university receives an average of 3 calls every 10 minutes. Assume that the number of calls follows a Poisson distribution.

(a) What is the probability that exactly 5 calls are received in a 10-minute interval?

(b) What is the probability that at most 2 calls are received in a 10-minute interval?

(c) What is the expected number and variance of calls received in a 10-minute interval?

## Solution

Let $X$ be the number of calls received in a 10-minute interval. Given that $X \sim \text{Poisson}(\lambda = 3)$.

**(a) Probability of exactly 5 calls:**

$$P(X = 5) = \frac{e^{-3} \cdot 3^5}{5!} = \frac{e^{-3} \cdot 243}{120} \approx 0.1008$$

**(b) Probability of at most 2 calls:**

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$P(X = 0) = \frac{e^{-3} \cdot 3^0}{0!} = e^{-3} \approx 0.0498$$
$$P(X = 1) = \frac{e^{-3} \cdot 3^1}{1!} = 3e^{-3} \approx 0.1494$$
$$P(X = 2) = \frac{e^{-3} \cdot 3^2}{2!} = \frac{9e^{-3}}{2} \approx 0.2240$$
$$P(X \leq 2) \approx 0.0498 + 0.1494 + 0.2240 = 0.4232$$

**(c) Mean and Variance:**

$$\text{Mean} = \lambda = 3, \quad \text{Variance} = \lambda = 3$$

# Question 42

The time $T$ (hours) that a machine operates before needing repair is modelled by an exponential distribution with rate $\lambda = 0.05$ repairs per hour.

## (a)

What is the probability that the machine operates for at least 20 hours before needing a repair?

## (b)

What is the probability that the machine operates for less than 10 hours before needing a repair?

## (c)

Given that the machine has already operated for 5 hours without a repair, what is the probability that it will last for at least another 5 hours?

## Solution

For an exponential random variable with rate $\lambda$,

$$P(T > t) = e^{-\lambda t}, \qquad P(T \le t) = 1 - e^{-\lambda t}.$$

**(a)**

$$P(T \ge 20) = P(T > 20) = e^{-\lambda \cdot 20} = e^{-0.05 \times 20} = e^{-1}.$$

Numerically $e^{-1} \approx 0.3678794412$.

$$\boxed{P(T \ge 20) = e^{-1} \approx \mathbf{0.3679}}$$

**(b)**

$$P(T < 10) = 1 - P(T \ge 10) = 1 - e^{-\lambda \cdot 10} = 1 - e^{-0.5}.$$

Numerically $e^{-0.5} \approx 0.6065306597$, so

$$P(T < 10) = 1 - 0.6065306597 \approx 0.3934693403.$$

$$\boxed{P(T < 10) = 1 - e^{-0.5} \approx \mathbf{0.3935}}$$

**(c)** Use the memoryless property:

$$P(T \ge 10 \mid T > 5) = P(T \ge 5) = e^{-\lambda \cdot 5} = e^{-0.25}.$$

Numerically $e^{-0.25} \approx 0.7788007831$.

$$\boxed{P(\text{last at least another 5 hrs} \mid \text{already 5 hrs}) = e^{-0.25} \approx \mathbf{0.7788}}$$

# Question 43

Let $X \sim \text{Binomial}(n = 5, p = 0.3)$.

### (a)

Write the PMF formula.

### (b)

Calculate $P(X = 2)$.

### (c)

Find $P(X = 0)$.

### (d)

Calculate $P(X \leq 2)$ by summing PMF values.

### (e)

Find $P(X \geq 4)$.

## Solution

The binomial PMF is

$$P(X = k) = \binom{5}{k} p^k (1 - p)^{5-k}, \qquad k = 0, 1, \ldots, 5.$$

**(a)**

$$P(X = k) = \binom{5}{k} (0.3)^k (0.7)^{5-k}.$$

**(b)** For $k = 2$:

$$\binom{5}{2} = 10, \quad (0.3)^2 = 0.09, \quad (0.7)^3 = 0.343.$$

$$P(X = 2) = 10 \times 0.09 \times 0.343 = 10 \times 0.03087 = 0.3087.$$

$$\boxed{P(X = 2) = \mathbf{0.3087}}$$

**(c)** For $k = 0$:

$$P(X = 0) = (0.7)^5 = 0.16807.$$

$$\boxed{P(X = 0) = \mathbf{0.16807}}$$

**(d)** We already have $P(0) = 0.16807$ and $P(2) = 0.3087$. Now $P(1)$:

$$P(1) = \binom{5}{1} (0.3)^1 (0.7)^4 = 5 \times 0.3 \times 0.2401 = 0.36015.$$

$$P(X \leq 2) = 0.16807 + 0.36015 + 0.3087 = 0.83692.$$

$$\boxed{P(X \leq 2) = \mathbf{0.83692}}$$

**(e)** $P(X \geq 4) = P(4) + P(5)$.

$$P(4) = \binom{5}{4}(0.3)^4(0.7)^1 = 5 \times 0.0081 \times 0.7 = 0.02835.$$

$$P(5) = (0.3)^5 = 0.00243.$$

$$P(X \geq 4) = 0.02835 + 0.00243 = 0.03078.$$

$$\boxed{P(X \geq 4) = \mathbf{0.03078}}$$

## Question 44

The scores of 5 students in a mathematics test are: 72, 85, 90, 78, and 88.
Calculate the mean score.

### Solution:

$$\begin{aligned} \text{Mean} &= \frac{72 + 85 + 90 + 78 + 88}{5} \\ &= \frac{413}{5} \\ &= \boxed{82.6} \end{aligned}$$

## Question 45

A machine produces 4 items with weights (in grams): 102, 100, 98, and 100.
Find the variance of the weights.

### Solution:

$$\begin{aligned} \text{Mean} &= \frac{102 + 100 + 98 + 100}{4} = \frac{400}{4} = 100 \\ \text{Variance} &= \frac{(102 - 100)^2 + (100 - 100)^2 + (98 - 100)^2 + (100 - 100)^2}{4} \\ &= \frac{(2)^2 + (0)^2 + (-2)^2 + (0)^2}{4} \\ &= \frac{4 + 0 + 4 + 0}{4} = \frac{8}{4} \\ &= \boxed{2} \end{aligned}$$

## Question 46

The daily number of customers visiting a shop over 5 days are: 30, 28, 35, 32, and 25.
Calculate the standard deviation.

**Solution:**

$$\begin{aligned}
\text{Mean} &= \frac{30+28+35+32+25}{5} = \frac{150}{5} = 30 \\
\text{Variance} &= \frac{(30-30)^2+(28-30)^2+(35-30)^2+(32-30)^2+(25-30)^2}{5} \\
&= \frac{0^2+(-2)^2+5^2+2^2+(-5)^2}{5} \\
&= \frac{0+4+25+4+25}{5} = \frac{58}{5} = 11.6 \\
\text{Standard Deviation} &= \sqrt{11.6} \approx \boxed{3.41}
\end{aligned}$$

# Question 48

Although most of us buy milk by the quart or gallon, farmers measure daily production in pounds. Ayrshire cows average 47 pounds of milk a day, with a standard deviation of 6 pounds. For Jersey cows, the mean daily production is 43 pounds, with a standard deviation of 5 pounds. Assume that Normal models describe milk production for these breeds.

**a)** If we select an Ayrshire at random, what's the probability that she averages more than 50 pounds of milk a day?

**Solution:**
Let $X \sim N(47, 6^2)$.
Compute the Z-score:

$$Z = \frac{50-47}{6} = \frac{3}{6} = 0.5$$

Using standard normal tables:

$$P(Z < 0.5) = 0.6915$$

Thus,

$$P(X > 50) = 1 - 0.6915 = 0.3085$$

$$\boxed{P(X > 50) = 0.3085}$$

**b)** A farmer has 20 Jerseys. What's the probability that the average production for this small herd exceeds 45 pounds of milk a day?

**Solution:**
Let $X \sim N(43, 5^2)$ and consider the sample mean $\bar{X}$ for $n = 20$.
Compute the standard error (SE):

$$SE = \frac{5}{\sqrt{20}} = \frac{5}{4.472} = 1.118$$

Compute the Z-score:

$$Z = \frac{45-43}{1.118} = \frac{2}{1.118} = 1.79$$

Using standard normal tables:

$$P(Z < 1.79) = 0.9633$$

Thus,

$$P(\bar{X} > 45) = 1 - 0.9633 = 0.0367$$

$$\boxed{P(\bar{X} > 45) = 0.0367}$$

**c)** A farmer has 20 Ayrshires. There's a 99% chance each day that this small herd produces at least how many pounds of milk?

**Solution:**
Let $X \sim N(47, 6^2)$ and consider the sample mean $\bar{X}$ for $n = 20$.
Compute the standard error (SE):

$$SE = \frac{6}{\sqrt{20}} = \frac{6}{4.472} = 1.342$$

Find the Z-score corresponding to the lower 1% tail:

$$z = -2.33 \quad (\text{since } P(Z < z) = 0.01)$$

Compute the lower bound for $\bar{X}$:

$$k = 47 + z \cdot SE = 47 + (-2.33)(1.342) = 47 - 3.126 = 43.874$$

Compute total milk production for 20 cows:

$$\text{Total} = 20 \times 43.874 = 877.5\,\text{pounds}$$

$$\boxed{\text{Total milk} = 877.5\,\text{pounds (99\% confidence lower bound)}}$$

# Question 49

Assume that a test has a mean score of 75 and a standard deviation of 10. Assume the distribution of scores is approximately normal.

**a)** What is the probability that a person chosen at random will make 100 or above on the test?

**Solution:**
Let $X \sim N(75, 10^2)$.
Compute the Z-score:

$$Z = \frac{100 - 75}{10} = \frac{25}{10} = 2.5$$

Using standard normal tables:

$$P(Z < 2.5) = 0.9938$$

Thus,

$$P(X \geq 100) = 1 - 0.9938 = 0.0062$$

$$\boxed{P(X \geq 100) = 0.0062}$$

**b)** What score should be used to identify the top 2.5%?

**Solution:**
We are looking for the score $x$ such that:

$$P(X \geq x) = 0.025 \quad \Longrightarrow \quad P(Z \leq z) = 0.975$$

From the Z-table:

$$z = 1.96$$

Compute the corresponding raw score:

$$x = \mu + z \cdot \sigma = 75 + (1.96)(10) = 75 + 19.6 = 94.6$$

$$\boxed{\text{Top 2.5\% cutoff score} = 94.6}$$

**c)** In a group of 100 people, how many would you expect to score below 60?

**Solution:**
Compute the Z-score:

$$Z = \frac{60 - 75}{10} = \frac{-15}{10} = -1.5$$

Using standard normal tables:

$$P(Z < -1.5) = 0.0668$$

Expected number of people scoring below 60:

$$100 \times 0.0668 = 6.68$$

$$\boxed{\text{Expected number} = 6.68 \approx 7 \text{ people}}$$

**d)** What is the probability that the mean of a group of 100 will score below 70?

**Solution:**
For the sample mean $\bar{X}$, the distribution is:

$$\bar{X} \sim N(75, \frac{10}{\sqrt{100}}) = N(75, 1)$$

Compute the Z-score:

$$Z = \frac{70 - 75}{1} = -5$$

Using approximations:

$$P(Z < -5) = 2.9 \times 10^{-7}$$

$$\boxed{P(\bar{X} < 70) = 2.9 \times 10^{-7}}$$