1/1/19

Module-II

Probability

Sample Space:

→ A set of all possible outcomes of an experiment is called sample space.

Ex: In tossing a coin, all possible outcomes are H & T.

i.e Sample space $S = \{H, T\}$

Event:

→ Every non-empty subset of a sample space of a random experiment

Ex: Tossing a coin and getting a head.

∴ Head is known as an event.

Mutually Exclusive Events:

→ If no two or more of, then can happen simultaneously in the same trial.

Ex: In throwing a die all possible cases are mutually exclusive.

Probability:

→ It is a measure of uncertainty and depends upon a chance either success or failure.

→ If an experiment is performed 'n', the no. of exhaustive cases and 'm' is no. of favourable cases of an event 'A'. probability of an event 'A' is defined by

$$P(A) = \frac{m}{n} = \frac{\text{no. of favourable cases}}{\text{Total no. of exhaustive cases.}} = \frac{n(A)}{n(S)}$$

Scanned by CamScanner

→ The probability value cannot exceed '1'

i.e. $0 \le P \le 1$

NOTE:

Axioms of Probability:

1) $0 \le P \le 1$
2) $P(S) = 1$ (total probability sample space is always 1)
3) If 'A' and 'B' are any two mutually exclusive events then $P(A \cup B) = P(A) + P(B)$

Find the probability of getting a head in tossing a coin

Total sample space $S = \{H, T\}$

$n(S) = 2.$ (no. of exhaustive cases)

No. of favourable cases = 1 (getting head)

$P(A) = \frac{m}{n} = \frac{1}{2}.$

Find the probability of getting one head in tossing two coins.

$S = \{HH, HT, TH, TT\}$

$n(A) = 2$

$n(S) = 4$

$P(A) = \frac{n(A)}{n(S)}$

$= \frac{2}{4} = \frac{1}{2}$

$P(A) = 0.5.$

Find the probability of getting a sum '9' if two dices are thrown.

$S = 6^2 = 36$

$\{(1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)$
$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$
$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$
$(4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6)$
$(5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (5,6)$
$(6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)\}$

$n(A) = 4$

$n(S) = 36$

$P(A) = \frac{4}{36} = \frac{1}{9}$

$= 0.1$

$n(A) = \{(3,6), (4,5), (5,4), (6,3)\}$

Combination:

→ To find the number of ways in which objects can be selected from a set of ~~sm~~ 'n' distinct objects is called the no. of combinations of 'n', objects taken 'r' at a time and is denoted by nC_r.

$$nC_r = \frac{n!}{(n-r)!\,r!} \quad (0, r)$$

→ unordered of the objects is called the combination.

Ex: In how many ways 3 students can be selected from 15 students.

$n = 15,\ r = 3$

$$nC_r = 15C_3 = \frac{15!}{12!\,3!}$$

$$= \frac{\cancel{15}^{5} \times \cancel{14}^{7} \times 13}{\cancel{6}\,\cancel{2}}$$

$= 3 \times 7 \times 13$

$= 455$

Scanned by CamScanner

Permutations:

→ Objects are choosen from a set of 'n' distinct objects any particular arrangment or order of this objects is called a permutation and is denoted by ^nP_r.

$$^nP_r = \frac{n!}{(n-r)!}$$

(or)

$$^nP_r = n(n-1)(n-2)\cdots(n-r+1).$$

→ If repetations are not allowed, how many four digits number can be formed from digits 1, 2, 3, 4, 5, 6, 7

$$^7P_4 = 7(6)(5)(4)$$

$$= 840$$

Find the probability of getting 2 diamonds if we draw two cards at random from a pack of 52 cards.

Sol Total no. of cards = 52.

(Total exhaustive cases) $n(S) = {}^{52}C_2$

no. of favourable cases $n(A) = {}^{13}C_2$

$$P(A) = \frac{n(A)}{n(S)} = \frac{^{13}C_2}{^{52}C_2} = \frac{1}{17}.$$

→ 3 light bulbs are choosen at random from 12 bulbs of which 5 are defective. Find the probability that.

i) all are defective.

ii) One is defective.

iii) Two are defective.

Sol Total exhaustive cases $n(S) = 12C_3 = \frac{12!}{9!\,3!} = \frac{\overset{2}{\cancel{12}} \times 11 \times 10 \times \cancel{9!}}{\cancel{9!} \times \cancel{6}}$

$= 220$

i) no. of favourable case $n(A) = 5C_3$

$$= \frac{5!}{2!\,3!} = \frac{120}{2 \times 6} = 10$$

$$P(A) = \frac{\frac{5!}{2!\,3!}}{12C_3} = \frac{\frac{120}{2\times 6}}{220} = \frac{10}{220} = \frac{1}{22}.$$

ii) One is defective.

no. of fav. cases $n(A) = 5C_1 \times 7C_2$

$$P(A) = \frac{5C_1 \times 7C_2}{12C_3} = \frac{21}{44}.$$

(iii) Two are defective.

no. of favourable case $n(A) = 5C_2 \times 7C_1$

$$P(A) = \frac{5C_2 \times 7C_1}{12C_3} = \frac{7}{22}$$

→ A bag contains 5 red balls, 8 blue balls, 11 white balls. Three balls are drawn together from the box. Find the probability of 1) one is red

one is blue

one is white

2) Two white and one red

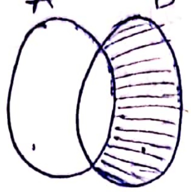
3) 3 white balls.

21/1/19

Addition theorem on Probability

St: if 'S' is a sample and A & B are any events in S then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof: A & $A^c \cap B$ are disjoint events.



$$P(A \cup (A^c \cap B)) = P(A) + P(A^c \cap B) \rightarrow ①$$

But $A \cup (A^c \cap B) = A \cup B \rightarrow ②$

① & ②

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

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

If A, B, C are any 3 events then probability of

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

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

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

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

$$= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C).$$

→ If $P(A \cup B) = 0.05$ and $P(A \cap B) = 0.15$ find $P(\bar{A}) + P(\bar{B})$

$$P(\bar{A}) = 1 - P(A)$$

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

$$P(A \cup B) = 1 - (P(\bar{A})) + 1 - P(\bar{B}) - P(A \cap B)$$

$$0.05 = 1 - 0.15 - (P(\bar{A}) + P(\bar{B})) + 1$$

$$P(\bar{A}) + P(\bar{B}) = 1.8$$

→ A card is drawn from a well shuffled pack of cards. what is the probability that it is either a spade (or) ace.

Total Sample space $n(S) = 52$

→ 3 students A, B, C are in running race. A and B have the same probability of winning and each is twice as likely to win as C. Find the probability that B or C wins.

$S = A \cup B \cup C$

$P(A) = 2P(C)$

$P(B) = 2P(C)$

$P(A) = P(B)$

$P(A) + P(B) + P(C) = 1$

$2P(C) + 2P(C) + P(C) = 1$

$5P(C) = 1$

$P(C) = 1/5$

$P(A) = 2P(C) = 2/5$

$P(B) = 2P(C) = 2/5$

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

$= \frac{2}{5} + \frac{1}{5} -$

Scanned by CamScanner

→ From a city 3 newspapers A, B, C are published. A is read by 20%, B is read by 16%, C is read by 14%. Both A and B are read by 8%, both A and C read by 5%. Both B and C are read by 4% and all A, B, C are read by 2%. What is the percentage of the population that reads atleast one paper.

$P(A) = \frac{20}{100}$, $P(B) = \frac{16}{100}$, $P(C) = \frac{14}{100}$.

$P(A \cap B) = 8\% = \frac{8}{100}$, $P(A \cap C) = \frac{5}{100}$, $P(B \cap C) = \frac{4}{100}$

$P(A \cap B \cap C) = \frac{2}{100}$.

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

$$= \frac{20}{100} + \frac{16}{100} + \frac{14}{100} - \frac{8}{100} - \frac{5}{100} - \frac{4}{100} + \frac{2}{100}$$

$$= \frac{20+16+14-8-5-4+2}{100} = \frac{52-17}{100} = \frac{35}{100}.$$

Conditional Probability:

→ If B is arbitary event in a sample space and $P(B) > 0$ then $P(A|B)$ is defined by

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}.$$

→ This is called conditional probability of $P(A|B)$, $P\left(\frac{A}{B}\right)$ represent the conditional probability of

→ occurrency of A when event B has already happen.

Multiplication theorem:

→ If B is arbitary event in sample space and $p(B)>0$ then probability of A given B is defined by

$\left[P\left(\frac{A}{B}\right) = \frac{P(A\cap B)}{P(B)}\right]$ this is called conditional probability of A given B, $P\left(\frac{A}{B}\right)$ represent the conditional probability of occurance of A when the event B has already happen.

* $P(A\cap B) = P(B)\cdot P\left(\frac{A}{B}\right)$. $P(B)>0$ (or)

(*) $P(A\cap B) = P(A)\cdot P\left(\frac{B}{A}\right)$ $P(A)>0$

we know that $P(A) = \frac{n(A)}{n(S)} \rightarrow ①$

$P(B) = \frac{n(B)}{n(S)} \rightarrow ②$

$P(A\cap B) = \frac{n(A\cap B)}{n(S)} \rightarrow ③$

$P(A\cap B)$. For the conditional event $\frac{A}{B}$, the favourable outcomes must be one of the sample point.

$P(A\cap B) = \frac{n(A\cap B)}{n(S)} \times \frac{n(B)}{n(B)}$

$= \frac{n(A\cap B)}{n(B)} \cdot \frac{n(B)}{n(S)}$

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

similarly $P(A\cap B) = \frac{n(A\cap B)}{n(S)} \cdot \frac{n(A)}{n(A)}$

$= \frac{n(A\cap B)}{n(A)} \cdot \frac{n(A)}{n(S)}$

$= P(B/A)\cdot P(A)$.

→ Two marbles are drawn in success from a box containing 10 red, 30 white, 20 blue and 15 orange marbles, with replacement begining made after each draw. Find the probability that (i) both are white (ii) 1 is red, 1 is white

Sol Total no. of marbles in the box = 75

(i) Let E_1 be the event of the first drawn marble is white then, $P(E_1) = \frac{30}{75}$.

Let E_2 be the even of the second drawn marble is also white.

$$\frac{P(E_2)}{P(E_1)} = \frac{30}{75}$$

$$P(E_1 \cap E_2) = ?$$

$$P(E_1 \cap E_2) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right)$$

$$= \frac{30}{75} \times \frac{30}{75}$$

$$= \frac{4}{25}$$

(ii) Let E_1 be the event i.e. first drawn marble is

$$P(E_1) = \frac{10}{75}$$

Let E_2 be the event that the drawn marble is white $P\left(\frac{E_2}{E_1}\right) = \frac{30}{75}$.

$$P(E_1 \cap E_2) = ?$$

$$P(E_1 \cap E_2) = P(E_1) \cdot P\left(\frac{E_2}{E_1}\right)$$

$$= \frac{10}{75} \times \frac{30}{75}$$

(Q) The probability of students A, B, C, D solved a problem are $\frac{1}{3}, \frac{2}{5}, \frac{1}{5}, \frac{1}{4}$ respectively. If all of them to solve the problem, what is the probability that the problem is solved.

Ans $P(A) = \frac{1}{3}$, $P(B) = \frac{2}{5}$, $P(C) = \frac{1}{5}$; $P(D) = \frac{1}{4}$

The probability that the problem is not solved by A, B, C, D are

$P(\bar{A}) = 1 - \frac{1}{3} = \frac{2}{3}$, $P(\bar{B}) = 1 - \frac{2}{5} = \frac{3}{5}$

$P(\bar{C}) = 1 - \frac{1}{5} = \frac{4}{5}$, $P(\bar{D}) = 1 - \frac{1}{4} = \frac{3}{4}$

The probability that problem is not solved by A, B, C and D together

$$P(\bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D}) = P(\bar{A}) \cdot P(\bar{B}) \cdot P(\bar{C}) \cdot P(\bar{D})$$

$$= \frac{2}{3} \times \frac{3}{5} \times \frac{4}{5} \times \frac{3}{4}$$

$$= \frac{6}{25}$$

$$\rightarrow 1 - \frac{6}{25} = \frac{19}{25}$$

(Q) 3 machines I, II, III produce 40%, 30% of the total number of items of factory. The percentage of defective items of these machine are 4%, 2%, 3%. If one item is selected at random, find the probability that the item is defective.

Sol Let A, B, C be the events that the machines 1, 2, 3 be choosen respectively and let D be the event which denote the defective items.

6 6

Scanned by CamScanner

$P(A) = \frac{400}{100}$; $P(B) = \frac{30}{100}$, $P(C) = \frac{30}{100}$

From the given data we have

$P\left(\frac{D}{A}\right) = \frac{4}{100}$, $P\left(\frac{D}{B}\right) = \frac{2}{100}$, $P\left(\frac{D}{C}\right) = \frac{3}{100}$

The probability that the selected item at random is defective is

$P(D) = P(A) \cdot P(D/A) + P(B) \cdot P\left(\frac{D}{B}\right) + P(C) \cdot P(D/C)$

$= \frac{40}{100} \times \frac{4}{100} + \frac{30}{100} \times \frac{2}{100} + \frac{30}{100} \times \frac{3}{100}$

$= \frac{31}{1000} = 0.031$

$= 31\%$.

(Q) 2 dice are thrown. Let A be the event that the sum of the points on the faces is 9. Let B be the sum event that atleast one number is 6.

(i) $P(A \cup B)$, (ii) $P(A \cap B)$, (iii) $P(\bar{A} \cup \bar{B})$

$P(A) = \frac{4}{36}$

$P(B) = \frac{11}{36}$

$P(A \cap B) = \frac{2}{36} = \frac{1}{18}$

$P(A \cup B) = \frac{4}{36} + \frac{11}{36} - \frac{2}{36}$

$= \frac{15-2}{36} = \frac{13}{36}$

$P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B)$

$= 1 - \frac{1}{18}$

$= \frac{17}{18}$

Scanned by CamScanner

Baye's Theorem:

Suppose $E_1, E_2, \ldots E_n$ are mutually designed event of a sample spaces such that $P(E_i) \neq 0, (i = 1, 2, 3, \ldots n)$ for any arbitary event A. which is a subset of $\sum_{i=1}^{n} E_i$

i.e. $P(A) > 0$, $P(E_i/A) = \dfrac{P(A/E_i) \cdot P(E_i)}{\sum_{i=1}^{n} P(A/E_i) \cdot P(E_i)}$

Proof: given that $A \subseteq \sum_{i=1}^{n} E_i$

$$A \cap \sum_{i=1}^{n} E_i = A$$

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) \rightarrow ①$$

since $E_1, E_2, \ldots E_n$ are mutually Disjoint events.

$A \cap E_1, A \cap E_2, \ldots A \cap E_n$ are also mutually disjoint events

$$P(A \cap E_i) = P(E_i) \cdot P(A/E_i) \rightarrow ②$$

$$P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A/E_i) \rightarrow ③$$

$$P(E_i/A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(A \cap E_i)}$$

$$= \frac{P(E_i) \cdot P(A/E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$$

(Q) In bolt factory. machines a, b, c manufactures 20%, 30%, 50% of the total of their output and 6%, 3%, 12% are defective. A bolt is drawn at random and find to be defect.

Find the probability that it is manufacture from

(i) Machine A (ii) Machine B (iii) Machine C.

Sol Let $P(A), P(B), P(C)$ be the probability of the events that the bolts are manufactured by the machine A, B, C resp.

$$P(A) = \frac{20}{100}, \quad P(B) = \frac{30}{100}, \quad P(C) = \frac{50}{100}.$$

* Let E denote that a bolt is defective, then

$$P(E/A) = 6/100, \quad P(E/B) = 3/100, \quad P(E/C) = \frac{2}{100}.$$

(i) If bolt is defective then the probability that is from machine A is

$$P(A/E) = \frac{P(A) \cdot P(E/A)}{P(A) \cdot P(E/A) + P(B) \cdot P\left(\frac{E}{B}\right) + P(C) \cdot P\left(\frac{E}{C}\right)}$$

$$= \frac{\frac{20}{100} \times \frac{6}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}}$$

$$= \frac{12}{31}$$

(ii) If bolt is defective then probability that is from machine B is

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

$$= \frac{\frac{30}{100} \times \frac{3}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}} = \frac{9}{31}$$

(iii) If bolt is defective then probability that it is from machine C is

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

$$= \frac{\frac{50}{100} \times \frac{2}{100}}{\frac{20}{100} \times \frac{6}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{50}{100} \times \frac{2}{100}} = \frac{10}{31}$$

(Q) A business man goes to hotel x, y, z - 20%, 50%, 30% of the time resp. It is known, that 5%, 4%, 8% of the rooms in x, y, z hotels have fault plumbing what is the pb of room having fault plumbing is assigned to hotel 'z'.

Sol $$P(Z/E) = \frac{P(Z) \cdot P(E/Z)}{P(A) \cdot P(E/A) + P(B) \cdot P(E/B) + P(C) \cdot P(E/C)}$$

$$P(A) = \frac{20}{100}, \quad P(B) = \frac{50}{100}, \quad P(C) = \frac{30}{100}.$$

$$P(E/A) = \frac{5}{100}, \quad P(E/B) = \frac{4}{100}, \quad P(E/C) = \frac{8}{100}.$$

$\therefore$ sub above values in $P(Z/E)$

$$\therefore P(Z/E) = \frac{24}{540}.$$

(Q) A bag 'A' contains 2 white and 3 red balls. A bag 'B' contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that the redball drawn is from 'Bag B'.

Sol $P(A) = 1/2, \quad P(B) = 1/2$

Red ball $\rightarrow R$

$$P(R/A) = \frac{3}{5}, \quad P(R/B) = \frac{5}{9}$$

$$P(B/R) = \frac{P(R/B) \cdot P(B)}{P(A) \cdot P(R/A) + P(B) \cdot P(R/B)}$$

$$= \frac{1/2 \times 5/9}{1/2 \times \frac{3}{5} + 1/2 \times 5/9}$$

$$= \frac{25}{52}.$$

Module - 2

Random Variables and Probability Distribution

Random Variables:

Suppose 'S' is the sample space of some experiment we know that outcomes of the experiment are the elements of the sample space and they need not be numbers. Sometimes, we wish to assign a specific number to each outcome.

Eg: The number of heads in tossing two coins or 3 coins such assignment is called a random variable.

In the above example, we may consider the random variable which is the no. of heads.

Outcomes	HH	HT	TH	TT
Variable X	2	1	1	0

Random variable:

Assigning a numerical values to each outcome in the experiment is called Random variable.

Random Variable

D.R.V finite → probability mass function

$$P(x) = \sum_{i=1}^{n}$$

Properties:

$$\sum_{i=1}^{P.n} P(x) = 1$$

$$P(x) \geq 0$$

↓

mean, variance, standard deviation of discrete probability distribution

↓

Binomial, poisson, Geometric distribution

C.R.V (infinite) → probability density function

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

properties

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

$$f(x) \geq 0$$

↓

mean, variance, Standard Deviation of continuous probability distribution

↓

Normal, exponential distribution

Types of Random Variables:

1) Discrete random variables
2) Continuous random variables

Discrete Random Variables:

A Random variables which can take only finite variables is called Discrete random variables

Eg: A coin is tossed, $X(H) = 1$ if head occurs.

$X(H) = 0$ if tail occurs

Here X take the value only countable number of values

$\therefore X$ is a discrete random variable

Probability Mass function (or) Discrete probability distribution:

We associate a number $p_i = P(x_i) = P(x = x_i)$, $i = 1, 2, \ldots$ is called probability of x_i must satisfy the following conditions

$$P(x_i) \geq 0 \ \forall \text{ value } i.$$

$$\sum_{i=1}^{n} P(x_i) = 1$$

$\Rightarrow$ The function P is called probability mass function of a random variable (X)

Mass of discrete Probability distribution:

Let $x_1, x_2, \ldots x_n$ are the values of the Random variables X which it can take corresponding probabilities $P(X_1), P(X_2), \ldots P(X_n)$, then the mean of random variables is

$$E(X) = \mu = \sum_{i=1}^{n} x_i P(x_i)$$

$E(x)$ expectation value of X.

Variance of discrete probability distribution.

If a random variable take the values $x_1, x_2 - - - x_n$ and the corresponding probabilities $P(x_1), P(x_2), - - - P(x_n)$ then the variance of the random variable.

$$V(x) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$$

$$= \sum_{i=1}^{n} P_i x_i^2 - \mu^2$$

$$= E(x^2) - (E(x))^2$$

$$V(x) = E(x^2) - \mu^2.$$

Standard Deviation:

It is the positive square root of the variance

$$SD = \sigma = \sqrt{V(x)}.$$

Theorem(1): If x is a random variable and K is a constant then $E(X+K) = E(X)+K$

Sol. By the def.

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

$$E(X+K) = \sum_{i=1}^{n} (x_i + K) P_i$$

$$= \sum_{i=1}^{n} (x_i P_i) + \sum_{i=1}^{n} K P_i$$

$$= \sum_{i=1}^{n} x_i P_i + K \sum_{i=1}^{n} P_i \quad \left(\sum_{i=1}^{n} P_i = 1\right)$$

$$E(X+K) = E(x) + K$$

Scanned by CamScanner

Theorem(2): If X is a random variable and a, b are constants then $E(aX+b) = aE(X)+b$

Proof: $E(X) = \sum_{i=1}^{n} x_i p_i$

$$E(aX+b) = \sum_{i=1}^{n} (ax_i+b) p_i$$

$$= \sum_{i=1}^{n} (ax_i p_i + bp_i)$$

$$= \sum_{i=1}^{n} ax_i p_i + \sum_{i=1}^{n} bp_i$$

$$= a\sum_{i=1}^{n} x_i p_i + b\sum_{i=1}^{n} p_i \qquad \left(\sum_{i=1}^{n} p_i = 1\right)$$

$$E(aX+b) = aE(X)+b$$

Theorem(3): If X and Y are two discrete random variables then $E(X+Y) = E(X)+E(Y)$ provides $E(X)$ and $E(Y)$ exists.

Proof: Let 'x' assume the values $x_1, x_2, ---- x_n$ and 'y' assume the values $y_1, y_2, --- y_n$ then the definitio

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

$$E(Y) = \sum_{j=1}^{n} y_j p_j$$

Let $p_{ij} = P(X = x_i \cap Y = y_j) = P(x_i, y_j)$

(This is called joint probability functions of X and Y)

The sum $(X+Y)$ is also a random variable which can take (x_i+y_j) $i = 1, 2, --- n$ & $j = 1, 2, --- n$.

(ctx.

By the definition

$$E(X+Y) = E(x_i+y_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i+y_j) P_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j P_{ij}$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} P_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} P_{ij}$$

$$\left(\sum_{i=1}^{n} P_j = 1 \quad P_{ij} = P_j \right)$$

$$= \sum_{i=1}^{n} x_i P_i + \sum_{j=1}^{m} y_j P_j$$

$$= E(X) + E(Y)$$

$$\left(\sum_{j=1}^{m} P_j = 1 \quad P_{ij} = P_i \right)$$

Theorem(4): If X and Y are two independent random variables then $E(XY) = E(X) \cdot E(Y)$

Proof: Let X assume the values $x_1, x_2, \ldots x_n$ and Y assume the values $y_1, y_2, \ldots y_n$ then by the definition

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

$$E(Y) = \sum_{j=1}^{m} y_j P_j$$

$$P_{ij} = P(X = x_i \cap Y = y_j)$$

$$= P(x_i) \cdot P(y_j)$$

$$P_{ij} = P_i \cdot P_j \longrightarrow (1)$$

XY is also random variable, which can take $x_i y_j$

$i = 1, 2, \ldots n$, $j = 1, 2, \ldots m$

by def.

$$E(X,Y) = E(x_i y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i y_j) P_{ij} \quad \text{(from equ (1))}$$

Scanned by CamScanner

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i \cdot y_j)\, P_i \cdot P_j$$

$$= \sum_{i=1}^{n} x_i P_i \cdot \sum_{j=1}^{m} y_j P_j$$

$$= E(X) \cdot E(Y)$$

Note:

1) $E(X+Y+Z) = E(X) + E(Y) + E(Z)$

2) $E(aX+bY) = aE(X) + bE(Y)$

3) $E(X-\bar{X}) = 0$

4) $E(XYZ) = E(X) \cdot E(Y) \cdot E(Z)$

(Q) A random variable X as the following probability function.

X	0	1	2	3	4	5	6	7
P(X)	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2+K$

Find

1) K 2) Evaluate $P(X<6)$, $P(X\geq 6)$, $P(0<X<5)$

3) If $P(X\leq K) > \frac{1}{2}$ find the minimum value of K.

4) Determine, the distribution function of X.

5) Mean

6) Variance

Sol (1) By def $\sum_{i=1}^{7} P_i = 1$

$0 + K + 2K + 2K + 3K + K^2 + 2K^2 + 7K^2 + K = 1$

$10K^2 + 9K = 1$

$10K^2 + 9K - 1 = 0$

$$10k^2 + 10k - k - 1 = 0$$

$$10k(k+1) - 1(k+1) = 0$$

$$(10k-1)(k+1) = 0$$

$$k = -1, \quad k = \frac{1}{10}$$

$k = -1$ is not possible

so, $k = 1/10$

2) (i) $P(X<6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$$= 0 + k + 2k + 2k + 3k + k^2$$

$$= k^2 + 8k$$

$$= \frac{1}{100} + \frac{8}{10} = \frac{81}{100} = 0.81$$

(ii) $P(X \geq 6) = P(X=6) + P(X=7)$

$$= 2k^2 + 7k^2 + k$$

$$= 9k^2 + k$$

$$= \frac{9}{100} + \frac{1}{10} = \frac{9+10}{100} = \frac{19}{100}$$

$$= 0.19$$

(iii) $P(0<X<5) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$$= k + 2k + 2k + 3k$$

$$= 8k = \frac{8}{10} = 0.8$$

3) The req. minimum value of k is obtained as below.

$$P(X \leq k) > \frac{1}{2}$$

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

$$= 0 + k$$

$$= 1/10 \not> \frac{1}{2}$$

$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$= 0 + k + 2k$

$= 3k = \frac{3}{10} \cancel{>} 1/2$

$P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 5k = \frac{5}{10} \cancel{>} \frac{1}{2}$

$P(X \le 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= 8k = \frac{8}{10} > \frac{1}{2}.$

$\therefore$ Min value of k for which $P(X \le k) > \frac{1}{2}$ is

$\boxed{k = 4}$

(4) The distribution function of X is given by the following table.

x	$P(x)$	$F(x) = P(X \le x_i)$
0	0	$0 = 0$
1	k	$k = 1/10$
2	$2k$	$3k = 3/10$
3	$2k$	$5k = 5/10$
4	$3k$	$8k = 8/10$
5	k^2	$8k + k^2 = 81/100$
6	$2k^2$	$8k + 3k^2 = 83/100$
7	$7k^2 + k$	$9k + 10k^2 = 1$

5) Mean $= \mu = \sum_{i=1}^{n} x_i P(x_i)$

$= 0 + k + 4k + 6k + 12k + 5k^2 + 12k^2 + 49k^2 + 7k$

$= 66k^2 + 30k$

$= \frac{66}{100} + \frac{30}{10} = 0.66 + 0.3 \boxed{\mu = 3.66}$

(6) variance $= \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$

$$v = \sum_{i=1}^{n} x^2 p_i - \mu^2$$

$$= 0(0) + 1^2(k) + 2^2(2k) + 3^2(2k) + 4^2(3k) + 5^2(k^2) + 6^2(2k^2) + 7^2(7k^2 + k) - \mu^2$$

$$= 124k + 440k^2 - \mu^2$$

$$= \frac{124}{10} + \frac{440}{100} - (3.66)^2$$

$$v = 3.4044$$

(Q) Let X denote the minimum of the 2 numbers that app when a pair of fair die is thrown once

(i) Determine the discrete probability distribution.

(ii) Expectation (iii) variance.

When two dices are thrown, total no. of outcomes is 36

The minimum number can be 1, 2, 3, 4, 5, 6

$$S = \{(1,1) \text{ ------ } \text{ ----- } (6,6)\} = 36$$

$x = 1$, $P((1,1); (1,2), (1,3), (1,4), (1,5), (1,6)$
$(2,1), (3,1), (4,1), (5,1), (6,1))$

$$P(X=1) = 11/36$$

$X = 2$ $P((2,2), (2,3), (2,4), (2,5), (2,6),$
$(3,2), (4,2), (5,2), (6,2))$

$$P(X=2) = 9/36.$$

$x=3$ $P((3,3),(3,4),(3,5),(3,6),$
$(4,3),(5,3),(6,3))$

$P(X=3) = 7/36$

$X=4$ $P((4,4),(4,5),(4,6),$
$(5,4),(6,4))$

$P(X=4) = 5/36.$

$X=5$ $P((5,5),(5,6),$
$(6,5))$

$P(X=5) = 3/36$

$X=6$ $P((6,6))$

$P(X=6) = 1/36.$

X	1	2	3	4	5	6
P(x)	11/36	9/36	7/36	5/36	3/36	1/36

Mean $\mu = \sum_{i=1}^{n} x_i p_i$

$= x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 + x_6 p_6$

$= 1(11/36) + 2(9/36) + 3(7/36) + 4(5/36) + 5(3/36) + 6(1/36)$

$= 91/36 = 2.52$

Variance $= \sum_{i=1}^{w} x^2 p_i - \mu^2$

$= 1/36 + \frac{4 \times 9}{36} + \frac{9 \times 7}{36} + \frac{16 \times 5}{36} + \frac{25 \times 3}{36} + \frac{36 \times 6}{36} - (2.52)^2$

$= 2.00$

Scanned by CamScanner

(Q) A sample of 4 items is selected at random from a box containing 12 items of which 5 are defective. Find the expected number of defective items.

Let 'X' denote the no. of defective items, among 4 items drawn from 12 items.

X can take the values of 0, 1, 2, 3 & 4

no. of good items = 7

no. of defective items = 5

$P(X=0)$ (good) $= \frac{7C_4}{12C_4}$

$= 7/99 = 0.07$

$P(X=1)$ (1 def & 3 good) $= \frac{7C_3 \times 5C_1}{12C_4}$

$P(X=2)$ (2 def & 2 good) $= \frac{5C_2 \times 7C_2}{12C_4}$

$= 0.4242$

$P(X=3)$ (3 def & 1 good) $= \frac{5C_3 \times 7C_1}{12C_4} = 0.1414$

$P(X=4)$ (4 def) $= \frac{5C_4}{12C_4}$

$= 0.0101$

x	0	1	2	3	4
$P(x)$	0.07	0.35	0.42	0.14	0.0101

$E(X) = \sum_{i=1}^{n} x_i P(x_i)$

$= 0 + 1(0.3535) + 2(0.4202) + 3(0.1414) + 4(0.0101)$

$= 1.666$

Scanned by CamScanner

Probability distribution function

Let X be a random variable then the function $F(x) = P(X \leq x)$, $-\infty < x < \infty$ is called the distribution function of X.

Properties of distribution function.

If F is the distribution function of the random variable 'x' and if $a < b$ then

1) $P(a < X \leq b) = F(b) - F(a)$

2) $P(a \leq X \leq b) = P(X = a + [F(b) - F(a)])$

3) $P(a < X < b) = [F(b) - F(a) - P(X = b)]$

4) $P(a \leq X < b) = [F(b) - F(a)] - P(X = b) + P(X = a)$

5) $F(-\infty) = \lim_{x \to 0} F(x) = 0$

6) $F(\infty) = \lim_{x \to \infty} F(x) = 1$

Q) Discrete Random variable has followed the distribution function

$$F(x) = \begin{cases} 0, & \text{for } x < 1 \\ 1/3 & \text{for } 1 \leq x < 4 \\ 1/2 & \text{for } 4 \leq x < 6 \\ 5/6 & \text{for } 6 \leq x < 10 \\ 1 & \text{for } x \geq 10 \end{cases}$$

Sol:

Bernouli Distribution:

A random variable 'x' is said to have a Bernoulli distribution with parameter P. If its probability mass function is given by

$$P(X=x) = \begin{cases} p^{x}(1-p)^{x-1} & \text{for } x=0,1 \\ 0 & \text{otherwise} \end{cases}$$

⇒ Mean of Bernoulli $\mu = P$

⇒ Variance $v(x) = pq$

⇒ Standard deviation $\sigma = \sqrt{pq}$.

~~Bernoulli~~

Binomial Distribution:

A random variable 'x' is said to follow binomial distribution if it assume only non-negative values and its probability mass function is given by

$$P(X=x) = P(x) = \begin{cases} nC_x p^{x} q^{n-x} & \text{for } x=0,1,2,\dots n \\ 0 & \text{otherwise} \end{cases}$$

n = total no. of trials, x = selecting outcomes

p = success of trials

q = failure of the trials

C = combinations

& $p+q=1$

Mean of the binomial Distribution:

The probability may function of binomial distribution is given by

$$P(x) = nC_x p^x q^{n-x} \quad \text{for } x = 0, 1, 2, \ldots n$$

$$\mu = E(X) = \sum_{x=0}^{n} x p(x)$$

$$= \sum_{x=0}^{n} x \cdot nC_x p^x q^{n-x}$$

$$= 0 + nC_1 p^1 q^{n-1} + 2nC_2 p^2 q^{n-2} + 3nC_3 p^3 q^{n-3} + \cdots + nC_n p^n$$

$$= npq^{n-1} + \frac{2n(n-1)}{2!} p^2 q^{n-2} + \frac{3 \cdot n(n-1)(n-2)}{3!} p^3 q^{n-3} + \cdots + n^n C_n p^n$$

$$= np\left(q^{n-1} + \frac{(n-1)}{1!} pq^{n-2} + \frac{(n-1)(n-2)}{2} p^2 q^{n-3} + \cdots + p^{n-1} \right)$$

$$= np(p+q)^{n-1} \qquad (p+q=1)$$

$$= np(1)^{n-1}$$

$$\boxed{\mu = np}$$

Variance of binomial distribution:

$$P(x) = nC_x p^x q^{n-x} \quad \text{for } x = 0, 1, 2, \ldots n$$

$$\mu = np$$

$$V(x) = E(X^2) - (E(X))^2$$

$$= \sum_{x=0}^{n} x^2 p(x) - \mu^2$$

$$= \sum_{x=0}^{n} (x(x-1) + x) p(x) - \mu^2$$

$$= \sum_{x=0}^{n} x(x-1)\, p(x) + \sum_{x=0}^{n} x\, p(x) - \mu^2$$

$$= \sum_{x=0}^{n} x(x-1)\, {}^nC_x p^x q^{n-x} + \mu - \mu^2$$

$$= \left(2\, nC_2 p^2 q^{n-2} + 3\times 2\, nC_3 p^3 q^{n-3} + 4\times 3\, nC_4 p^4 q^{n-3} + \cdots + n(n-1)\cdot nC_n p^n\right) + \mu - \mu^2$$

$$= \left(2\frac{n(n-1)}{2!} p^2 q^{n-2} + \frac{3\times 2 n(n-1)(n-2)}{3!} p^3 q^{n-3} + \cdots + n(n-1)p^n\right) + \mu - \mu^2$$

$$= n(n-1)p^2\left(q^{n-2} + (n-2)pq^{n-3} + \cdots + p^{n-2}\right) + \mu - \mu^2$$

$$= n(n-1)p^2 (p+q)^{n-2} + \mu - \mu^2$$

$$= n(n-1)p^2 (1)^{n-2} + \mu - \mu^2$$

$$= n(n-1)p^2 + \mu - \mu^2$$

$$= n^2p^2 - np^2 + (np) - (np)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p)$$

$$V(x) = npq$$

Recurrence relation for binomial distribution

We know that $P(x) = nC_x p^x q^{n-x}$ for $x = 0, 1, 2, \ldots n \rightarrow (1)$

$P(x+1) = nC_{x+1} p^{x+1} q^{n-(x+1)} \rightarrow (2)$

$(2) \div (1)$

$$\frac{P(x+1)}{P(x)} = \frac{nC_{x+1} p^{x+1} q^{n-(x+1)}}{nC_x p^x q^{n-x}}$$

$$P(x+1) = \frac{nC_{x+1}}{nC_x} \cdot \frac{p}{q} P(x)$$

$$P(x+1) = \frac{n-x}{(x+1)} \cdot \frac{p}{q} P(x)$$

(Q) A fair coin is tossed for 6 times. Find the probability of getting 4 heads.

p = probability of getting a head $= \frac{1}{2}$

q = probability of not getting head $= \frac{1}{2}$

$n = 6$

The probability mass distribution of a function is given by

we know that, $p(x) = {}^nC_x \, p^x q^{n-x}$

$$P(X=4) = 6C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{6-4}$$

$$= \frac{6!}{2!\,4!} \cdot \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2$$

$$= \frac{6 \cdot 5 \cdot \cancel{4!}}{2 \cdot \cancel{4!}} \left(\frac{1}{2}\right)^6$$

$$= 15 \cdot \left(\frac{1}{2}\right)^6$$

$$= \frac{15}{64} = 0.238$$

(Q) Determine the probability of getting the sum 6 exactly 3 times in 7 throws with a pair of fair die.

Sol. In a single throw of a pair of fair die, sum of six can occur (1,5), (2,4), (3,3), (4,2), (5,1)

$$p = 5/36$$

$$q = 1 - p = 1 - \frac{5}{36} = \frac{31}{36}$$

$$n = 7, \quad x = 3$$

$$P(X=3) = 7C_3 \, (5/36)^3 \left(\frac{31}{36}\right)^{7-3}$$

$$= 0.051$$

Scanned by CamScanner

(4) 10 coins are thrown simultaneously. find the probability of (i) getting atleast 7 heads

(ii) atleast 6 heads

$p = 1/2, \; q = 1/2, \; n = 10, \; x = 7, 8, 9, 10$

$P(x) = {}^nC_x p^x q^{n-x} \; ; \; n = 10$

(i) $P(x \geq 7) = P(x=7) + P(x=8) + P(x=9) + P(X=10)$

$= 10C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + 10C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + 10C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right)^1 + 10C_{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^0$

$= 0.1719$

(2) $P(X \geq 6) = P(X=6) + P(X \geq 7)$

$=$

(5) If the probability of defective bolt is 0.2. Find mean, standard deviation for the distribution of bolt in a total of 400

$p = 0.2$

$q = 1 - p$

$q = 1 - 0.2 = 0.8$

$n = 400$

mean $= np$

$= 400 \times 0.\cancel{8} = \cancel{320} \; 80$

$SD = \sqrt{npq} = \sqrt{400 \times 0.8 \times 0.2}$

$= 8$

Q) Out of 800 families with 5 children each. How many would you expect to have (i) 3 boys (ii) 5 girls (iii) either 2 or 3 boys. (iv) atleast 1 boy.

assume equal probability for boys & girls.

Sol probability of each boys $p = 1/2$

girl $= q = 1/2$

$n = 5$

$P(x) = nC_x \, p^x \, q^{n-x}$

$P(x) = 5C_x (1/2)^x (1/2)^{5-x}$

$P(x) = 5C_x (1/2)^5$

(i) 3 boys. $P(X=3) = 5C_3 (1/2)^5$

$= \frac{5}{16}$ for family

$= 800 \times \frac{5}{16} = 250$ families

Thus for 800 families the probability of no. of families having 3 boys.

(ii) 5 girls

$P(X=0) = 5C_0 (1/2)^5$

$= 1/32 = 0.03125$ per family

800 families $= 800 \times \frac{1}{32} = 25$ families

(iii) either 2 or 3 boys

$= P(X=2) + P(X=3)$

$= 5C_2 (1/2)^2 (1/2)^3 + 5C_3 (1/2)^3 (1/2)^2$

$= (5/16) + (5/16) = \frac{5}{8}$ per fam

800 family $800 \times \frac{5}{8} = 500$ family

Scanned by CamScanner

(iv) atleast one boy $= P(X \geq 0) = 1 - P(X=0)$

$= 1 - 5C_6 (1/2)^5$

$= 1 - 1/32 = 31/32 \times 800$

$= 775$

(2) The mean and variance of the binomial distribution is 4 and 4/3. Find $P(X \geq 1)$

$np = 4 \rightarrow (1)$, $npq = 4/3 \rightarrow (2)$

(2) ÷ (1) $\Rightarrow \frac{\cancel{n}\cancel{p}q}{\cancel{n}\cancel{p}} = \frac{\cancel{4}/3}{\cancel{4}}$

$q = 1/3$

$p = 1 - q = 1 - 1/3 = 2/3$

sub p in (1) $np = 4$

$n \times \frac{2}{3} = 4$

$n = 6$

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - nC_x p^x q^{n-x}$

$= 1 - 6C_0 (2/3)^0 (1/3)^{6-0}$

$= 1 - 6C_0 (1/3)^6 = 0.998$

(Q) The mean and variance of binomial distribution 'X' with parameters n and p are 16 and 8. find $P(X \geq 1)$ & $P(X > 2)$

Sol

$$np = 16 \rightarrow (1)$$

$$npq = 8 \rightarrow (2)$$

(2) ÷ (1)

$$\frac{npq}{np} = \frac{8}{16}$$

$$q = 1/2, \quad p = 1 - q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$np = 16$$

$$n\left(\frac{1}{2}\right) = 16$$

$$n = 32$$

$$P(X \geq 1) = 1 - P(X < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - nC_0\left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^{32}$$

$$= 1 - \left(\frac{1}{2}\right)^{32}$$

$$= 0.999$$

$$P(X > 2) = 1 - P(X \leq 2)$$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$= 1 - \left[32C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{32} + 32C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{31} + 32C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{30}\right]$$

$$=$$

⇒ Fit a binomial distribution for the following data

x	0	1	2	3	4	5
f	2	14	20	34	22	8

x	f	$x_i f_i$
0	2	0
1	14	14
2	20	40
3	34	102
4	22	88
5	8	40
	$N = 100$	284

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{284}{100} = 2.84$$

$np = 2.84$

$(5)p = 2.84$

$p = 0.568$ $\qquad q = 1 - p$

$q = 0.432$

$P(x) = nC_x p^x q^{n-x}$ $\quad N(p+q)^n$

$= 100[5C_0 \cdot p^0 (0.432)^5 + 5C_1 (0.568)^1 (0.432)^4 + 5C_2 (0.568)^2 (0.432)^3 +$

$5C_3 (0.568)^3 (0.432)^2 + 5C_4 (0.568)^4 (0.432)^1 +$

$5C_5 (0.568)^5 (0.432)^0]$

$= 100[0.0752 + 0.09891 + 0.2601 + 0.3419 + 0.2248 +$

$0.0591]$

$= 7.59 + 9.891 + 26.01 + 34.19 + 22.42 + 5.91$

∴ Observed frequencies are

x	0	1	2	3	4	5
f	2	14	20	34	22	8
ob f	8	10	26	34	22	6

Poisson Distribution:

A random variable X is said to follow a poisson distribution if it assumes only non-negative values and its probability distribution is given by

$$P(x, \lambda) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Here $\lambda > 0$ is called the parameter of the distribution.

Note:

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} P(x)$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \quad \left(e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \right)$$

$$= e^{-\lambda} \cdot e^{\lambda} = 1$$

∴ 1) This eqn is known as probability of function.

2). The distribution function is

$$F(x) = P(X \le x) = \sum_{r=0}^{x} P(r) = e^{-\lambda} \sum_{r=0}^{x} \frac{\lambda^r}{r!} \cdot \quad x = 0, 1, 2, \ldots$$

Constance of poisson distribution:-

Mean:

$$E(X) = \sum_{x=0}^{\infty} x\,(p(x))$$

$$= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\cancel{x}\,\lambda^{x}}{\cancel{x}\,(x-1)!}$$

$$= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!}$$

Let $x-1=y \Rightarrow x=y+1$

$$= e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^{y+1}}{y!}$$

$$= e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!} \cdot \lambda$$

$$= \lambda e^{-\lambda} \cdot e^{\lambda}$$

$$= \lambda$$

$\therefore$ This parameter is called the arithmetic mean of the poisson distribution.

Note: Sometime $\lambda = np$

Variance of a poisson distribution

$$V(x) = E(X^2) - (E(X))^2$$

$$= \sum_{x=0}^{\infty} x^2 p(x) - \lambda^2$$

$$= \sum_{x=0}^{\infty} \frac{x^2 e^{-\lambda}\lambda^{x}}{x!} - \lambda^2$$

$$= \sum_{x=0}^{\infty} \frac{x^2 e^{-\lambda}\lambda^{x}}{x(x-1)!} - \lambda^2$$

$$= \sum_{x=0}^{\infty} \left((x-1)+1\right) \frac{e^{-\lambda}\cdot \lambda^{x}}{(x-1)!} - \lambda^2$$

$$= \sum_{x=1}^{\infty} (x-1)\frac{e^{-\lambda}\cdot\lambda^{x}}{(x-1)!} + \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{(x-1)!} - \lambda^2$$

$$= \sum_{x=1}^{\infty} \frac{(x-1)e^{-\lambda}\cdot\lambda^{x}}{(x-1)(x-2)!} + e^{-\lambda}\cdot\sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!} - \lambda^2$$

$$= \sum_{x=1}^{\infty} \frac{(x-1)e^{-\lambda}\cdot\lambda^{x}}{(x-1)(x-2)!} + e^{-\lambda}\cdot\sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!} - \lambda^2$$

$$= \sum_{x=2}^{\infty} \frac{e^{-\lambda}\cdot\lambda^{x}}{(x-2)!} + e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!} - \lambda^2$$

let $\quad x-2=y \Rightarrow x=y+2$

$\quad x-1=z \Rightarrow x=z+1$

$$= \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda}{y!} + e^{-\lambda}\cdot\sum_{z=0}^{\infty} \frac{\lambda^{z+1}}{z!} - \lambda^2$$

$$= e^{-\lambda}\left[\lambda^2 \sum_{y=0}^{\infty} \frac{\lambda^{y}}{y!} + \lambda \sum_{z=0}^{\infty} \frac{\lambda^{z}}{z!}\right] - \lambda^2$$

$$= e^{-\lambda}\left[\lambda^2 e^{\lambda} + \lambda e^{\lambda}\right] - \lambda^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$V(x) = \lambda.$$

Recurrence Relation for Poisson's distribution.

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

$$P(x+1) = \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!} = \frac{e^{-\lambda}\lambda^{x}\cdot\lambda}{(x+1)\,x!}$$

$$= \frac{\lambda}{(x+1)} \cdot \frac{e^{-\lambda}\cdot\lambda^{x}}{x!}$$

$$P(x+1) = \frac{\lambda}{(x+1)}\cdot P(x)$$

$$P(x) = \frac{\lambda}{x}\cdot P(x-1)$$

(Q) If the probability that an individual suffer a bad reaction from a certain injection is 0.001. Determine the probability that out of 2000 individuals

(i) exactly 3, (ii) more than 2 individuals, (iii) none

(iv) more than 1 individual suffer a bad reaction.

sol $p = 0.001$ $n = 2000$

$$\lambda = np$$

$$= 2000 \times 0.001$$

$$\lambda = 2$$

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-2}2^{x}}{x!}$$

1) $P(x=3) = \dfrac{e^{-2}2^{3}}{3!} = 0.1804$

2) $P(x>2) = 1 - \left[\dfrac{e^{-2}2^{0}}{0!} + \dfrac{e^{-2}2^{1}}{1!} + \dfrac{e^{-2}2^{2}}{2!}\right]$

$$= 0.324$$

3) $P(X=0) = \dfrac{e^{-2}2^0}{0!} = e^{-2} = 0.135$

4) $P(X \geq 1) = 1 - [P(X=0) + P(X=1)]$

$= 1 - \left[\dfrac{e^{-2}2^0}{0!} + \dfrac{e^{-2}2^1}{1!}\right]$

$= 0.5939$

(Q) If a random variable as a poisson distribution such that $P(1) = P(2)$. find (i) mean of the distribution (ii) $P(4)$ (iii) $P(X \geq 1)$. (iv) $P(1 < x < 4)$

sol (i) $P(x+1) = \dfrac{\lambda}{x+1} \cdot P(x)$

$x = 1$

$P(2) = \dfrac{\lambda}{2} P(1)$

$\lambda = \dfrac{2P(2)}{P(1)}$

mean $\lambda = 2$

2) $P(X=4) = \dfrac{e^{-\lambda} \cdot \lambda^4}{4!} = \dfrac{e^{-2} 2^4}{4!} = 0.09$

3) $P(X \geq 1) = 1 - P(X=0)$

$= 1 - \dfrac{e^{-2} 2^0}{0!} = 0.865$

4) $P(1 < x < 4) = P(X=2) + P(X=3)$

$= \dfrac{e^{-\lambda}\lambda^2}{2!} + \dfrac{e^{-\lambda}\lambda^3}{3!}$

$= \dfrac{e^{-2}2^2}{2!} + \dfrac{e^{-2}2^3}{3!}$

$= 0.451$

Moment Generating Function:

The moment generating function of random vari[able] X, about the origin.

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum e^{tx} p(x) & \text{in case of discrete r.v} \\ \int_{-\infty}^{\infty} e^{tx} f_x(x) & \text{in case of continuous r.v} \end{cases}$$

(whose probability density function $f_x(x)$)

is given by

(i) discrete

(ii) continuous.

Moment generating function of BD:

If (X) is a binomial distribution, variable then the moment generating function is $(pe^t+q)^n$

Proof:

We know that, the Binomial density function of density is given by

$$P(X=x) = p(x) = \begin{cases} nC_x p^x q^{n-x} & \cdot f_x(x) \\ 0 & \text{otherwise} \end{cases}$$

$x = 0, 1, 2, \ldots$

A moment generating function is

$$M_x(t) = E(e^{tx})$$

$$= \sum e^{tx}(p(x))$$

$$= \sum e^{tx}\, nC_x p^x q^{n-x}$$

$$= \sum nC_x (e^t p)^x q^{n-x}$$

$$\boxed{\therefore m_x(t) = (pe^t+q)^n}$$

Moment generating function of poisson Distribution:

If (X) is a poisson distribution, variable then its moment generating function is $e^{\lambda(e^t-1)}$ $e^{\lambda(e^t-1)}$

Proof: Let X be a random variable having the poisson distribution with probability density function is given by

$$P(X=x) = p(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \forall x(n) \\ & x = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$$

$$M_x(t) = E(e^{tx})$$

$$= \Sigma e^{tx} p(x)$$

$$= \Sigma e^{tn} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \Sigma \frac{e^{tx}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \left(\frac{e^t\lambda}{x!}\right)^x$$

$$= e^{-\lambda}\left(1 + \frac{e^t\lambda}{1!} + \left(\frac{e^t\lambda}{2!}\right)^2 + \ldots\right)$$

$$= e^{-\lambda}\left(e^{e^t\lambda}\right)$$

$$= e^{-\lambda} e^{e^t\lambda}$$

$$= e^{-\lambda + e^t\lambda}$$

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

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

Geometry Distribution:

A random variable (X) is said to have a geometric distribution if it assumes only non-negative values & probability mass function is given by

$$P(X=x) = p(x) = \begin{cases} q^x p & \text{if } x=0,1,2,\dots \quad 0 \le p \le 1 \\ 0 & \text{otherwise} \quad q = 1-p \end{cases}$$

Mean of G.D:

$$\mu = E(x) = \sum_{x=0}^{\infty} x\,p(x)$$

$$= \sum_{x=0}^{\infty} x q^x p$$

$$= pq \sum_{x=1}^{\infty} x q^{x-1}$$

$$= pq\,(1+2q+3q^2+4q^3+5q^4+\dots)$$

$$= pq\,(1-q)^{-2}$$

$$= \frac{pq}{(1-q)^2} = \frac{\cancel{p}q}{p^{\cancel{2}}}$$

$$\boxed{\mu = \frac{q}{p}}$$

Variance:

$$V(x) = E(x^2) - (E(x))^2$$

$$= \sum_{x=0}^{\infty} x^2 p(x) - (E(x))^2$$

$$= \sum_{x=0}^{\infty} x^2 q^x p - \left(\frac{q}{p}\right)^2$$

$$= \sum_{x=0}^{\infty} x(x-1)+1)\, q^x p - \frac{q^2}{p^2}$$

$$= \sum_{x=2}^{\infty} x(x-1) q^x p + \sum_{x=0}^{\infty} x q^x p - \frac{q^2}{p^2}$$

$$= 2pq^2 \sum_{x=2}^{\infty} \frac{x(x-1)}{2} q^{x-2} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= 2pq^{2}\left[1 + 3q + 6q^2 + 10q^3 + \cdots\right] + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= 2pq^2 (1-q)^{-3} + \frac{q}{p} - \frac{q^2}{p^2},$$

$$= \frac{2pq^2}{(1-q)^3} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= \frac{2pq^2}{p^3} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2}$$

$$= \frac{q}{p}\left(1 + \frac{q}{p}\right)$$

$$= \frac{q}{p}\left(\frac{p+q}{p}\right) = \frac{q}{p}\left(\frac{1}{p}\right) = \frac{q}{p^2}$$

Negative Binomial Distribution:

A random variable is to negative BD. If parameter r and p is pb mass function.

$$P(X=x) = p(x) = \begin{cases} {}^{x+r-1}C_{r-1}\, p^r q^x & x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$\text{mean} = rq/p$$

$$\text{valience} = \frac{rq}{p^2}$$

Scanned by CamScanner

Continuous probability distribution:

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

2) Variance $= V(x) = \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$

3) Standard deviation $(\sigma) = \sqrt{V(x)}$

Q) If a random variable of a probability density distribution as

$$f(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Find the probability that it will take a value b/w

i) 1 & 3 (ii) greater than 0.5

Sol: The probability that take the values b/w 1 and 3

$$P(1 \leq x \leq 3) = \int_{-\infty}^{\infty} f(x)\, dx$$

$$= \int_{1}^{3} f(x)\, dx$$

$$= \int_{1}^{3} 2e^{-2x}\, dx$$

$$= 2\left[\frac{e^{-2x}}{-2}\right]_1^3$$

$$= -\left[e^{-2x}\right]_1^3$$

$$= -\left[e^{-6} - e^{-2}\right]$$

$$= e^{-2} - e^{-6}$$

$$= 0.135 - 0.0024$$

$$= 0.1326$$

(ii) $P(x) \geq 0.5 = \int_{0.5}^{\infty} f(x)\,dx$

$$= \int_{0.5}^{\infty} 2e^{-2x}\,dx$$

$$= 2\left[\frac{e^{-2x}}{-2}\right]_{0.5}^{\infty} = -\left[e^{-2x}\right]_{0.5}^{\infty}$$

$$= \left[e^{-\infty} - e^{-2\times 0.5}\right]$$

$$= \left[e^{-1} - 0\right]$$

$$= 1/e = 0.367.$$

Q) probability density function of a random variable X is

$$f(x) = \begin{cases} \frac{1}{2}\sin x & \text{for } 0 \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

i) Find mean and variance of the probability function.

sol $f(x) = \frac{1}{2}\sin x$

Mean: $\mu = \int_{-\infty}^{\infty} x f(x)\,dx$

$$\mu = \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{\pi} x f(x)\,dx + \int_{\pi}^{\infty} x f(x)\,dx$$

$$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{\pi} x \cdot \frac{1}{2}\sin x\,dx + \int_{\pi}^{\infty} 0\,dx$$

$$\left[\because \int uv = u\int v\,dx - \left[\int \frac{d}{dx}u \cdot \int v\,dx\right]dx\right]$$

$$= \frac{1}{2}\int_{0}^{\pi} x \sin x\,dx$$

$$= \frac{1}{2}\left[-x\cos x - \int -\cos x\,dx\right]_{0}^{\pi}$$

$$= \frac{1}{2}\left[-x\cos x + \sin x\right]_{0}^{\pi}$$

Scanned by CamScanner

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

$$= \frac{1}{2}\left[-\pi\cos\pi + \sin\pi + 0\cos 0 - \sin 0\right]$$

$$\mu = \frac{1}{2}[\pi] = \pi/2$$

Variance:

$$v(x) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$= \int_{-\infty}^{0} x^2 f(x)\,dx + \int_{0}^{\pi} x^2 f(x)\,dx + \int_{\pi}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$= \int_{-\infty}^{0} 0\cdot dx + \int_{0}^{\pi} x^2 \frac{1}{2}\sin x\,dx + \int_{\pi}^{\infty} 0\,dx - \mu^2$$

$$= \int_{0}^{\pi} x^2 \frac{1}{2}\sin x\,dx - \mu^2$$

$\int uv\,dx = uv_1 - u'v_2 + u''v_3 - \ldots$

$v_1 = \int v\,dx$

$u' = \frac{du}{dx}$

$$= \frac{1}{2}\int_0^{\pi} x^2 \sin x\,dx - \mu^2$$

$$= \frac{1}{2}\left[x^2[-\cos x] - \int 2x(-\cos x)\,dx\right]_0^{\pi} - \mu^2$$

$$= \frac{1}{2}\left[\left[-x^2\cos x + 2\left[x(\sin x) - \int \sin x\,dx\right]_0^{\pi}\right]_0^{\pi}\right] - \mu^2$$

$$= \frac{1}{2}\left[-x^2\cos x + 2x\sin x + 2\cos x\right]_0^{\pi} - \mu^2$$

$$= \frac{1}{2}\left[-\pi^2\cos\pi + 2\pi\sin\pi + 2\cos\pi + 0\cos(0) - 2(0)\sin(0) - 2\cos(0)\right] - \mu^2$$

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

$$= \frac{\pi^2 - 4}{2} - \left(\frac{\pi}{2}\right)^2$$

$$= \frac{\pi^2 - 4}{2} - \frac{\pi^2}{4}$$

$$= \frac{\pi^2}{2} - 2 - \frac{\pi^2}{4} \qquad = \frac{\pi^2}{4} - 2.$$

Q) A continuous random variable has the probability density function

$$f(x) = \begin{cases} kxe^{-\lambda x} & \text{for } x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) Determine k

(ii) Mean (iii) Variance

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

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

$$\int_{-\infty}^{\infty} kxe^{-\lambda x}\,dx = 1$$

$$k\int_{-\infty}^{\infty} xe^{-\lambda x}\,dx = 1 \Rightarrow k\int_{-\infty}^{0} xe^{-\lambda x}\,dx + k\int_{0}^{\infty} xe^{-\lambda x}\,dx = 1$$

$$k\left[\frac{xe^{-\lambda x}}{-\lambda} - \int 1\cdot\frac{e^{-\lambda x}}{-\lambda}\,dx\right]_0^{\infty} = 1$$

$$k\left[\frac{xe^{-\lambda x}}{-\lambda} - \frac{e^{-\lambda x}}{\lambda^2}\right]_0^{\infty} = 1$$

$$k\left[\frac{\infty\cdot e^{-\lambda(\infty)}}{-\lambda} - \frac{e^{-\lambda(\infty)}}{\lambda^2} - \frac{0\cdot e^{-\lambda(0)}}{-\lambda} + \frac{e^{-\lambda(0)}}{\lambda^2}\right] = 1$$

$$\Rightarrow k\left[0 - 0 - \left(0 - \frac{1}{\lambda^2}\right)\right] = 1$$

$$k\left(\frac{1}{\lambda^2}\right) = 1$$

$$\boxed{k = \lambda^2}$$

$$\therefore f(x) = \begin{cases} \lambda^2 xe^{-\lambda x} & \text{for } x \geq 0, \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

Scanned by CamScanner

$$\mu = \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{\infty} x f(x)\,dx$$

$$= 0 + \int_{0}^{\infty} \lambda^2 x^2 e^{-\lambda x}\,dx$$

$$= \lambda^2 \left[\frac{x^2 e^{-\lambda x}}{\lambda} - \frac{2x}{\lambda^2} e^{-\lambda x} - \frac{2e^{-\lambda x}}{\lambda^3} \right]_{0}^{\infty}$$

$$= \lambda^2 \left(\frac{2}{\lambda^3} \right) = 2/\lambda$$

$=$

Q) If 'x' is a continuous random variable & $Y = ax + b$. prove that $E(Y) = aE(X) + b$

$V(Y) = a^2 V(x)$

where V stands for variance a, b are constants.

Sol

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

$$E(Y) = \int_{-\infty}^{\infty} y f(x)\,dx$$

$$E(ax+b) = \int_{-\infty}^{\infty} (ax+b) f(x)\,dx$$

$$= \int_{-\infty}^{\infty} ax f(x)\,dx + \int_{-\infty}^{\infty} b f(x)\,dx$$

$$= a \int_{-\infty}^{\infty} x f(x)\,dx + b \int_{-\infty}^{\infty} f(x)\,dx$$

$$= aE(X) + b(1)$$

$$= aE(X) + b$$

2) $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \rightarrow ①$ $\quad \int_{-\infty}^{\infty} f(x)\,dx = 1 \rightarrow ②$

$V(x) = \int_{-\infty}^{\infty} x^2 f(x) - (E(x))^2 \rightarrow ③$

Scanned by CamScanner

$$v(Y) = (ax+b) = \int_{-\infty}^{\infty} (ax+b)^2 f(x)\,dx - (E(ax+b))^2$$

But $\boxed{E(ax+b) = a\,E(X)+b.}$

$$= \int_{-\infty}^{\infty} (a^2x^2 + 2abx + b^2) f(x)\,dx - (aE(X)+b)^2$$

$$= a^2 \int_{-\infty}^{\infty} x^2 f(x)\,dx + 2ab \int_{-\infty}^{\infty} x f(x)\,dx + b \int_{-\infty}^{\infty} f(x)\,dx$$

$$- \left(a^2 (E(x))^2 + 2ab\,E(x) + b^2\right)$$

from equ ① & ②

$$= a^2 \int_{-\infty}^{\infty} x^2 f(x)\,dx + 2ab\,E(X) + b^2 - a^2 (E(X))^2 - 2ab\,E(X) - b^2$$

from equ ③

$$= a^2 \left[\int_{-\infty}^{\infty} x^2 f(x)\,dx \, (E(X))^2 \right]$$

$$v(Y) = a^2 v(X).$$

5) If 'X' is a continuous random variable then prove that (1) $v(X+k) = v(x)$

(2) $v(kx) = k^2 v(x)$

sol: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \longrightarrow ①$ $\quad \int_{-\infty}^{\infty} f(x)\,dx = 1 \longrightarrow ②$

$$v(x) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - (E(X))^2 \rightarrow ③$$

$$v(X+K) = \int_{-\infty}^{\infty} (x+k)^2 f(x)\,dx - (E(X+K))^2$$

$\boxed{\text{But } E(aX+b) = (a\,E(X)+b)}$

Scanned by CamScanner

$$= \int_{-\infty}^{\infty} (x^2 + 2kx + k^2) f(x) dx - (E(X) + k)^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx + 2k \int_{-\infty}^{\infty} x f(x) dx + k^2 \int_{-\infty}^{\infty} f(x) dx - (E(X))^2 - 2kE(X) - k^2$$

from ① & ②

$$= \int_{-\infty}^{\infty} x^2 f(x) dx + \cancel{2kE(X)} + \cancel{k^2} - (E(X))^2 - \cancel{2kE(X)} - \cancel{k^2}$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2$$

$$= V(x)$$

2) $V(kx) = k^2 V(x)$

$$V(x) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2$$

$$V(kx) = \int_{-\infty}^{\infty} k^2 x^2 f(x) dx - (E(X))^2$$

$$E(kx) = k E(X)$$

$$V(kx) = k^2 \int_{-\infty}^{\infty} x^2 f(x) dx - k^2 (E(X))^2$$

$$= k^2 \left[\underbrace{\int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2}_{V(x)} \right]$$

$$= k^2 V(x)$$

Constants of Normal Distributions Mean :-

Consider the Normal distribution mean (μ), standard deviation (σ) are parameters.

$$f(x) = f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{\mu - x}{\sigma}\right)^2}$$

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

$$= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\left(\frac{\mu - x}{\sigma}\right)^2} \cdot dx$$

Let $\frac{\mu - x}{\sigma} = y \Rightarrow x = \mu - \sigma y$

$$-\frac{dx}{\sigma} = dy \Rightarrow dx = -\sigma dy$$

Let $x = -\infty$ $\quad y = \infty$

$x = \infty$ $\quad y = -\infty$

$$= \int_{\infty}^{-\infty} (\mu - \sigma y) \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \cdot e^{-1/2\, y^2} (-\cancel{\sigma} dy)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu - \sigma y) \cdot e^{-y^2/2}\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \mu e^{-y^2/2} - \sigma \int_{-\infty}^{\infty} y e^{-y^2/2} dy \right]$$

$$= \frac{1}{\sqrt{2\pi}} \left[2 \int_{0}^{\infty} \mu e^{-y^2/2} dy - 0 \right]$$

$$\left[\int_{-a}^{a} f(x)\, dx = 2\int_0^a f(x)\, dx \text{ even}; \quad 0 \text{ odd} \right]$$

$f(-x) = f(x)$ even

$f(-x) = -f(x)$ odd

$\Rightarrow y^2/2 = z \Rightarrow y = \sqrt{2z}$

$$2y/2\, dy = dz$$

$$y\, dz = dz$$

$y=0 \Rightarrow z=0$

$y=\infty \Rightarrow z=\infty$

$$\frac{1}{\sqrt{2}\sqrt{\pi}}\int_0^\infty \mu \cdot e^{-z}\,\frac{dz}{y}$$

$$\frac{2}{\sqrt{2}\sqrt{\pi}}\cdot\mu\int_0^\infty e^{-z}\frac{dz}{\sqrt{2z}}$$

$$= \frac{2\mu}{2\sqrt{\pi}}\cdot\int_0^\infty e^{-z} z^{-1/2}\,dz$$

$$= \frac{\mu}{\sqrt{\pi}}\cdot\int_0^\infty e^{-z} z^{1/2-1}\,dz \qquad \left[P(n) = \int_0^\infty e^{-x} x^{n-1}\,dx\right]$$

$$= \frac{\mu}{\sqrt{\pi}}\cdot P(1/2) \qquad \left(P(1/2)=\sqrt{\pi}\right)$$

$$= \frac{\mu}{\sqrt{\pi}}(\sqrt{\pi})$$

$$\boxed{\text{Mean} = \mu}$$

Variance of Normal distribution

$$f(x) = f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-1/2\left(\frac{\mu-x}{\gamma}\right)^2}$$

$$\left(\frac{x-\mu}{\sigma}\right)^2$$

$$V(E) = E(x^2) - (E(x))^2$$

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

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\left(\frac{\mu-x}{\sigma}\right)^2} dx$$

Let $\frac{\mu-x}{\sigma} = y \Rightarrow x = \mu - \sigma y$

$dx = -\sigma\,dy$

limits $x=\infty \Rightarrow y=-\infty$

$x=-\infty \Rightarrow y=\infty$

$$= \int_{-\infty}^{\infty} (\mu - \sigma y \cdot \mathcal{Q} - \mu)^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} y^2} \cdot (-\sigma dy)$$

$$= \int_{-\infty}^{\infty} \sigma^2 y^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} y^2} dy$$

$$\left[\begin{array}{c} f(-y) = f(y) \\ \int_{-\infty}^{\infty} f(x) dx = 2 \int_0^{\infty} f(x) dy \end{array} \right]$$

$$= 2 \int_0^{\infty} \sigma^2 y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot y^2} dy$$

let $\frac{y^2}{2} = z \Rightarrow y = \sqrt{2z}$

$$dy = \frac{1}{\sqrt{z}} dz$$

$$y = 0 \Rightarrow z = 0$$

$$y = \infty \Rightarrow z = \infty$$

$$= 2\sigma^2 \int_0^{\infty} z \cdot \frac{1}{\sqrt{2}\sqrt{\pi}} \cdot e^{-z} \cdot \frac{1}{\sqrt{2}\sqrt{z}} dz$$

$$= \sigma^2 \int_0^{\infty} \frac{\sqrt{z}}{\sqrt{\pi}} \cdot e^{-z} dz$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \cdot \int_0^{\infty} e^{-z} z^{1/2} dz$$

$$\left[\gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx \right]$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \cdot \int_0^{\infty} e^{-z} \cdot z^{3/2 - 1} dz$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \cdot \gamma\left(\frac{3}{2}\right)$$

$$\boxed{\gamma(n+1) = n\gamma(n)}$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \gamma\left(\frac{1}{2}+1\right)$$

$$= 2 \frac{\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \gamma\left(\frac{1}{2}\right)$$

$$= \cancel{2} \frac{\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \frac{1}{\cancel{2}} \cancel{\sqrt{\pi}} = \sigma^2$$

Median of normal distribution:

If 'M' is Median of normal distribution is

$$\int_{-\infty}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2} = \frac{1}{2} = \int_{-\infty}^{M} f(x)\,dx.$$

$$f(x) = f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2}$$

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

$$\int_{-\infty}^{\mu} f(x)\,dx + \int_{\mu}^{M} f(x)\,dx = \frac{1}{2} \longrightarrow (1)$$

Case(i):

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

$$\int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2} dx$$

let $\frac{\mu-x}{\sigma} = y$

$x = \mu - \sigma y$

$dx = -\sigma dy$

but $x = -\infty \Rightarrow y = \infty$

$x = \mu \Rightarrow y = 0$

$$= \int_{\infty}^{0} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-y^2/2} (-\sigma dy)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-y^2/2} dy$$

let $\frac{y^2}{2} = z$

$y = \sqrt{2}\sqrt{z}$

$$dy = \frac{\sqrt{2}}{2\sqrt{z}}\,dz$$

$y=0 \rightarrow z=0$

$y=\infty \rightarrow z=\infty$

$$= \frac{1}{\sqrt{2}\sqrt{\pi}} \cdot \int_0^\infty e^{-z} \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{z}}\,dz$$

$$= \frac{1}{2\sqrt{\pi}} \cdot \int_0^\infty e^{-z} \cdot z^{-1/2}\,dz$$

$$= \frac{1}{2\sqrt{\pi}} \int_0^\infty e^{-z} \cdot z^{1/2 - 1}\,dz \qquad \left[\gamma(n) = \int_0^\infty e^{-x} x^{n-1}\,dx\right]$$

$$= \frac{1}{2\sqrt{\pi}} P(1/2)$$

$$= \frac{1}{2\sqrt{\pi}} (\sqrt{\pi})$$

$$\int_{-\infty}^{\mu} f(x) \cdot dx = 1/2 \longrightarrow 2$$

sub equ 2 in 1

$$\int_{-\infty}^{\mu} f(x)\,dx + \int_{\mu}^{M} f(x)\,dx = 1/2$$

$$\Rightarrow \int_{-\infty}^{\mu} f(x)\,dx + \int_{\mu}^{M} f(x)\,dx = 1/2$$

$$\Rightarrow \frac{1}{2} + \int_{\mu}^{M} f(x)\,dx = 1/2$$

$$\Rightarrow \int_{\mu}^{M} f(x)\,dx = 0$$

$$\Rightarrow \int_{\mu}^{M} \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \left(\frac{\mu - x}{\sigma}\right)^2}\,dx = 0$$

$$\mu = M \qquad \left[\because \int_a^b f(x)\,dx = 0 \quad a = b\right]$$

Mean = Median

Mode of Normal distribution:

Mode is a value of x for which $f(x)$ is maximum

(Mode is a solution) $f'(x)=0$ & $f''(x)<0$

$$f(x) = f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{\mu-x}{\sigma}\right)^2}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\,\sigma^2(\mu-x)^2}$$

$$f'(x) = 0$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\sigma^2(\mu-x)^2}\left(-\frac{2(\mu-x)(-1)}{2\sigma^2}\right) = 0$$

$$f'(x) \Rightarrow \frac{f(x)}{\sigma^2}(\mu-x) = 0$$

$$\mu - x = 0$$

$$\boxed{x = \mu}$$

$$f'(x) = \frac{1}{\sigma^2} f(x)(\mu-x)$$

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

$$f''(x) = \frac{1}{\sigma^2}\left[f'(x).(\mu-x) - f(\mu)\right]$$

$$f''(x) = \frac{1}{\sigma^2}\left[f'(x)(\mu-x) - f(\mu)\right]$$

$$= \frac{1}{\sigma^2}\left[0 - f(\mu)\right]$$

$$= \frac{1}{\sigma^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} < 0$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}} < 0$$

μ = Mode

Median = Mode

Hence proved

$$\begin{bmatrix} f''(x) < 0 & \text{maxima} \\ f''(x) > 0 & \text{minima} \end{bmatrix}$$

→ Mean deviation from Mean of Normal distribution:

$$\int_{-\infty}^{\infty} |x-\mu| \cdot f(x)\, dx$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$= \int_{-\infty}^{\infty} |(x-\mu)| \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $\frac{\mu-x}{\sigma} = y$

$x = \mu - \sigma y$

$dx = -\sigma dy$

$x = -\infty \Rightarrow y = \infty$

$x = \infty \Rightarrow y = -\infty$

$$\int_{\infty}^{-\infty} \sigma y \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 y^2} (\sigma dy)$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (y) e^{-1/2 y^2} dy$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \cdot \int_{0}^{\infty} (y) e^{-y^2/2} \cdot dy$$

$\Rightarrow \frac{y^2}{2} = z$

$y^2 = \sqrt{2}\sqrt{z}$

$dy = \frac{\sqrt{2}}{2\sqrt{z}} dt$

$y = 0 \Rightarrow z = 0$

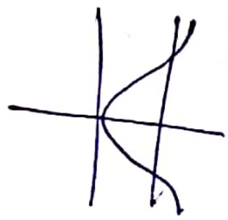
$y = \infty \Rightarrow z = \infty$

$$\Rightarrow \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} \sqrt{2}\sqrt{z} e^{-z} \frac{\sqrt{2}}{2\sqrt{z}} dz$$

Characteristic of Normal distribution (properties) (or) Importance

1) The curve is bell-shaped and symmetric about the line $x = \mu$

2) Mean, Median and mode of distribution coincide



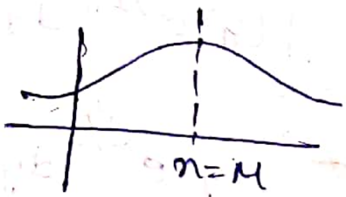
3) As x increases numerically, $f(x)$ decreases rapidly.

4) The maximum probability occurs at the point $x = \mu$ and is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}\left(\frac{\mu - x}{\sigma}\right)^2$$

$$\left[P(x)\right]_{max} = \frac{1}{\sigma\sqrt{2\pi}}$$

⇒ x-axis is an asympton



⇒ Linear combination of independent normal varient is also a normal varient.

⇒ The points of intersection of curves on $x = \mu \pm \sigma$

⇒ The probability that the normal varience occurence & standard variation σ between x_1 and x_2 is given by

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}\left(\frac{\mu - x}{\sigma}\right)^2 dx$$

⇒ Normal distribution find large applications and statistical quality control.

⇒ since, Normal distribution is limiting case of binomial distribution for expectational values of large numbers.

⇒ Normal distribution is also used in t-test, F-test, χ^2 distribution.

⇒ It is applicable to many applied problems in kinetic theory of gases & fluctuations in the magnitude of an electric current.

Q) Find the probability density of Normal curves

1) perform the change of scale $z = \frac{x-\mu}{\sigma}$ & find z_1 and z_2 corresponding to the values of x_1 and x_2 respectively.

2) to find $p(x_1 \leq x \leq x_2) = P(z_1 \leq Z \leq z_2)$ if both z_1 & z_2 both positive or both negative. then

$$P(x_1 \leq x \leq x_2) = |A(z_2) - A(z_1)|$$

⇒ If one is positive & one is negative $(z_1 < 0$ & $z_2 > 0)$ then $p(x_1 \leq x \leq x_2) = |A(z_2) + A(z_1)|$

⇒ If only one value is given (+ve or -ve) then

$$P(Z \geq z_1) = 0.5 - A(z_1)$$

Q) For a normally distributed varient with mean & standard duration 1 & SD 3. Find the probability that

(i) $3.43 \le x \le 6.19$, (ii) $-1.43 \le x \le 6.19$

Sol (i) $\mu = 1$, $\sigma = 3$

$3.43 \le x \le 6.19$

$x_1 = 3.43$, $x_2 = 6.19$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{3.43 - 1}{3} \qquad z_2 = \frac{x_2 - \mu}{\sigma} = \frac{6.19 - 1}{3}$$

$$= 0.31 \qquad = 1.73$$

So, probability of $P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$

$$= |A(z_2) - A(z_1)|$$

$$= |A(1.73) - A(0.8)|$$

$$= |0.4582 - 0.2910|$$

$$= 0.1672$$

(ii) $-1.43 \le x \le 6.19$

$x_1 = -1.43$, $x_2 = 6.19$

$$z_1 = \frac{x_1 - \mu}{\sigma} \qquad z_2 = \frac{x_2 - \mu}{\sigma}$$

$$= \frac{-1.43 - 1}{3} = -0.81 \qquad = \frac{6.19 - 1}{3} = 1.73$$

$P(x_1 \le x \le x_2) = P(z_1 \le z \le z_2)$

$$= |A(z_2) + A(z_1)|$$

$$= |A(1.73) + A(-0.81)|$$

$$= |0.4582 + 0.2910|$$

$$= \cancel{0.1672}$$

$$0.7492$$

$A(-z) = A(z)$

Scanned by CamScanner

Q) A normal distribution 7% of the items are under 35 and 89% are under 63. Find mean and variance of the distribution

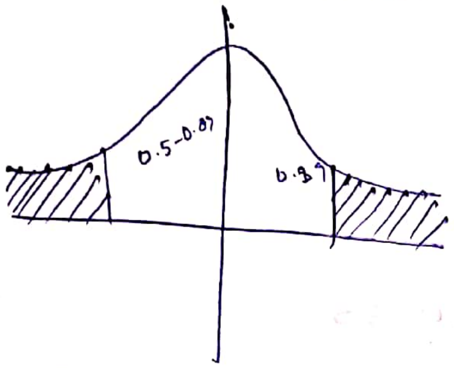
Sol

$P(X < 35) = -0.07$

$P(X \leq 63) = 0.89 \Rightarrow P(X = 63) = 1 - P(X \leq 63)$

$= 1 - 0.89$

$P(X > 65) = 0.11$



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

$$\frac{x_1 - \mu}{\sigma} = z_1$$

$$\frac{35 - \mu}{\sigma} = z_1 \rightarrow (1)$$

$A(z_1) = -0.43 \Rightarrow z_1 = 1.48$

$A(z_2) = 0.89 \Rightarrow z_2 = 1.23$

$$\frac{x_2 - \mu}{\sigma} = z_2$$

$$\frac{63 - \mu}{\sigma} = z_2 \rightarrow (2)$$

Sub z_1 and z_2 in eqn (1) & (2)

$$\frac{35 - \mu}{\sigma} = -148 \rightarrow (3)$$

$$\frac{63 - \mu}{\sigma} = 1.28 \rightarrow (4)$$

Scanned by CamScanner

(4) - (3)

$$\frac{63}{\sigma} - \frac{\mu}{\sigma} - \frac{35}{\sigma} + \frac{\mu}{\sigma} = 1.23 + 1.48$$

$$\frac{63-35}{\sigma} = 2.71$$

$$\sigma = \frac{63-35}{2.71}$$

$$\sigma = \frac{28}{2.71} = 10.33$$

sub 'σ' in equ (3)

$$\frac{35-\mu}{\sigma} = -1.48$$

$$35 - \mu = -1.48 \times 10.33$$

$$\mu = 35 + 1.48 \times 10.33$$

$$\mu = 50.29 \simeq 50.3$$

2) X is a normal varient mean 30, SD - 5, find the probabilities that (1) $26 \leq x \leq 40$ (2) $x \geq 45$

sol $x_1 = 26$, $x_2 = 40$

$$z_1 = \frac{x_1 - \mu}{\sigma} \qquad z_2 = \frac{x_2 - \mu}{\sigma}$$

$$= \frac{26-30}{5} \qquad = \frac{40-30}{5}$$

$$z_1 = -0.6 \qquad z_2 = 2$$

$$P(x_1 \leq x \leq x_2) = P(z_1 \leq z \leq z_2)$$

$$P(26 \leq x \leq 40) = P(-0.8 \leq z \leq 2)$$

$$\Rightarrow |A(z_1) + A(z_2)|$$

$$\Rightarrow |A(0.8) + A(2)|$$

$$\Rightarrow 0.2881 + 0.4772$$

$$\Rightarrow 0.7653$$

Scanned by CamScanner

② $x_1 \leq 45$

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

$$= \frac{45-30}{5}$$

$$z_1 = 3$$

$$P(x \geq x_1) = P(z \geq z_1)$$

$$P(x \geq 45) = P(z \geq 3)$$

$$= 0.5 - A(z_1)$$

$$= 0.5 - A(3)$$

$$= 0.5 - 0.49865$$

$$= 0.00135$$

Q) A normal distribution 35% are under 45 and 5% are over 64. Find mean & variance of distribution.

moment
~~No.~~ ~~Event~~ Generating function of Normal distribution.

$$f(e^{xt})$$

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

$$\int_{-\infty}^{\infty} e^{xt} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\frac{x-\mu}{\sigma} = y$$

$$x = \mu + \sigma y$$

$$dx = \sigma dy$$

$$x = -\infty \rightarrow y = -\infty$$
$$x = \infty \rightarrow y = \infty$$

$$\int_{-\infty}^{\infty} e^{(\mu+\sigma y)t} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2 y^2} (\sigma dy)$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu t} \cdot e^{\sigma y t} \cdot e^{-y^2/2} dy$$

$$= \frac{e^{-\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2(y^2 - 2\sigma y t)} dy$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-1/2(y^2 - 2\sigma y + (\sigma t)^2 - (\sigma t)^2)} dy$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-1/2}\left[(y-\sigma t)^2 - \sigma^2 t^2\right] dy$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-1/2}(y-\sigma t)^2 \, e^{\frac{\sigma^2 t^2}{2}} dy$$

$$= \frac{e^{\mu t} \cdot e^{\frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2(y-\sigma t)^2} dy$$

Scanned by CamScanner

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-1/2 (y - \sigma t)^2} dy$$

$$y - \sigma t = z$$

$$y = z + \sigma t$$

$$dy = dt$$

$y = -\infty \qquad z = -\infty$

$y = \infty \qquad z = \infty$

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{z}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz$$

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \, 2 \int_{0}^{\infty} e^{-z^2/2} dt$$

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \cdot 2 \cdot \sqrt{\frac{\pi}{2}}$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Exponential distribution:

A random variable 'x' is said to be an exponential distribution with parameter $\theta > 0$, if its probability density function is given by

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

$$\text{Mean} = 1/\theta$$

$$\text{variance} = 1/\theta^2$$

Scanned by CamScanner

β distribution of first kind:

A random variable 'x' is said to have a β distribution of 1^{st} with parameters u & v $(u>0, v>0)$ if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta(u,v)} x^{u-1}(1-x)^{v-1} ; & u, v > 0 \\ 0 & \text{otherwise} \end{cases}$$

β distribution of 2^{nd} kind:

A random variable 'x' is said to have a β distribution of 2^{nd} kind with parameters u & v, $u \geq 0$, $v>0$ with its probability density function.

$$f(x) = \begin{cases} \frac{1}{\beta(u,v)} \cdot \frac{x^{u-1}}{(1+x)^{u+v}} ; & u, v > 0 \\ 0 & \text{otherwise.} \end{cases}$$

γ distribution

A random variable x is said to have a γ distribution with parameters $\lambda > 0$ if its probability density function is given by

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

Scanned by CamScanner

Module - III

Sampling Distributions

Introduction:

The total outcomes of a statistical experiments may be recorded either as a numerical value or a discriptive representation.

Eg: In real life we may have to take some samples to get the information about the population.

population (universe): It is defined as a large collection of individuals or attributes or numerical data. It is also known as a population (or) the collection of objects.

→ A population may be finite or infinite.

→ A population contains a finite no. of units then it is called finite population.

ex: No. of students in a college.

→ If the population contains an infinite no. of units then it is called infinite population.

ex: stars in the sky

$x_1, x_2, \ldots, x_n$ are the population units.

SAMPLE:

A finite subset of statistical individuals in a population is called a sample (or) subset.

ex: blood testing.

The no. of objects in the sample is called sample set.

Types of Samples:

1) purposive sampling: If the sample element are selected with a definite purposive in mind, then the Sample selected is called purposive sampling

Ex: Blood testing

2) Random sampling:

A sample is said to be Random sample if ea elements of it has an equal chance of designing included in the sample (probability sampling)

Eg: selecting randomly 20 words from a dictionary is a random sample

3) simple sampling:

It is a random sampling in which each element of the population has an equal and independent chance of begining included

Ex: for an infinite population, any random sampling is simple.

4) ~~Statisfied~~ stratified sampling:

The sample which is the aggregate of the sample units of each of the leveling series is called the satisfied sampling. It is a heterogenous (population may be divided into sets population of struta)

The population is first sub-divided into several parts called strata which are more heterogenous with in itself. Then the entire population.

Ex: the people in a small village may be divided into 2 strata wise (a) literate (b) illiterate. After dividing the population into strata we can select individuals at random from each of stratum.

Classification of Sampling:

Samples are classified into 2 ways.

1. Large sample: If the size of the sample $n \geq 30$, a sample is said to be large sample.
2. Small sample: If the size of the sample $n < 30$, a sample is said to be a small sample.
3. Statistic sample: A statistical measures computed from the sample observation is known as statistics

→ let $x_1, x_2, ---x_n$ are n sample observations mean $(\bar{x})$ and S.D [S] are known as statistics

Sample mean:

If $x_1, x_2, x_3, --x_n$ represents the random sample of size n, then the sample mean is defined by the statistics

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Scanned by CamScanner

Sample variance: If $x_1, x_2, --- x_n$ represents a random sample of size n then the sample variable is defined by the statisty

$$\boxed{s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}}$$

	Population	sample
Mean	μ	$\bar{x}$
variance	σ^2	s^2
SD	σ	s

→ population correction factor:- $\frac{N-n}{N-1}$ where N is the size of the finite population, n is the size of the sample.

→ Finite population of sample size (without replacement):- NC_n

→ Infinite population of sample size (with replacement) :- N^n

Standard Errors

standard deviation of statistics is known as standard error. The standard error gives some idea about the precision of the estimate. As the sample size 'n' increases standard error decreases. Standard error plays a very important role in large sample decision theory and forms the bases in testing of hypothesis.

⇒ standard error of sample space is $\bar{x} = \frac{\sigma}{\sqrt{n}}$.

Sampling distribution of properties.

Population is infinite than the probability of its occurrences of event (success) is p. while the Probability of non-occurrence of the event is $q = 1 - p$, consider all possible samples of size N.

Sampling with replacement $\sigma^2 p = \frac{pq}{n}$

sampling without replacement $\sigma^2 p = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$

Standardize sample mean $= Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

Population in finite size (sampling without replacement)

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

population is finite & {sampling with replacement}

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Q) What is the value of population correction factor is n=10 and N=1000

$$C.F = \frac{N-n}{N-1}$$

$$= \frac{1000-10}{1000-1}$$

$$= \frac{990}{999} = 0.99$$

Q) How many different samples of size 2 can be choosen from the finite population of size 25.

Sol $^{N}C_{n}$ = N=25, n=2

$$\Rightarrow {}^{25}C_2$$

$$= \frac{25 \times 24}{2 \times 1} = 25 \times 12$$

$$= 300$$

Q) The population consists of 5 members 2, 3, 6, 8, 11 Consider all possible samples of size 2 which can be drawn with replacement from the population. Find

(i) the mean of the population

2) Standard deviation of population

3) mean of the sampling distribution of means

Scanned by CamScanner

4) The SD of sampling distribution of means

Sol) Mean of population is given by

$$\mu = \frac{2+3+6+8+11}{5}$$

$$\mu = 6$$

Standard deviation of the population (σ^2)

$$\sigma^2 = \frac{\Sigma (x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5}$$

$$\sigma^2 = 10.8$$

$$SD(\sigma) = \sqrt{10.8}$$

$$= 3.28$$

(6) $N^n = 5^2 = 25$

(2,2), (2,3), (2,6), (2,8) (2,11)
(3,2), (3,3), (3,6), (3,8), (3,11)
(6,2), (6,3), (6,6), (6,8), (6,11)
(8,2), (8,3), (8,6), (8,8), (8,11)
(11,2), (11,3), (11,6), (11,8), (11,11)

Means of sample (A.M)

$$\begin{Bmatrix} 2 & 2.5 & 4 & 5 & 6.5 \\ 2.5 & 3 & 4.5 & 5.5 & 7 \\ 4 & 4.5 & 6 & 7 & 8.5 \\ 5 & 5.5 & 7 & 8 & 9.5 \\ 6.5 & 7 & 8.5 & 9.5 & 11 \end{Bmatrix}$$

$$\mu_{\bar{x}} = \frac{2+2.5+4+5+6.5+ \cdots + 9.5+11}{25}$$

$$\mu_{\bar{x}} = \frac{150}{25} = 6$$

$$\sigma^2_{\bar{x}} = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2 + (2.5-6)^2 + (4-6)^2 + (5-6)^2 + \cdots + (9.5-6)^2 + (11-6)^2}{25}$$

=

Q) Solve the above problem without replacement

(i) mean

$$\mu = \frac{2+3+6+8+11}{5}$$

$$\mu = 6$$

2) variance

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

$$\sigma = 3.28$$

3) $^NC_n = {^5C_2} = 10$

$$\left\{ \begin{array}{llll} (2,3), & (2,6), & (2,8), & (2,11) \\ & (3,6), & (3,8), & (3,11) \\ & & (6,8), & (6,11) \\ & & & (8,11) \end{array} \right\}$$

Means of sample (AM)

$$\left\{ \begin{array}{llll} 2.5 & 4 & 5 & 6.5 \\ 4.5 & 5.5 & 7 & \\ 7 & 8.5 & & \\ 9.5 & & & \end{array} \right\}$$

Scanned by CamScanner

$$\mu_{\bar{x}} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$\mu_{\bar{x}} = \frac{60}{10} = 6$$

(d) $$\sigma_{\bar{x}}^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$$

$$= \frac{(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+--+(9.5-6)^2}{10}$$

$$\sigma_{\bar{x}}^2 = \frac{40.5}{10}$$

$$\sigma_{\bar{x}} = \sqrt{4.05}$$

$$\sigma = 2.012$$

Q) What is the effect on standard error if a sample is taken from an infinite of population of sample size is increased for 400 to 900

Ans Standard error of mean $= \frac{\sigma}{\sqrt{n}}$

$$n = n_1 = 4000, \quad n_2 = 900$$

$$SE_1 = \frac{\sigma}{\sqrt{n_1}} = \frac{\sigma}{\sqrt{400}} = \frac{\sigma}{20}$$

$$SE_2 = \frac{\sigma}{\sqrt{n_2}} = \frac{\sigma}{\sqrt{900}} = \frac{\sigma}{30}$$

$$=> \frac{1}{2} : \frac{1}{3}$$

$$\frac{SE_1}{SE_2} = \frac{3}{2} \Rightarrow \frac{SE_1}{SE_2} = 3/2$$

Scanned by CamScanner

Q) The mean height of students in a clg is 15.5 cm and standard deviation is 15. What is the probability that the mean height of 36 student that is less than 157 cms

Sol μ = Mean of the population = 155 cm

$\bar{x}$ = mean of sample = 157 cm

SD of population = 15 $\boxed{n=36}$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{157-155}{15/\sqrt{36}}$$

$Z = 0.8$

$Z \le Z_1$

$Z \le 0.8$

$P(Z \le 0.8) \Rightarrow 0.5 - A(0.8)$

$= 0.5 - 0.2881$

$= 0.2119$

Q) The mean of voltage battery is 15 and standard deviation is 0.2. Find probability that four such batteries connected in series will have combined voltage of 60.8 (or) more voltage.

Sol Let A, B, C, D be the mean of voltage of batteries $\bar{x} = 15$

$$\mu_{\bar{x}} = \bar{x}_A + \bar{x}_B + \bar{x}_C + \bar{x}_D$$

$$= 15 + 15 + 15 + 15 = 60$$

$$\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}_A} + \sigma^2_{\bar{x}_B} + \sigma^2_{\bar{x}_C} + \sigma^2_{\bar{x}_D}}$$

$$= \sqrt{4(\sigma_{\bar{x}})^2}$$

$$= \sqrt{4(0.2)^2}$$
$$= \sqrt{4 \times 0.04}$$
$$= \sqrt{0.16}$$
$$\sigma_x = 0.4$$

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} \qquad (x = 60.8,\ n = 1)$$

$$Z = \frac{60.8 - 60}{0.4}$$

$$P(Z \geq z_1)$$

$$P(Z \geq 2) = 0.5 - A(2)$$
$$= 0.5 - 0.477$$
$$= 0.023$$

ESTIMATION

Estimate:

An estimate is a statement to find an unknown population parameters.

Estimator:

The method of determining unknown population parameter is called Estimator.

The estimator can be done in two ways.

(i) point estimator and (ii) interval estimator.

(i) **point estimator:** If a single value is calculated as an estimate from an unknown population parameter

The Procedure to find parameter is called point estimator.

2) Interval Estimator: In General, point estimator does not coincide with true values parameter. So, it is preferred to obtain arrange of values in an interval which the parameter may be consider to lie, then the estimator is called interval estimate of the parameter

Properties of good Estimator:

An estimator is said to be a good estimator if it is. (1) unbiased (2) consistent (3) efficient (4) sufficient.

1) Unbiased estimator:

A statistic $\hat{\theta}$ is said to be an unbiased estimator if and only if the mean of the sampling distribution of estimator is equal to the parameter θ.

$$E(\hat{\theta}) = \theta$$

2) consistent: Any statistic $\bar{\theta}$ (or) $\hat{\theta}$ is said to be consistent if and only if $E(\hat{\theta}) = \theta$ & $V(\hat{\theta}) \Rightarrow 0$ as $n \to \infty$

3) Efficient: Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of population parameter θ and $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$ be the variance of statistics if $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ then $\hat{\theta}_1$ is said to be efficient or more efficient unbiased estimator of θ.

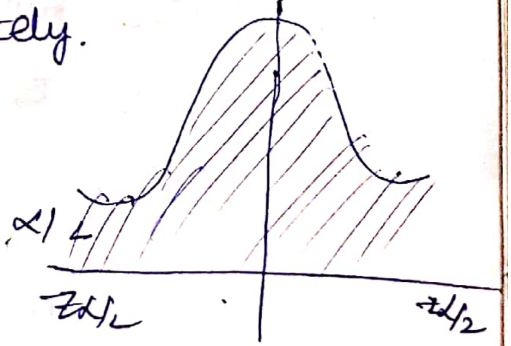
Scanned by CamScanner

a) <u>sufficient:</u>

$T = t(x_1, x_2, \ldots x_n)$ is an estimate of a parameter 'θ', based on a sample $(x_1, x_2, \ldots x_n)$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $(x_1, x_2, \ldots x_n)$ given T is independent of given θ then T is sufficient estimator of θ.

<u>Maximum error of estimate 'e' for large samples:</u>

The sample mean estimate very rarely equals to the mean of population μ. A point estimate is generally acompained with a statements of error which gives difference b/w estimate and the quantity to be estimated the estimator. thus error is $|\bar{x} - \mu|$ for large 'n' the random vaiables $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ is normal variate approximately.



then. $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

$$P\left(\frac{z\alpha}{2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{z_\alpha}{2}\right) = 1 - \alpha$$

$$P\left(-z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \bar{x} - \mu < z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$$

$$P\left(-x - \frac{z_\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right) < -\mu < \frac{z_\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right) - x\right) = 1 - \alpha$$

$$P\left(x - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \frac{z_\alpha}{2}\left(\frac{\sigma}{\sqrt{n}}\right) + x\right) = 1 - \alpha$$

→ confidence interval for μ and σ known:

ϵ (maximum error estimate) $= z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

→ <u>Sample size</u>: When α, ϵ, σ, are known that sample size n is given by $n = \left(\frac{z_{\alpha/2}\,\sigma}{\epsilon}\right)^2$. When σ is unknown symbol, σ is replaced by 's'. s is the standard deviation of sample to determine E.

* Thus the maximum error estimate $E = t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$ with $1-\alpha$ pb

→ <u>Maximum error of estimate ϵ for small sample</u>:

$$P\left(\bar{x} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) + \bar{x}\right) = 1-\alpha$$

→ confidence interval for μ, σ unknown (unknown variance, $(1-\alpha)\,100\%$ confidence interval for μ).

$E = t_{\alpha/2}\frac{S}{\sqrt{n}}$ where n = small sample size
S = standard deviation of sample

→ sample size for estimation population proportion

$$E = z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$E = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$$

$$n = \frac{z^2 PQ}{E^2}$$

→ confidence intervals

$(1-\alpha)\,.01$ $\quad (1-\alpha)\,100\%$ $\longrightarrow z_{\alpha/2}$

(5%) 95% $\longrightarrow \pm 1.96$

(2%) 98% $\longrightarrow \pm 2.33$

(1%) 99% $\longrightarrow \pm 2.5$

Q) A random sample of size 100 has mean 10, the population variance begins 25. Find the interval estimate of the population mean with confidence level (i) 95%. (ii) 99%.

S) $\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$\bar{x} = 10, \quad n=100, \quad \sigma^2 = 25, \quad \sigma = 5$

(i) 95%.

$Z_{\alpha/2} = 1.96$

$10 \pm 1.96 \times \frac{5}{\sqrt{100}}$

$10 \pm 1.96 \times 0.5$

10 ± 0.98

$10-0.98, \quad 10+0.98$

$(9.02, 10.98)$

(ii) 99%.

$Z_{\alpha/2} = 2.53$

$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 10 \pm 2.53 \times 0.5$

$= 10 \pm 1.26$

$= 10-1.26, \quad 10+1.26$

$= (8.74, 11.26)$

Q) If we assert 95% that the maximum error is 0.05 and p = 0.2. Find the size of the sample.

$e = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

$n = \frac{PQ \times Z^2_{\alpha/2}}{e^2}$

Scanned by CamScanner

$\epsilon = 0.05$, $P = 0.2$, $Q = (1-P) = 0.8$, $Z_{\alpha/2} = 1.96$

$$= \frac{0.2 \times 0.8 \times (1.96)^2}{(0.05)^2}$$

$$n = 245.86$$

$$n = 246$$

Q) In a study of an automobile insurance here ran sample of 80, body repair cost, at a mean of rupees 472.36 and the standard deviation of rupees 62.35. If $\bar{x}$ is used as a point estimate to the true average repair cost, with what confidence we can assert that the maximum error does not exceed rupees 10.

Sol) size of a random sample, $n = 80$

Mean $\bar{x} = 472.36$

$\sigma = 62.35$

Maximum error estimate, $\epsilon = 10$

and $z = ?$

$$\epsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha/2} = \frac{\epsilon\sqrt{n}}{\sigma}$$

$$= \frac{10 \times \sqrt{80}}{62.35}$$

$$z_{\alpha/2} = 1.4345$$

$\alpha/2 = 0.4236$ (from table)

$\alpha = 2 \times (0.4236)$

$\alpha = 0.8472.$

Scanned by CamScanner

Bayesian estimation:

It is used to update mean and variance of "posterior" distribution of a population.

* If the prior distribution parameters mean μ_0 and variance σ_0^2 of a population are known, then find the posterior distribution parameters of a given population. This is called Bayesian estimation.

$$\mu_1 = \frac{n\bar{x}\,\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2 \cdot \sigma^2}{n\sigma_0^2 + \sigma^2}}$$

where n = sample size, $\bar{x}$ = sample mean

s = standard deviation of sample ($s = \sigma$)

μ_0 = prior mean

→ Bayesian interval for μ is given by

$$\mu_1 - z_{\alpha/2}\,\sigma_1 < \mu < \mu_1 + z_{\alpha/2}\,\sigma_1$$

Q) A professor feeling about the mean mark in the final examination in probability of a large group of students is expressed subjectively by normal distribution $\mu_0 = 67.5$ $\sigma_0 = 1.5$

(i) If the mean mark lies in the interval 65, 70. determine the clear probability should assign to the mean mark.

(ii) Find the professor mean μ_1 and the posterior standard deviation σ_1. If the examination are conducted in a random sample of 40 students yielding mean 74.9 & standard deviation use $s = 7.4$ as an estimate. σ

(iii) Determine the posterior probability which he will assign to the mean mark being the interval 65, 70 using results obtained in (ii).

(iv) construct a 95% bayesian interval for μ.

Sol) (i) $\mu_0 = 67.5$

$\sigma_0 = 1.5$

Standard variable

corresponding to 65 (65.0, 70.0)

$$Z_1 = \frac{x - \mu}{\sigma}$$

$$= \frac{65 - 67.5}{1.5}$$

$$Z_1 = -1.66$$

$$Z_2 = \frac{70 - 67.5}{1.5}$$

$$Z_2 = 1.66$$

let X be the mean mark obtained in the final examination

prior probability $= P(Z_1 \le Z \le Z_2) = A(Z_1) + A(Z_2)$

$P(-1.66 \le Z \le 1.66) = A(1.66) + A(1.66)$

$= 0.4515 + 0.4515$

$= 0.9030$

(ii) $\bar{x} = 74.9$, $\sigma = 7.4$, $\sigma_0 = 1.5$ $n = 40$

posterior mean $= \dfrac{n\bar{x}\sigma + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$

$$\mu_1 = \frac{40(74.9)(1.5)^2 + (67.5)(7.4)^2}{40(1.5)^2 + (7.4)^2}$$

$\mu_1 = 72.107$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\cdot\sigma_0^2 + \sigma^2}}$$

$$= \sqrt{\frac{(1.5)^2(7.4)^2}{40(1.5)^2 + (7.4)^2}}$$

$= 0.927$

(iii)

$\mu_1 = 72.10$

$\sigma_1 = 0.922$

$$Z_1 = \frac{x - \mu_1}{\sigma_1}$$

$$= \frac{65 - 72.10}{0.922}$$

$= -7.70$

$$Z_2 = \frac{70 - 72.10}{0.922}$$

$$Z_2 = -2.28$$

posterior probability

$$P(Z_1 \le Z \le Z_2) = A(Z_1) - A(Z_2)$$

$$P(-7.77 \le Z \le -2.28) = A(7.7) - A(2.28)$$

$$= 0.5 - 0.48871$$

$$= 0.0113$$

(iii) 95% Bayesian limits are

$$\mu_1 \pm Z_{\alpha/2} (\sigma_1)$$

$$\mu_1 = 72.107, \; Z_{\alpha/2} = 1.96, \; \sigma_1 = 0.922$$

$$72.107 \pm (1.96)(0.9225)$$

$$72.107 \pm 1.807$$

$$(72.107 - 1.8081, \; 72.107 + 1.8081)$$

$$= (70.29989, \; 73.915)$$

Testing of Hypothesis:

When parametric values are unknown, we estimate them through sample values. But the problem arises when the sample provides a value, which is neither exactly equal to the parameter value, not too far. In that situation one has to develop some procedure which enables one to decide whether to accept a value or not on the basis of sample values, such a procedure is known as testing of hypothesis.

Statistical hypothesis:

In many circumstances to arrive at destination about the population on the basis of sample information we make assumptions (guesses) about the population parameters involved, such an assumptions is called a statistical hypothesis.

Testing of statistical hypothesis:

A statistical test of hypothesis is a rule or procedure which makes I to decide about the acceptancy or rejection of the hypothesis. It can be denoted by 'H'.

1) NULL hypothesis
2) Alternate hypothesis.

NULL hypothesis:

For applying the test of significancy, we first set up a hypothesis which is a statement about the population parameter such hypothesis which is a hypothesis of no differences is called a null hypothesis i.e. there is no significance between 2 parameters. It is denoted by Ho (Initial statement)

Alternate hypothesis:

Any hypothesis which is complimentary to the NULL hypothesis is called an alternative hypothesis. It is denoted by H_1. i.e. there is significantly between 2 parameter. (opposite of initial statement).

Critical region:

let $x_1, x_2, ---x_n$ be the sample observations and 'S' be the sample space. 'S' into 2 disjoint parts 'W' & 'S-W'. A region in the sample space S which amount & rejection to Ho is called critical region or rejection of region. W is the critical region and S-W is the accept region. If the sample parts falls in the subset W. Ho is rejected otherwise Ho is accepted.

level of significancy:

The pb of type1 error is known as level of significancy. The level of significancy usually employed in testing of hypothesis or 5% and 1%, α is always fixed in advance before collecting the sample information.

Errors in sampling:

There are 2 types of error sampling

1) Type1 error   2) Type2 error.

⇒ Type (1) error (or) first kind of error:

Reject H_0 when H_0 is true. It is also known as rejection error. pb of type 1 error is denoted by α.

* The medicine is good, even though we reject.

⇒ Type2 error (or) second kind of error:

Accept H_0 when H_0 is false. It is also known as acceptancy error. pb of type 2 error is denoted by β.

Ex: Type 2 error is more serious than type1 error.

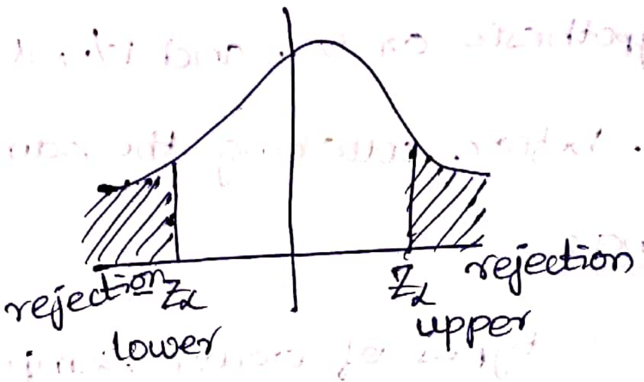
* Smoking is injurious to health even though we accept.

* The medicine is bad even though we accept.

Scanned by CamScanner

⇒ 2-tailed test:

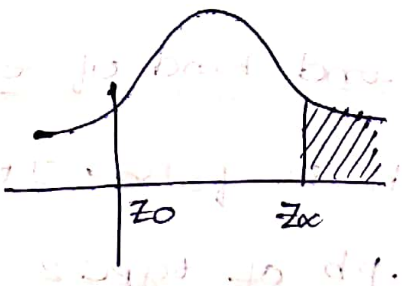
In alternative hypothesis, leads two alternative to the Ho, It is said to be 2 tailed test.



One tailed test:

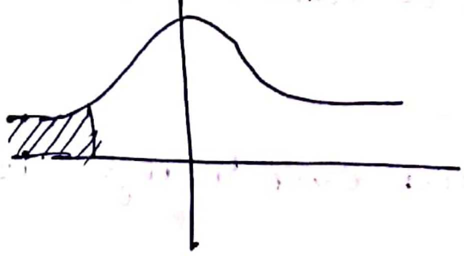
Right tailed test:

If alternative hypothesis, H_1, $\mu > \mu_0$ to the H_0, it is said to be right tailed test. In this situation half of the area of critical region lies half on the right



* left tailed test

$\mu < \mu_0$



Procedure for Testing of Hypothesis:

step 1: set up the NULL hypothesis has no significance difference b/w static and parameter.

step 2: setup of alternate hypothesis H_1 that we could decide whether we should use one-tailed or two-tailed test.

step 3: choose the appropriate level of significancy α (1% or 5% or any percentage)

step 4: calculate the value of Z, test statisty under the NULL hypothesis.

$$|Z| = \frac{t - E(t)}{S - E(t)}$$

where t = statisty . S.E(t) . standard error (t).

conclusion:

compare calculated values of Z with the tabulated values at LO; LOS (level of significance)

If Z calculated value < Z table value then we accept our NULL hypothesis as L L.O.S.

If Z calculated value > Z tabular value then NULL hypothesis is rejected at L L.O.S we accept alternate hypothesis H_1

critical value of Z.

L.O.S

Z_α	1%	5%	10%
two-tailed test	2.58	1.96	1.645
Right-tailed test	2.33	1.645	1.28
Left-tailed test	-2.33	-1.645	-1.28

Test of significance for large sample:

standard Normal variance $Z = \frac{x-\mu}{\sigma}$

where $\mu = np$

$\sigma = \sqrt{npq}$

Q) A coin was tossed 980 times are returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance (L.O.S)

Sol 1) NULL hypothesis (Ho):

The coin is unbiased

2) Alternate hypothesis (H_1):

The coin is biased

3) Level of significance:

$\alpha = 1.96$

4) Test of statisty:

standard Normal variate $Z = \frac{x-\mu}{\sigma}$

$n = 960$, $x = 183$, p = probability of getting head

$p = 1/2$

$Z = 1/2$

$\mu = np$

$= 960 \times 1/2$

$\mu = 480$

$\sigma = \sqrt{npq}$

$\sigma = \sqrt{960 \times \frac{1}{4}}$

$\sigma = \sqrt{240}$

$\sigma = 15.49$

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

$= \frac{183 - 480}{15.49}$

$Z = -19.17$

$|Z| = 19.17$

Step4: $|Z|$ cal $>$ $|Z|$ table

Step5: Ho is rejected

$\therefore$ The coin is biased.

(Q) A die is tossed 960 times and it falls with 5 upwards 184 times. Is the dice unbiased at the level of significancy of 0.01

(b) Test of significance of a single mean - large samples.

$\therefore$ The test of statisty $= Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ = we use

$(S = \sigma)$

(8) An oceanographer wants to check whether the depth of the ocean is at certain region is 57.4 fathoms as had a professionally has been recorded. What can he conclude at the 0.05 level of significancy, if readings taken as 40 random locations in the given region yeilded. A Mean of 59.1 fathoms with a standard deviation or 5.2 fathoms.

Sol

(i) NULL hypothesis. $\mu = 57.4$

2) Alternate hypothesis $H_1: \mu \neq 57.4$

3) Level of significance $\alpha = 5\%$.

$Z_\alpha = 1.96$, $|Z|_{0.05}$.

4) Test of statisty

$$|Z| = \frac{x - \mu}{\sigma / \sqrt{n}}$$

$\bar{x} = 59.1$, $\mu = 57.4$

$|Z|_{cal} = 2.06$

$|Z|_{0.01} > |Z|_{table}$

i.e. $2.06 > 1.96$

Step 5: Conclusion

There Ho (NULL hypothesis) is rejected i.e. H_1 is accepted (Alternate hypothesis). Then the depth of the ocean $\mu \neq 57.4$

Scanned by CamScanner

Test of significancy for difference of mean of 2 samples

Test of statisty:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If the sample have been drawn from the population with common standard deviation σ then

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

if $\sigma_1^2 = \sigma_2^2 = \sigma$

If σ is not known we can use and estimate σ^2 is given by

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

<u>Notes:</u>

If the two samples are drawn from two population with unknown standard deviations σ_1^2 & σ_2^2 then σ_1^2 & σ_2^2 can be replaced by sample variances s_1^2 & s_2^2 provided both the samples n_1 and n_2 are large then test of statisty:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}}$$

Scanned by CamScanner

(5) The means of two large samples 1000 and 2000 members are 67.5 inches and 68 respectively can the samples be regarded as drawn from the same population of standard deviation of 2.5 inches.

Sol

$n_1 = 1000$ $\quad n_2 = 2000$

$\bar{x}_1 = 67.5$ $\quad \bar{x}_2 = 68$

$\sigma = 2.5$

NULL hypothesis Ho the samples have been drawn from the same population of standard deviation 2.5 inches.

$\mu_1 = \mu_2$

Alternate hypothesis

$\mu_1 \neq \mu_2$

Level of significance

$\alpha = 5\% = 1.96$

Test of statisty:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$Z = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$Z = -5.16$

$|Z| = 5.16$

$|Z|_{table} = 1.96$

$|Z|_{cal} > |Z|_{table}$

Scanned by CamScanner

∴ Ho is rejected H_1 is accepted.

∴ The samples are not drawn from the same population of standard deviation 2.5.

Q) The samples are drawn from the 2 universities and from their weights in kilograms mean & SD are calculated as shown below. Make a large sample test to test the significance of the difference b/w the mean

	mean	SD	size of sample
university A	55	10	400
university B	57	15	100

sol $\bar{x}_1 = 55$, $\bar{x}_2 = 57$,

$SD_1 = 10$, $SD_2 = 15$

$n_1 = 400$, $n_2 = 100$

Ho: null hypothesis $\bar{x}_1 = \bar{x}_2$ (there is no difference)

H1: alternate hypothesis $\bar{x}_1 \neq \bar{x}_2$

level of significance $Z_\alpha = 1.96$

Test of statisty. $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$Z = \frac{55-57}{\sqrt{\frac{10^2}{400} + \frac{15^2}{100}}}$$

$Z = -1.26$

$|Z| = 1.26$

$|Z|_{cal} < |Z|_{table}$

$\Rightarrow$ Ho is accepted

$\therefore$ There is no significance difference b/w mean

Test of significancy for single proportion of large sample:

Suppose a large sample of size n is taken from a normal population to test the significance difference b/w sample proportions p and population P

we use the statistic

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

where n is the sample size

Note:

1) limits for population parameter P are given by $p \pm 3\sqrt{\frac{pq}{n}}$ where $q = 1-p$

2) confidence interval for population P for sample at α level of significance 'S'

$$p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}} < P < p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}} \quad \text{where } Q = 1-P$$

Scanned by CamScanner

Q) A manufacturer claim that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment reveal that 18 were fault. Test his claim at 5% level of significance

Sol. sample size $n = 200$

No. of pieces conforming two specifications

$= 200 - 18 = 182.$

p – proportion of pieces conforming two specifications.

P – population proportion $= 95\% = 0.95$

$Q = 1 - 0.95 = 0.05$

H_0 – Null hypothesis.

The proportion of pieces conforming for specifications.

$P = 95\% = 0.95.$

H_1 = alternate hypothesis

$P < 0.95$ (left tail test)

Level of significance 5%

$Z_{\alpha} = 1.1645$

Test of statisty

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

Scanned by CamScanner

$$= \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.01}{200}}}$$

$$Z = -2.59$$

$$|Z| = 2.59$$

$$|Z|_{cal} > |Z|_{table}$$

Ho is rejected, H1 is accepted

∴ The manufacturer claim is rejected.

Q) A random sample of 500 pine apples was taken from a large consonmunt was 65, were found to be bad. Find the percentage of bad pineapple in the consonment.

$n = 500$

P = population of bad pineapples in the sample $= \frac{65}{500} = 0.13$

$q = 1 - P = 1 - 0.13$

$= 0.87$

we know that the limits for population proportion

$$P = P \pm 3\sqrt{\frac{Pq}{n}}$$

$$= 0.13 \pm 3\sqrt{\frac{0.13 \times 0.89}{500}}$$

$= 0.13 \pm 0.0451$

$= 0.13 + 0.0451,\ 0.13 - 0.0451$

$= (0.1751, 0.0849)$

Scanned by CamScanner

∴ The percentage of bad pineapples in the consignment lies between 8.5 and 17.5.

Q) Among 900 people in a state 90 are found to be chapathi eaters. Construct 99% confidence interval for the true proportion.

Sol $n = 900$, $x = 90$

$$p = \frac{90}{900} = 0.1$$

$$Q = 1 - 0.1 = 0.9$$

∴ confidence interval is $P \pm 3\sqrt{\frac{PQ}{n}}$

$$= 0.1 \pm 3\sqrt{\frac{0.1(0.9)}{900}}$$

$$= 0.1 \pm 0.03$$

$$= (0.13, 0.07)$$

Test of significance of difference b/w two sample proportions - large samples.

Let P_1 and P_2 be the sample proportions in two large random samples of size n_1 and n_2 drawn from two populations having proportions P_1 and P_2 then test of statisty.

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

where P_1 and P_2 are unknown

when the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known in the case. we have two methods to estimate P_1 and P_2.

1) Method of substitution:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

2) Method of pooling:

In this method two sample proportions p_1 and p_2 into a single proportions p

$$\therefore p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$Z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Q) A random sample of 400 men and 600 women were asked whether they would like to have a fly over near their residincy 200 men and 325 women were in favour of the proportional. Test the hypothesis that proportions of men and women in favour of the proposal or sum at 5% level.

sol

sample size $n_1 = 400$

$n_2 = 600$

proportion of men $P_1 = \frac{200}{400} = 0.5$

" " women $P_2 = \frac{325}{600} = 0.54$

Ho - Null hypothesis.

There is no significant difference b/w the option of men and women as far as proposal of flyover is concerned

$P_1 = P_2 = P$

H1 : Alternate hypothesis

$P_1 \neq P_2$

level of significancy 5%

$Z_\alpha = 1.96$.

Test of statisty

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

$P_1 = 0.5 \quad q_1 = 0.5$

$P_2 = 0.541 \quad q_2 = 0.459$

$$Z = \frac{0.5 - 0.541}{\sqrt{\frac{(0.5)(0.5)}{400} + \frac{(0.541)(0.459)}{600}}}$$

$Z - 1.272$

$|Z|_{cal} = 1.27$

$|Z|_{tab} = 1.96$

$|Z|_{cal} < |Z|_{table}$

$\therefore$ Ho is accepted.

Scanned by CamScanner