5 Probability and Statistics

Objective

Upon completion of this chapter you will be able to:

- Understand the key terms associated with Probability i.e. Sample Space, Event etc.
- Determine the probability of occurrence of an event.
- Understand different discrete and continuous probability distributions.
- Understand the basics of statistics i.e. mean, standard deviation and variance.

Introduction

The foundations of probability theory were built by a mathematical study of games of chance. Today a huge gambling industry rests on the foundations of probability theory. Casinos make sure that people win just enough to keep them hooked but the odds are slightly balanced in favor of casinos so that they always come ahead. Similarly, in the Stock Market investors are always engaged in predicting the random fluctuations in the market. also, insurance business relies heavily on the probability theory. similarly in communication engineering digital communication is based on probability theory.

Definitions

Sample Space and Event:

If the output of any experiment is not known with certainty like tossing a coin or rolling a dice, then such an experiment is called as random experiment. But we can assume that the set of all possible outcomes is known like the outcome of a coin toss can be either heads or tails and the outcome of dice rolling can be from 1-6.

The set of all possible outcomes of an experiment is called as the sample space of an experiment and is represented by S. Thus, the set {H,T} for toss outcome will be the sample space for coin toss. Also, the set {1,2,3,4,5,6} is the sample space for dice rolling.

Any subset of sample space is called as an Event. As for example, the occurrence of head on a coin toss is an event. Similarly, the possibility of an outcome of an even number i.e. {2,4,6} is an event for rolling a dice.

Union and intersection:

For any two events A and B we define a new event $A \cup B$ called as "A union B" which consists of outcomes that are either in A or B or in both A and B. This also means $A \cup B$ will occur if A or B or both have occurred.

Similarly, we define an event $A \cap B$ called as "A intersection B" which consists of outcomes that are in both A and B. This means $A \cap B$ will occur if both A and B occur.

Complementary event:

The event E^c is called a complementary event for the event E. It consists of all outcomes not in E, but in S. As for an example, the occurrence of head is a complementary event to occurrence of tail. Similarly, the occurrence of even number on a dice roll is complementary to the occurrence of an odd number on a dice roll.

Equally likely events:

Two events are said to be equally likely if their probability of occurrence is equal. As for an example the probability of occurrence of tail on a coin toss is the same as occurrence of head and both are equal to ½ so both events are equally likely.

Mutually Exclusive Events:

Two events are mutually exclusive if they have no common outcome. As an example occurrence of a head on a coin toss and the occurrence of tail are both mutually exclusive events. Similarly, the occurrence of odd number and an even number on a coin toss are both mutually exclusive events. But the occurrence of a number divisible by 2 on a dice roll i.e. {2,4,6} and the occurrence of a number divisible by 3 i.e. {3,6} have the common outcome {6}. Thus, they are not mutually exclusive events.

In mathematical terms, we can say A and B are two mutually exclusive events if, $P(A \cap B) = 0$

Collectively Exhaustive Events

Two events E and F are collective exhaustive if they combined form the sample space of an event. For example occurrence of head on a coin toss and occurrence of tail are collective exhaustive events as they form the complete sample space i.e. {H,T} of a coin toss event.

Independent events:

Two events A and B are said to be independent if the probability of occurrence of one event does not depend on the probability of occurrence of the second. This, is expressed in terms of conditional probability, $P(A \mid B)$ which represents the probability of occurrence of A given that B has occurred.

If A and B are independent, then $P(A \mid B) = P(A)$ and in a similar manner $P(B \mid A) = P(B)$.

This is also expressed as, $P(A \cap B) = P(A) \times P(B)$

De Morgan's:

This law is used to find the probability of complement of a group of events as shown below,

$$\left(\bigcup_{i=1}^{n} E_1\right)^C = \bigcap_{i=1}^{n} E_i^C$$

This indicates that the complement of the event that at least one of the events occur is the same as the probability that none of the events occur.

$$\left(\bigcap_{i=1}^{n} E_i\right)^C = \bigcup_{i=1}^{n} E_1^C$$

Axioms of probability:

Consider an event A for an experiment where the sample space is represented as S. Then, the following axioms must be satisfied,

Axiom-1: $0 \le P(A) \le 1$

Here, $P(A) = 0$ indicates an impossible event.

$P(A) = 1$ indicates a certain or sure event.

Axiom-2: $P(S) = 1$, which indicates that the probability of sample space is 1 which means that certainly one of the events from the sample space will occur.

Axiom-3: If there are 'n' mutually exclusive events A_1 , A_2 , ... , A_n which means $P(A_i \cap A_j) = 0$. Then, P

$$\text{Then, } P\left(\bigcap_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

There are certain rules of the probability of calculation of compound event involving events A and B,

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

If A and B are mutually exclusive events, $P(A \cap B) = 0$

Then, $P(A \cup B) = P(A) + P(B)$

Also, the probability of intersection can be expressed in terms of conditional probability,

$P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$

Here $P(A)$ and $P(B)$ are called as marginal probabilities of A and B and $P(A \cap B)$ is called as joint probability of A and B.

If A and B are independent events, then joint probability is the same as product of marginal probabilities.

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

Some rules on complementary probability are,

$$P\left(A^{c}\right)=1-(A)$$

$$P\left(A^{c}\cap B\right)=P(B)-P(A\cap B)$$

$$P\left(A^{c}\cap B^{c}\right)=P\left(\overline{A\cup B}\right)=1-P(A\cup B)$$

$$P(A\cap B)^{c}=P\left(A\cap B^{c}\right)+P\left(A^{c}\cap B\right)$$

$$P(A^{c}\mid B=\frac{P\left(A^{C}\cap B\right)}{P(B)}=\frac{P(B)-(A\cap B)}{P(B)}$$

$$=\frac{1-P(A\cap B)}{P(B)}=1-P(A/B)$$

$$P(A\mid B^{c}=\frac{P\left(A\cap B^{c}\right)}{P\left(B^{c}\right)}=\frac{P(A)-P(A\cap B)}{1-P(B)}\because (P(B)\neq 1$$

$$P\left(A^{c}\mid B^{c}\right)=\frac{P\left(A^{C}\cap B^{c}\right)}{P\left(B^{C}\right)}=\frac{1-P(A\cup B)}{1-P(B)}\because P(B)\neq 1$$

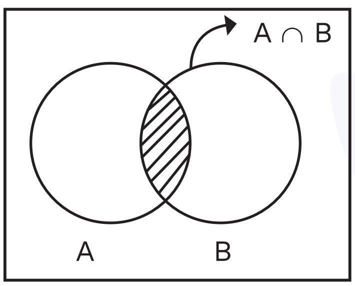

Fig. 5.1 A intersection B

If A and B are independent events, the probability of ($A\cap B^{c}$), P($A^{C}\cap B$) and P($A^{C}\cap B^{C}$) are also zero and these events are also independent.

Probability of an Event

The probability of an event can be computed by the classical approach in terms of favorable cases and sample space i.e. total outcomes.

Assume, m represents the number of favorable cases for an event E happening and n represents the total number of possible cases.

Then, the probability of occurrence of E is,

$$P(E)=\frac{m}{n}$$

As for an example, if we wish to compute the probability of occurrence of an odd number on a dice roll then the set {1,3,5} represents favorable cases and the set {1,2,3,4,5,6} represents sample space.

Thus, P (Odd Number) $=\frac{3}{6}=\frac{1}{2}$

Solved Examples

Example: If 3 coins are tossed at a time find P(getting at most one head)?

Solution: The 3 coin tosses are independent events. So we can multiply their probabilities.

Probability of occurrence of at most one Head is equal to the probability that no Head occurs and probability that exactly one Head occurs.

If a fair coin is tossed, P (Head) = P (Tail)$=\frac{1}{2}$

P (No-Head) = P (All Tails) $=\frac{1}{2}\times\frac{1}{1}\times\frac{1}{2}=\frac{1}{8}$

If exactly one Head occurs then there can be 3 cases that Head occurs on first toss, on second toss or on third toss.

$$P(\text{One}-\text{Head})=3\times P(\text{OneHead}\ \&\ 2\,\text{Tails})$$

$$=3\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{3}{8}$$

$$P(\text{At Most One}-\text{Head})=\frac{1}{8}+\frac{3}{8}=\frac{1}{2}$$

Example: Find the probability of occurrence of at least one Tail if 3 coins are tossed simultaneously.

Solution: Here also all tosses are independent so the probability of intersection can be calculated by multiplication of individual probabilities.

The complementary event of at least one tail would be no tail and thus, we can write,

P(At Least One - Tail) = 1 - P(No Tail) = 1 - P (All Heads)

P (At Least - One Tail) $= 1 - \frac{1}{2}\times\frac{1}{2}\times\frac{1}{2} = \frac{7}{8}$

Example: In the same problem find the probability of getting at least one head and at least one tail?

Solution: The sample space of 3 coin tosses is {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} The favorable cases are {HHT, HTH, HTT, THH, THT, TTH}

P (At Least One Head and One Tail) $= \frac{6}{8} = \frac{3}{4}$

This can also be computed as

1 -P(No Head) -P (No Tail) $= 1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4}$

Example: A player tosses 6 coins, Find the probability that number of heads is more than number of tails.

Solution: The favorable cases are when there are 4 Heads and 2 Tails or when there are 5 Heads and 1 Tails or when there are all 6 heads.

P (4 Heads & 2 Tails) $= {}^6C_2\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2 = \frac{15}{64}$

Here, we have considered the combination term as the two tails can occur on any 2 of the 6 tosses.

P (5 Heads & 1 Tails) ${}^6C_1\left(\frac{1}{2}\right)\left(\frac{1}{1}\right) = \frac{6}{64}$

$P(6\ \text{Heads}) = \left(\frac{1}{2}\right)^6 = \frac{1}{64}$

Thus, $P(\text{Heads} > \text{Tails})$

$$= \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{22}{64} = \frac{11}{32}$$

This problem can also be dealt with in terms of binomial random variable which will be covered later.

Example: Two dice are thrown 2 times. Find the probability of getting a sum of 7.

(A) At least once

(B) Only once

(C) Twice

Solution: Consider A as the event when a sum of 7 occurs on the first throw and B as the event when a sum of 7 occurs on the second throw. Here A and B are both independent events.

There are 36 possible outputs when two dice are thrown as 6 outputs are possible on each dice.

The following combinations will yield a sum of 7: {1,6}, {2,5}, {3,4}, {4,3}, {5,2}, {6,1} i.e. a total of 6 favorable cases.

$$P(A) = \frac{6}{36} = \frac{1}{6}\ \&\ P(A^c) = 1 - P(A) = \frac{5}{6}$$

$$P(B) = \frac{6}{36} = \frac{1}{6}\ \&\ P(B^c) = 1 - P(B) = \frac{5}{6}$$

(A) P (At least One)

$= P(A \cup B) = 1 - P(A^c \cap B^c) = 1\ [P(A^c).P(B^c)]$

P(At lease Once) $= 1 - \frac{5}{6} \times \frac{5}{6} = \frac{11}{36}$

(B) $P(\text{only once}) P(A \cap B^c) + P(A^c \cap B)$

$$= P(A)P(B^c) + P(A^c)P(B) = \frac{1}{6}\times\frac{5}{6} + \frac{5}{6}\times\frac{1}{6} = \frac{10}{36}$$

(C) $P(\text{twice}) = P(A \cap B) = P(A)P(B) = \frac{1}{6}\times\frac{1}{6} = \frac{1}{36}$

Example: Two dice are rolled simultaneously. Find the probability of getting a prime number on the first or a sum of 8 on both.

Solution: There are 3 possible prime number on a dice roll viz. {2,3,5}

$$P(A) = (\text{prime number}) = \frac{3}{6} = \frac{1}{2}$$

There are a total of 36 combinations on the two dice roll and the following combinations will yield a sum of 8: {2,6}, {3,5}, {4,4}, {5,3}, {6,2} i.e. a total of 5 combinations

$P(B) = P(\text{sum of } 8) = \frac{5}{36}$

The probability that first dice shows prime number and sum is 8 has three cases {3,5}, {5,3}, {2,6}

In our case, we need to compute

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

$$=\frac{1}{2}+\frac{5}{36}-\frac{1}{12}=\frac{5}{9}$$

Example: A determinant is chosen from a set of all determinants of order 2 with the elements 0 (&) or 1. Find the P (the chosen determinant is non zero)

Solution: Each element in 2×2 matrix can be zero or one. Thus there are 16 such possible matrices.

$$n(S)=16$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \Delta = ad-bc \neq 0$$

$$\text{Case (i)}: \Delta=+1[a=d]$$

$$=1 \text{ and at least one of } b=c=0$$

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 3$$

$$\text{Case (ii)}: \Delta=-1[b=c]$$

$$=1 \text{ at least one of } a=d=0$$

$$\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = 3$$

$$P(\text{nonzero}\,\Delta)=\frac{6}{16}=\frac{3}{8}$$

Example: Four cards are drawn at random from the pack of 52 cards.

(A) Find P(all 4 cards are drawn from the same suit)

(B) P(no two cards are drawn from the same suit)

Solution: If all four cards are from same suit then there are four possibilities that all four are from spades, diamonds, hearts, clubs.

$$P(\text{All 4 from suit})=\frac{{}^{13}C_4+{}^{13}C_4+{}^{13}C_4}{{}^{52}C_4}=\frac{44}{4165}$$

If all 4 cards are from different suits then each card is drawn from a different suit

P(no two cards from the same suit)

$$=\frac{{}^{13}C_1\times{}^{13}C_1\times{}^{13}C_1\times{}^{13}C_1}{{}^{52}C_4}=\frac{2197}{20825}$$

Example: A card is drawn from the pack of 52 cards. Find the probability that it is neither a diamond nor an ace card.

Solution: There are 13 possible cards in diamond, so the probability that the card drawn is a diamond,

$$P(\text{Diamond})=\frac{13}{52}$$

There are 4 aces, one in each suit, so the probability that the card drawn is ace

$$P(\text{ace})=\frac{4}{52}=\frac{1}{13}$$

The probability that the drawn card is an ace of diamonds

$$P(D\cap\text{ace})=\frac{1}{52}$$

$$P(D^c\cap\text{ace}^c)=1-(D\cup\text{ace})=1-\left[\frac{13}{52}+\frac{4}{52}-\frac{1}{52}\right]$$

$$=\frac{36}{52}=\frac{9}{13}$$

Example: A and B are the two players rolling a dice on the condition that one who gets the 2 first win the game. If A starts the game what are the winning chances of player A, B.

Solution: Probability of getting a 2 on a dice roll is,

$$P(2)=\frac{1}{6}=p$$

Probability of getting any number other than 2 on dice roll is,

$$P(2^c)=\frac{5}{6}=q$$

If A starts the game then he can win in the following cases,

- A gets 2 on the first roll.
- A gets any number other than 2 on first roll, then B gets any number other than 2 on second roll then A gets 2 on the third roll.

- Similarly, A can get 2 on odd number of rolls and on rolls previous to winning one there should be any number other than 2.

$$P\left(A_{win}\right) = p + q^2p + q^4p + q^6p + \ldots\ldots\ldots\ldots$$

$$P\left(A_{win}\right) = \frac{p}{1-a^2} = \frac{1/6}{1-25/36} = \frac{6}{11}$$

If we wish to calculate the probability of B winning,

$$P\left(B_{win}\right) = 1 - P\left(A_{win}\right) = 1 - \frac{6}{11} = \frac{5}{11}$$

Example: A, B & C are tossing a coin on the condition that one who gets the head first wins the game. If A starts the game, what are the winning chances of C in the 3rd trial?

Solution: If a fair coin is tossed then, probability of head and teal occurrence is,

$$P\left(H\right) = \frac{1}{2} = p \qquad p\left(T\right) = \frac{1}{2} = q$$

Now, if C wins on the third trial, then all must get tails on first and second trial and then on third trial A and B must get Tails but C must get Heads.

$$P\left(winC\right) = q^3q^3\ qqp =$$

$$\left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^3 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{512}$$

Conditional Probability

Conditional Probability represents the probability of an event occurrence given that some other event has already occurred.

P(A|B) represents the probability of occurrence of A given that B has already occurred.

Total probability theorem:

Suppose there are n events which that are mutually exclusive and collectively exhaustive then probability of occurrence of an event E is given by,

$P(E) = P(A_1 \cap E) + P(A_2 \cap E) + \ldots\ldots\ldots\ldots + P(An \cap E)$

in terms of conditional probability,

$P(E) = P(A_1)P(E \mid A_1) + P(A_2)P(E \mid A_2) + \ldots\ldots\ldots\ldots P(A_n)P(E \mid A_n)$

Bayes' Theorem

This theorem is used for finding reverse probability,

$$P\left(A_i \middle| E\right) = \frac{\left(A_1 \cap E\right)}{P(E)}$$

If we represent P(E) in terms of the total probability theorem

$$P\left(A_i \middle| E\right) = \frac{p\left(A_i \cap E\right)}{P\left(A_1\right)P\left(E \middle| A_1\right) + P\left(A_2\right)P\left(E \middle| A_2\right) + \ldots\ldots\ldots\ldots P\left(A_n\right)P\left(E \middle| A_n\right)}$$

Solved Examples

Example: A number is drawn from the 100 numbers those are 0, 1, 2, 3, ... 99. Let x denote the sum of digitson a number and y denotes the product of the digits on the number. Find P(x=9), given y=0.

Solution: The following cases can be formed where the product of digits of a number drawn is 0.

00	10
01	20
02	30
03	40
04	50
05	60
06	70
07	80
08	90
09	

In the following cases the sum of digits is 9; 09 and 90 i.e. 2 cases are there where the sum of digits are 9 and product of digits is 0

$$P\left(x = 9 \middle| y = 0\right) = \frac{P\left(x = 9 \cap y = 0\right)}{P\left(y = 0\right)} = \frac{2/100}{19/100} = \frac{2}{19}$$

Example: 60% of the employees of the company are college graduates of these 10% are in the sales department. Of the employees who didn't graduate from the college 80% are in the sales department. A person is selected at random. Find the probability that

(A) The person is in the sales department.

(B) Neither in the sales department nor a college graduate.

Solution: Assuming that the total number of employees is 100. Then, college graduates are 60.

Number of college graduates in sales department = 10% of 60 = 6

Number of people who are not college graduates = 40

Number of non college graduates in sales department = 80% of 40 = 32

Total number of employees in sales department = 32+ 6 =38

(A) Probability of a randomly selected person to be from the Sales Department,

P(sales) = 38/100

(B) Number of people not from college and not in sales = 40 -32 = 8

Probability that the randomly selected person is non college graduate and not in sales, P(N-S & N-G) = 8/100

Example: In answering a question on multiple choice test, the students either know the answer or guess the answer. Let P be the probability that student know the answer and 1-P that of guessing the answer. Assume that the student guesses the answer to a question will be correct with a probability of 1/5. What is the conditional probability that the student knew the answer to a question given that he answered correctly?

Solution: Probability of knowing the answer

P(K) = P

Probability of guessing the answer

P(G) = 1-P

Let us define the event E as the answer is correct.

Probability that the answer is correct given that student knows the answer P(E|K) = 1

Probability that the answer is correct given that student is guessing the answer

P(E|G)= $\frac{1}{5}$

By Total Probability Theorem,

$$P(E)=(K\cap E)+P(G\cap E)=P(K)P(E|K)+P(G)P(E|G)$$

$$P(E)=P\times 1+(1-P)\frac{1}{5}=\frac{4P+1}{5}$$

By Baye's theorem $P(K|E)=$

$$\frac{P(K\cap E)}{P(E)}=\frac{P(K)P(E|K)}{P(E)}=\frac{P\times 1}{\frac{4P+1}{5}}=\frac{5P}{4P+1}$$

Example: There are 3 coins out of which 2 are unbiased and one is biased coin with two heads. A coin is drawn at random and tossed 2 times. If head appears on at both the times, find the probability that the selected coin is a biased coin?

Solution: Number of Unbiased Coins = 2 Number of biased coins = 1

Probability of selecting an unbiased or a biased coin

$$P(UB)=\frac{2}{3}, P(B)=\frac{1}{3}$$

Let us define an event E as getting two heads on two tosses Probability that we get two heads using an unbiased coin,

$$P(E)=P(UB\cap E)+P(B\cap E)=$$
$$P(UB)P(E|UB)+P(B)P(E|B)$$

$$P(E)=\frac{2}{3}\times\frac{1}{4}\times\frac{1}{3}\times 1=\frac{2}{12}+\frac{1}{3}=\frac{6}{12}=\frac{1}{2}$$

$$\frac{P(B\cap E)}{P(E)}=\frac{\frac{1}{3}\times 1}{\frac{1}{2}}=\frac{2}{3}$$

Example: Player A speaking truth $\frac{4}{7}$ times and the card is drawn from the pack of 52 cards, he reports that there is a diamond. What is the probability that actually there was a diamond?

Solution: Probability of a person speaking the truth and lie are,

$$P(T)=\frac{4}{7}; P(L)=\frac{3}{7}$$

Let us define an event E as the Player A reports that there is a diamond. Probability that there is a diamond given that Player A is speaking the truth,

$$P(E|T) = \frac{13}{52} = \frac{1}{4}$$

Probability that there is a diamond given that he is lying,

$$P(E|T) = \frac{39}{52} = \frac{3}{4}$$

By total probability theorem,

$$P(E) = P(T \cap E) + P(L \cap E) = P(T)P(E|T)$$

$$+P(L)P(E|L) = \frac{4}{7} \times \frac{1}{4} \times \frac{3}{7} \times \frac{3}{4} = \frac{13}{27}$$

By Baye's theorem,

$$P(T|E) = \frac{P(T \cap E)}{P(E)} = \frac{\frac{4}{7} \times \frac{1}{4}}{\frac{13}{28}} = \frac{4}{13}$$

Example: A letter is known to have come from Tatanagar (or) Calcutta on the envelope, the just two consecutive letters T and A are visible. Find the probability that the letter has come from Calcutta?

Solution: Probability that the word selected is Tatanagar or Calcutta is,

$$P(T) = \frac{1}{2} \quad P(C) = \frac{1}{2}$$

There are 8 pairs of consecutive letters in Tatanagar i.e. "TA"; "AT"; "TA"; "AN"; "NA"; "AG"; "GA"; "AR". There are two cases where "TA" are consecutive

Similarly, in Calcutta there are 7 pairs of consecutive letters i.e. "CA"; "AL"; "LC"; "CU"; "UT"; "TT"; "TA" There is only one case where "TA" are consecutive.

Define an event E: Getting a 'TA'

$$P(E|T) = \frac{2}{8}$$

$$P(E|C) = \frac{1}{7}$$

By Total Probability Theorem,

$$P(E) = \frac{1}{2} \times \frac{2}{8} + \frac{1}{2} \times \frac{1}{7} = \frac{11}{56}$$

$$P(C|E) = \frac{P(C \cap E)}{P(E)} = \frac{\frac{1}{2} \times \frac{1}{7}}{\frac{11}{56}} = \frac{4}{11}$$

Example: There are 3 bags which contain blue, red and green color balls in the form of

	B	R	G
(A)	1	2	3
(B)	2	3	1
(C)	3	1	2

A bag is drawn at random and two balls are drawn from it. They are found to be 1 blue and 1 red. Find the probability that the selected balls are from bag C?

Solution: Since all bags are equally likely to be selected

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Let us define on event E : Getting one blue and one red ball

$$P(E|A) = \frac{1C1 \times {}^2C_1}{C_2} = \frac{2}{15}$$

$$P(E|B) = \frac{6C2 \times {}^3C_1}{{}^6C_2} = \frac{6}{15}$$

$$P(E|C) = \frac{{}^3C_1 \times {}^1C_1}{{}^6C_2} = \frac{3}{15}$$

By Total Probability Theorem,

$$P(E) = P(A \cap E) + P(B \cap E) + P(C \cap E) =$$

$$P(A)P(E|A) + P(B)P(E|B) + P(C)P(E|C)$$

$$P(E) = \frac{1}{3}\left[\frac{2}{15} + \frac{6}{15} + \frac{3}{15}\right] = \frac{11}{45}$$

By Bayes' theorem

$$P(C|E) = \frac{P(C \cap E)}{P(E)} = \frac{\frac{1}{3} \times \frac{3}{15}}{\frac{11}{45}} = \frac{3}{11}$$

Statistics

Statistics is a branch of mathematics that deals with summarizing large quantities of data into some meaningful parameters like mean, standard deviation etc.

Quantities like Mean, Median and Mode quantify the central value around which the data points are centered and quantities like standard deviation, variance and coefficient of variation quantify how scattered the data points are around central point.

Types of data

- Grouped and ungrouped
- Open and closed

Grouped data: If the data is in the form of class intervals and frequency together, then the data is known as grouped data or distributing the frequencies to their corresponding class intervals is known as frequency distribution.

Closed data: If the class intervals are in continuous form without any discontinuity, the data is known as closed data. Otherwise open data.

Ungrouped data: If the data contains only observations without any class intervals, then the data is known as ungauged data or raw data.

Mean (Average)

x_i represents the individual observation. Then, the mean value for grouped and ungrouped data is given by,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ for ungrouped data}$$

For grouped data we associate a frequency to each interval represented by 'f'. In any interval we also define the lower limit and upper limit of an interval.

$$\bar{x} = \frac{\sum_{i=1}^{n} fx_i}{N} \text{ for grouped data}$$

Where

$$X_i = \frac{\text{Upper limit} + \text{lower limit}}{2}$$

n= no. of observations

N= sum of frequencies

Median

Mean represents a central value in the sense that positive and negative deviations from the arithmetic mean balance each other.

On the other hand, median is the central value of the distribution such that a number of values less than median is equal to number of value greater than the median.

If n is odd → The middle observation itself is the median

If n is even → Average between the middle observations, provided

1. Data is rearranged either in ascending or in a descending order.
2. The number of observations above the middle is equal to the number of observations below the middle.
3. For the grouped data, the median is given by

$$\text{Median} = L + \left[\frac{\frac{N}{2} - F}{f_m}\right]h$$

Where,

L → lower limit for the median class

f_m → frequency for median class

F → cumulative frequency for above the median class.

h → size of the median class

Solved Examples

Example: Find the median for the following grouped data?

Class Interval	Frequency	Cumulative frequency
0 – 10	3	3
10 – 20	5	8
20 – 30	7	15
30 – 40	2	17
40 – 50	1	18
	N = 18	

Table 5.1

Solution: The middle value i.e. 9th and 10th values both lie in the class 20-30 and hence 20-30 is the median class. Lower limit of median class, L=20

Frequency of median class = 7

Cumulative frequency of class above median class = 8 Width of median class = 30-20 = 10

$$\text{Median} = 20 + \frac{9-8}{7} \times 10 = 20 + 1.4 = 21.4$$

Note: If the first class itself is median class, the cumulative frequency and frequency (f) are equal (m = f).

Mode

The most frequently repeated observation is known as the mode If there is only one value with the highest frequency then it is unimodal.

If there are 2 such values having the highest but equal frequencies of occurrence then it is bimodal signal. Eg: 1, 2, 3, 4, 5, 2, 3, 6, 7, 2, 3, 11, 14, 2, 21, 23, 3, 36

In this dataset 2 and 3 both occur 4 times so it is bimodal having two mode values i.e. 2 and 3. For grouped data, mode can be computed as,

$$\text{Mode} = L + \left[\frac{\Delta_1 + h}{\Delta_1 \, \Delta_2}\right]$$

Modal class is defined as the class having highest frequency.

L = lower limit of modal class

h = size of modal class

f = frequency of modal class

$\Delta_1 = f - f_{-1}$

$\Delta_2 = f = f_{+1}$

f-1 = frequency of class preceding the modal class

f+1 = frequency of class succeeding modal class

For grouped data, the relation between mean, median and mode is given by, Mode = 3 Median – 2 Mean

Solved Examples

Example: Find the mode for the following frequency and data

Class Interval	Frequency
0-2	11
2-4	14
4-6	17
6-8	08
8-10	03

Table 5.2

Solution: The class 4-6 has the highest frequency of 17 and thus it is the modal class.

Lower Limit = 4

$f_{-1} = 14$

$f_{+1} = 8$

$\Delta_1 = f - f_{-1} = 17 - 14 = 3$

$\Delta_2 = f - f_{+1} = 17 - 8 = 9$

$$M_0 = 4 + \left[\frac{3}{3+9}\right]2 = 4 + \frac{6}{12} = 4.5$$

Note: If all the frequencies are equal mode is undefined $\left(\frac{0}{0}\text{ form}\right)$

Measures of Dispersion / Variability

To check the consistency, uniformity etc of the data measures of dispersion is used. They are

- Range
- Quartile Deviation (QD)
- Mean Deviation (MD)
- Standard Deviation (SD)
- Coefficient of Deviation (c, u)

Range: It is defined as the difference between the maximum and minimum values.

Standard Deviation

Instead of taking absolute deviation from the mean of the data, we square each deviation and obtain arithmetic mean of squared deviations. The mean of squared deviations is called as Variance of the data and square root of variance is called as standard deviation.

Standard Deviation SD= $\sqrt{\text{Vanance}}$

Variance = $(SD)^2$

Standard Deviation, $\sigma_x = \sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$

Lesser value of standard deviation or variance indicates that data is more uniform. Thus, the variance of constant is 0.

Note:

- Variance can never be negative.
- Sum of the deviations from the mean is always zero. $\sum\left(x_i - x\right) = 0$
- Sum of the squares of the deviation from mean is minimum.
- If the variance are equal for the different groups greater mean is more consistent
- For grouped data, variance is given by

$$\sigma_x^2 = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i^2\right)^2$$

- Relation between quartile deviation, mean deviation and standard deviation
  6QD=5MD=4SD
  QD = $\frac{4}{6}$ SD = $\frac{2}{3}$ σ
  5MD=4SD
  MD = $\frac{4}{5}$ SD
- Comparison of variability of two sets of data is done in terms of coefficient of variation.
  Coefficient of variation,

$$C.V = \frac{\text{Standard Deviation}}{\text{mean}} \times 100 = \frac{\sigma}{\mu} \times 100$$

- Coefficient of variation is the best measure to determine data consistency.

Solved Examples

Example: Find the mean and variance of 1st 'n' natural numbers?

$$\bar{x} = \frac{1+2+3+n}{n}$$

$$= 1/n\left[\frac{n(n+1)}{2}\right] = \frac{n+1}{2}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(xi)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}2x\right)^2$$

$$\frac{1}{n}\sum xi^2 = \frac{1}{n}\left[12 + 22 + \ldots\ldots\ldots\ldots\ldots + n^2\right]$$

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

$$\sigma^{x_2}\frac{(n+1)(2n+1)}{6} = \left(\frac{n+1}{2}\right)^2$$

$$\sigma^{x_2} = \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n+1}{2}\left[\frac{4n+2-3n-3}{6}\right]$$

$$\sigma^{x_2} = \frac{n+1}{2}\left[\frac{n-6-n^2-2^1}{6}\right]$$

Skewness

It is the measurement of lack of symmetry. Pearson's Coefficient of skewness.

$$S_{kp} = \frac{M - M_0}{\sigma}\left(\frac{3\,\text{Mean} - \text{Median}}{\sigma}\right)$$

$$-3 \le S_{kp} \le +3$$

If $-3 \le S_{kp} \le 0$, the data is said to be negatively skewed.

S_{kp} =0, the data is said to be symmetric

0 ≤ S_{kp} 3, the data is said to be positively skewed.

Random Variables

It is frequently the case when an experiment is performed that we are mainly interested in some function of the outcome as opposed to the actual outcome itself.

For instances, in tossing dice we are often interested in the sum of two dice and are not really concerned about the separate value of each die. That is, we may be interested in knowing that the sum is 7 and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). These quantities of interest, or more formally, these real valued functions defined on the sample space, are known as random variables.

Because the value of the random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Connecting the outcomes of an expectation with the real values is known as random variable (R-V) (1-D, R-V). The corresponding data is known as univariate data.

2-D R-V i.e. Two dimensional random variable is defined as connecting the two outcomes to a real value provided those two outcomes are drawn from the same sample space. The corresponding data is known as bivariate data.

Types of Random Variable

Random variable may be discrete or continuous.

Discrete random variable: A variable that can take one value from a discrete set of values.

Let x denotes sum of 2 dice. Number x is a discrete random variable as it take one value from the set {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}, since the sum of 2 dice can only be one of these values.

Continuous random variable: A variable that can take one value from a continuous range of values. Let X denotes the volume of Pepsi in a 500ml cup. Now x may be a number from 0 to 500, any of which value, x may take.

Note: Similar to Random Variables we can define Discrete and Continuous Distributions.

Properties of Discrete Distribution

- $\sum P(x) = 1$
- $E(x) = \sum xP(x)$
- $V(x) = E(x^2) - (E(x))^2 = \sum x^2P(x) - \left[\sum xP(x)\right]^2$

E(x) denotes the expected value or average value of the random variable x, while V(x) denotes the variance of the random variable x.

Probability

Distribution and Cumulative Distribution Function

The probability distribution function or the probability density function represents the distribution of probability

i.e. how the probability varies as the random variable takes on different values. It is represented as f(x). Cumulative Distribution Function represents the probability that the random variable is less than a certain value.

$\frac{d}{dx} F(x) = f(x)$ Probability Density Function (PDF)

$F(x) = \int_{-\infty}^{x} f(x)dx =$ Cumulative Distribution Function (CDF)

Note: Both these terms are defined for continuous distributions.

Properties of Continuous Distribution

- $\int_{-\infty}^{\infty} f(x)dx = 1 \text{ or } F(\infty) = 1$
- $E(x) = \int_{-\infty}^{\infty} xf(x)dx$
- $E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)f(x)dx$
- $V(x) = E(x^2) - [E(x)]^2$

  $= \int_{-\infty}^{\infty} x^2f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2$
- $P(a < x < b) = P(a \le x < b)$

  $= P(a < x \le b) = \int_{a}^{b} f(x)dx$

Properties of Expectation

- If ‘X’ is a random variable and ‘a’ constant E(ax) = aE(x)
- if X and Y are two random variables

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

  E(X - Y) = E(X) – E(Y)

  E(X . Y) = E(X) E(Y | X)= E(Y) E(X | Y)

Here, E(Y | X) and E(X | Y) are called as Conditional Expectations.

- If X and Y are independent Random Variables E(X. Y) = E(X) E(Y)
- If Y = aX + b, a, b → constants

  E(Y) = E(aX + b) = E(aX) +E(b)= a E(X) + b

  E(constant) = constant
- E(E(E(x))) = E(x)

Properties of Variance

- If X is a Random Variable and 'a' is constant $V(aX) = a^2 v(X)$

  V(-Y) = (-1)2 V(Y) = V(Y)
- If X and Y are independent Random Variables V(X + Y) = V(X) + V(Y)

  V(X – Y) = V(X) + V(-Y) = V(X) + V(Y)

  V(X.Y) = V(X) .V(Y)
- If a and b are constants, X and Y are independent Random Variables

  $$V\left(\frac{x}{a}-\frac{Y}{b}\right)=\frac{1}{a^2}V(X)+\frac{1}{b^2}V(y)$$
- If Y = aX + b, where a, b are constants

  V(Y) = V(aX+b) = V(aX) + V(b) = a2V(x) because V(constant)=0
- If X and Y are two Random Variables V(X . Y) = V(X) +V(Y) + 2cov(X, Y)

  cov(X, Y) = E(X.Y) – E(X) E(Y)

  cov(a, b) = E(ab) – E(a) E(b) = ab – ab = 0 where a and b → constant
- If x and y are independent Random Variables, covariance of x and y cov(x, y)=0. But the converse of statement is not true.

Thus, E(X . Y) = E(X).E(Y)

Note:
- Variance and covariance are independent of change of origin and dependent on change of scale.
- Expectation (mean) dependent on change of origin as well as dependent on change of scale.

Skewness

Skewness is defined as the lack of symmetry.

$$\beta_1=\frac{\mu^2{}_3}{\mu^3{}_2}$$

μ_3 : 3^{rd} central moment $\mu_3=E\left((x-\mu)^3\right)$

$\mu 2$: variance

- If μ_3 value is negative, then the curve is known as negatively skewed.
- If μ_3 value is positive, then the curve is known as positively skewed.
- If $\mu_3=0$, then the curve is known as symmetric β_1 = 0

Solved Examples

Example: Find the expectation of the number on a die when it is thrown?

x R.V	1	2	3	4	5	6
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Solution: E(X) = mean

$$=\sum_{x=1}^{6}xP(x)=$$

$$1\times P(1)+2P(2)+3P+4P(4)+5P(5)+6P(6)$$

$$E(x)=\frac{1}{6}[1+2+3+4+5+6]=\frac{21}{6}=\frac{7}{2}$$

Example: Find the variance of outcome of a single dice throw.

$$\text{Solution}: V(x)=E(x^2)-[E(x)]^2$$

$$E(x^2)=\sum_{x=1}^{6}x^2P(x)=\frac{1}{6}$$

$$[1^2+2^2+3^2+4^2+5^2+6^2]=\frac{91}{6}$$

$$V(x)=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{91}{6}-\frac{49}{4}=\frac{182-147}{12}=\frac{35}{12}$$

Note: The mean and variance for the sum of the numbers on the dice is

$$E(x)=\frac{7n}{2}$$

$$V(x)=\frac{35n}{12}$$

where, n represents the number of dice thrown.

To derive this, we need to construct a probability distribution table, where we list down the probability of occurrence of each value of sum. As for an example, for two dice, the sum can vary from 2 to 12.

Example: 3 unbiased dice are thrown. Find the mean and variance of the sum of the numbers on them?

Solution:

$$E(x)=\frac{7}{2}\times=\frac{21}{2}$$

$$V(x)=\frac{35}{12}\times 3=\frac{35}{4}$$

Example: When two dice are rolled, find the expectation for sum 7?

Solution: E(x) = x.P(x)

The sum 7 can be obtained for the following outcome (1,6); (2,5); (3,4); (4,3); (5,2); (6,1). Thus, 6 cases out of 36 yield the sum of 7.

$$E(7)=7.P(7)=7.\frac{6}{36}=\frac{7}{6}$$

Example: A player tosses 3 coins. He wins Rs.500 if 3 heads occur. Rs.300 if 2 heads occur, Rs.100 if only 1 head occur. On the other hand, if 3 tails occur he losses Rs.1500. Find expected value of the game?

Solution: Constructing the Probability distribution table,

x	+500	+300	+100	-1500
No. of head	3	2	1	${}^{3}C_{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{3}$
P(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
	${}^{3}C_{3}\left(\frac{1}{2}\right)^{3}$	${}^{3}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{1}$	${}^{3}C_{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{2}$	

Value of the game = Gain – loss

$$\text{Value}=500\times\frac{1}{8}+300\times\frac{3}{8}+100\times\frac{3}{8}-1500\times\frac{1}{8}$$

$$=\frac{1700}{8}-\frac{1500}{8}=\text{Rs. }25$$

Note: If the game is said to be balanced or fair, then the value of game = 0 (No loss and no gain)

Example: A man was given n keys of which 1 fits the lock. He tries them successively without replacement to open the lock. What is the probability that the lock will be open after the nth trial. Also determine mean and variance?

Solution: There can be two cases for opening the lock, With replacement $\Rightarrow$ Independent trials

Without replacement $\Rightarrow$ dependent trials

P(opening lock), 1st trial = $\frac{1}{n}$, Since there are n keys available

P(opening lock), 2st trial = $\frac{1}{n-1}$, Since there are (n-1) keys available

P(opening lock), 3rd trial = $\frac{1}{n-2}$, Since there are (n-2) keys available

P(opening lock, 1st failure, 2nd trial)

$$\left(1\frac{1-1}{n-1}\right)=\frac{n-1}{n}\times\frac{1}{n-1}=1$$

P(opening lock, 1st failure, 2nd failure 3rd trial)

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

Similarly, P(opening lock 1st success, 9th trial) = $\frac{1}{n}$

Thus, the probability of success on each trial is the same and we can call it uniform distribution.

$$E(x)=\sum_{i=1}^{n} i\times P(x=i)=\sum_{i=1}^{n} i\times\frac{1}{n}=\left(\frac{n-n+n}{2n}\right)$$

$$=\frac{n+1}{2}$$

Similiarly, $V(x)=\frac{n^2-1}{12}$

Example: A man was given 900 keys. Find the probability that the lock will be open at 450th trial by with and without replacement.

Solution: From the previous example,

450th trial lock open without replacement= $\frac{1}{n}=\frac{1}{900}$

Lock open 450th trial with replacement, then there must be failure on previous 449 trials and since the trials are with replacement, the number of keys in each trial is 900.

P (success) $= \frac{1}{900}$

P (failure) = 1-1/900 $\frac{1}{900} = \frac{899}{900}$

P(450th trial with replacement)

$$= \left(1 - \frac{1}{900}\right)^{499} \times \frac{1}{900} = 6.74 \times 10^{-4}$$

Example: If X is continuous Random Variable and $f(x) = ke^{-x^2}/2 \quad -\infty < X < +\infty$

Find

(A) K

(B) E(x), V(x)

Solution: Probability of sample space is 1 or the total probability is always one.

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

$$\int_{-\infty}^{\infty} Ke^{-x^2/2}dx = 1$$

$$K\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = 1$$

Since, the integrand is an even function,

$$2K\int_{0}^{\infty} e^{-x^2/2}\,dx = 1$$

$\text{Put } x^2 = 2u$

$2xdx = 2du$

$$dx = \frac{2du}{2x} = \frac{du}{\sqrt{2u}}$$

$$2K\int_{0}^{\infty} e^{-u}u^{-1/2}\frac{du}{\sqrt{2}} = 1$$

$$\sqrt{2K}\int_{0}^{\infty} e^{-u}\,u^{-1/2}\,du = 1$$

$$\sqrt{2K}\int_{0}^{\infty} e^{-u}\,u^{1/2-1}\,du = 1$$

By definition, gamma function is defined as,

This is due to the fact that integrand is an odd function and hence the integration in positive and negative limits cancels out.

$$\overline{|n} = \int_{0}^{\infty} e^{-u}u^{n-1}\,du$$

Therefore, $\sqrt{2K}\int_{0}^{\infty} e^{-u1} - u\,2^{-1}du = \sqrt{2K}\overline{\left|\frac{1}{2}\right.} = 1$

$$\sqrt{2K}\sqrt{\pi} = 1$$

$$K = \frac{1}{\sqrt{2\pi}}$$

Calculating expectation and variance of the random variable,

$$E(x)\int_{\infty}^{0} xf(x)\,dx = \frac{-1}{\sqrt{2□}}\int_{-\infty}^{\infty} x.^{-x^2}x^{-2}\,dx = 0$$

This is due to the fact that integrand is an odd function and hence the integration in positive and negative limits cancels out.

$$E(x) = \int_{-\infty}^{\infty} x.^2\,f(x)\,dx = \frac{1}{\sqrt{1Å}}\int_{-\infty}^{\infty} x^{2-x^2}e^{-2}dx$$

Again substitute $x^2 = 2^1$ ud

$$\left(Ex^2\right) = \frac{\sqrt{2ue^{-u}}}{\sqrt{\pi}}\int_{0}^{\infty}\frac{2}{\sqrt{2}} =$$

$$\frac{2}{\sqrt{\pi}}\int_{0}^{\infty} u^{-2}e\ du\frac{2}{\sqrt{□}}\overline{\left|\frac{3}{2}\right.} \times \frac{2}{\sqrt{\pi}} \times \frac{1}{2} \times \overline{\left|\frac{1}{2}\right.} = 1$$

$$V(x) = \left(x^2\right) - \left[E(x)\right]^2 = 1 - 0 = 1$$

Example: If X is a continuous Random Variable and f(x) = |x|, =1 < x < 1 find V(X)?

Solution: $E(x) = \int_{-1}^{1} xf(x)\,dx = \int_{-1}^{1} x|x|\,dx = 0$

Because, here the integrand is an odd function of x. $\left[\frac{x^4}{4}\right]_0^1 = \frac{2}{4} = \frac{1}{2}$

$$E(x^2) = \int_{-1}^{1} x^2 f(x)\,dx = \int_{-1}^{1} x^2\,|x|\,dx = 2\int_{0}^{1} x^2.x\,dx = 2$$

$$V(x) = E(x^2) - (E(x))^2 = \frac{1}{2} - 0\frac{1}{2}$$

Example: If x and y are the R.V, mean(x) = 10, variance (x) = 25, find +ve values of a, b such that Y=aX-b has E(Y)

= 0, V(Y) = 1

Solution: Variance of the two random variables are related as,

$$V(Y) = a^2 V(X)$$
$$1 = 25a^2$$
$$a^2 = \frac{1}{25}$$
$$a = \frac{1}{5}$$
$$\text{Since}, Y = aX - b$$
$$E(Y) = aE(X) - b$$
$$0 = 10a - b$$
$$b = 10a = 10 \times \frac{1}{5} = 2$$

Bivariate Data

Case 1: Continuous Random Variable

If x and y are 2-D continuous Random Variables and its probability function is known as joint probability density function and is denoted by f(x, y).

The marginal density functions are

$$f(x) = \int_y f(x,y)\,dy$$
$$f(y) = \int_x f(x,y)\,dy$$

The marginal density functions represent the 1-D density function for 2-D random variable.

If X and Y are 2-D, continuous, independent Random Variables if f(x,y) = f(x).f (y) is

JPDF = mdf (x) mdf (y)

Where JPDF = Joint Probability Density Function MDF = Marginal Density Function

Relation between JDF and JPDF

$$\frac{d^2}{d \times dy} F(x,y) = f(x,y)$$
$$\text{or} F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dxdy$$

Conditional PDF is $f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f(y)} \left(d(y) \neq 0\right)$

Conditional expectation

$$E\left(\frac{x}{y}\right) = \frac{E(xy)}{E(y)} \left(E(y) \neq 0\right)$$

Case-2: Discrete Random Variable

If x and y are 2-D discrete Random Variable and its probability function is known as joint probability mass function denoted by P(x y).

The marginal mass function is

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

Covariance

Covariance is a measure of the joint variability of two random variables For Discrete random variables

$$\text{Cov}(X,Y) = \frac{\sum_{i=1}^{n} (x_i - \cancel{x})(y1 - \cancel{y})}{n}$$

For Continues real valued random variables

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

Solved Examples

Example: If x and y are 2-D continuous Random variables and their corresponding joint probability function is

f (x, y) = xy . x → 0 to 1, y → 0 to 1

i) Find the mean and variance of y

ii) E(x, y), Cov(x, y)

iii) f(x/y) = E(x/y)

iv) Check whether x and y are independent variable or not

Solution: The marginal PDF can be obtained from Joint PDF as,

$$f(x) = \int_y f(x,y)\,dy = \int_0^1 xydy = x\int_0^1 ydy = x\left[\frac{y^2}{2}\right]_c^1$$
$$= \frac{x}{2}$$
$$\text{Similarly}, f(y) = \frac{y}{2}$$

$$E(X) = \int_0^1 x f(x) dx = \int_0^1 x.\frac{x}{2} dx = \frac{1}{6}$$

$$E(Y) = \frac{1}{6}$$

Variance can be calculated as $V(Y)$

$$V(Y) = E(Y^2) - [E(Y)]^2$$

$$E(y^2) = \int_0^1 y^2 f(y) dy = \int^1 yy^2 \frac{y}{2} dy = \frac{1}{8}$$

$$V(Y) = \frac{1}{8} - \left(\frac{1}{6}\right)^2 = \frac{1}{8} - \frac{1}{36} = \frac{9-2}{72} = \frac{7}{72}$$

$$E(XY) = \int_0^1\int_0^1 xyf(x,y) dxdy = \int_0^1\int_0^1 x^2y^2 dxdy$$

$$= \int_0^1 x^2 \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{9}$$

Conditional PDF and Conditional Expectation are,

$$f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f(y)} = \frac{xy}{y/2} = 2x$$

$$E\left(\frac{x}{y}\right) = \frac{E(XY)}{E(Y)} = \frac{1/9}{1/6} = \frac{2}{3}$$

Since,

$$f(x,y) \neq f(x) f(y)$$

$$xy \neq \frac{x}{y} \cdot \frac{y}{2}$$

$\therefore$ x and y are dependent Random Variable

Example: If x and y are 2-D discrete R.V, the corresponding probability function is

xy	- 1	0	+ 1
-1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Find $P\left(\frac{x+y=2}{x}\right)$

Solution:

$$P\left(\frac{x+y=2}{x-y=0}\right) = \frac{P(x+y=2 \cap x-y=0)}{P(x-y=0)}$$

$$P\left(\frac{x+y=0}{x-y=0}\right) =$$

$$\frac{P(x=1, y=1)}{P(x=-1, y=-1) + P(x=0, y=0) + P(x=1, y=1)}$$

$$P\left(\frac{x+y=2}{x-y=0}\right) = \frac{1}{4} - \frac{1-1-1}{4+2+4} = \frac{1}{4}$$

Binomial distribution

Suppose that a trial or an experiment, whose outcome can be classified as either a success or a failure is performed. As an example, when we toss a coin we can call occurrence of Head as success and occurrence of tail as failure.

Suppose now that n independent trials, each of which results in a success with probability p and in a failure with probability 1-p are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be binomial random variable with parameters (n, p).

The Binomial distribution occurs when the experiment performed satisfies the three assumptions of bernoulli trials, which are:

1. Only 2 outcomes are possible, success and failure
2. Probability of success (p) and failure (1-p) remains the same from trial to trial
3. The trials are statistically independent. i.e. The outcome of one trial does not influence subsequent trials. i.e., No memory.

The Probability Mass Function for x-success in n-trials is,

$P(x) = {}^nC_x p^x (1-p)^{n-x}$

Properties

$$E(x) = \text{mean} = np$$

$$V(x) = \mu_2 = npq$$

$$\mu_s = npq(q-p) = npq(1-2p)$$

$$\text{Skewness, } \beta_1 = \frac{\mu^{3^2}}{\mu_2} = \frac{n^2p^2q^2(1-2p)^2}{n_3p_3q_3} = \frac{(1-2p)^2}{npq}$$

Moment generating function

$m_x(t) = E(e^{tx}) = (q+pe^t)^n$

Characteristic Function

$\phi_x = E(e^{itx}) = (q+pe^{it})^n$

> **Note:**
>
> For $p = \frac{1}{2}$, then the binomial distribution is symmetric (μ_3=0).
>
> For $p < \frac{1}{2}$, $\Rightarrow \mu_3$ is positive, then the curve is positively skewed.
>
> For $p < \frac{2}{2}$, $\Rightarrow \mu_3$ is negative, then the curve is negatively skewed.

Sum of the independent binomial Random Variables is also a Binomial Random Variable.

The moment generating function is used to find addition and difference between Random Variable with their probability function.

The characteristic function is used for finding the convolution between the Random Variable and ratio between the Random Variable.

Solved Examples

Example: Find the probability of getting a 9 exactly twice in 3 times with a pair of dice?

Solution: Number of Trials, n = 3

Number of Success, x = 2

The sum can be 9 for the following combinations (3, 6); (4, 5); (5,4); (6, 3)

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

$$q=\left(1-\frac{1}{9}\right)=\frac{8}{9}$$

$$P(x=2)={}^3C_2\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)=\frac{8}{243}$$

Example: The probability of a man hitting the target is $\frac{1}{3}$.

(i) If he fires 5 times, what is the probability of his hitting the target at least twice.

(ii) How many times must he fire so that the probability of his hitting the target at least once is more than 90%.

Solution: The probability of success is, $p = \frac{1}{3}$

Thus, the probability of failure is, $q = (1-p) = \frac{2}{3}$

(i) Number of Trials, n = 5

$$P(x \ge 2) = 1 - P(x <_5 2) =$$

$$1-\left[P_1\left(x =_4 0\right)+P\left(x=1\right)\right]$$

$$P(x\ge 2)=1-\left[\left(\frac{2}{3}\right)+{}^5C_1\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\right]=\frac{131}{243}$$

(ii) $P(x \ge 1) > 90\%$

$$1-P(x=0)>0.9$$

$$P(x=0)<0.1$$

$${}^nC_0\,P_n^0 q^n < 0.1$$

$$\left(\frac{2}{3}\right) < 0.1$$

$$n > 5.67 = 6$$

Example: Two dice are rolled 120 times. Find the average number of times in which the number on the first dice exceeds the number on 2nd dice.

Solution: Number of Trials, n = 120

The cases in which number of first dice exceeds the number on 2nd dice

(2, 1); (3, 1); (3, 2); (4, 1); (4, 2); (4, 3); (5, 1); (5, 2); (5, 3); (5, 4); (6, 1); (6, 2); (6, 3); (6, 4); (6, 5);

Thus, there are 15 cases out of 36 in which we have success

Probability of Success,

$$P=\frac{15}{36}=\frac{5}{12}$$

$$E(x)=np=120\times\frac{15}{36}=50$$

Example: If x and y are the binomial Random Variables x is B(2, P), y is B(4, P) if $P(x \ge 1) = \frac{5}{9}$ find $P(y \ge 1)$?

Solution: The number of trials for both experiments,

$$n_x = 2 \text{ and } n_y = 4$$

$$\text{Given}, P(x \geq 1) = \frac{5}{9}$$

$$P(x \geq 0) = 1 - P(x = 0) = \frac{5}{9}$$

$$\text{Thus}, P(x = 0) = \frac{4}{9}$$

In terms of Binomial Random Variable,

$$P(x = 0) = {}^nC_0 p^0 q^n$$

$$\text{Therefore}, q^n = \frac{4}{9}$$

$$\text{For 2 trials}, q^2 = \frac{4}{9}$$

$$q = \frac{2}{3} \Rightarrow P = \frac{1}{3}$$

$$P(y \geq 1) = 1 - P(y = 0) = 1 - q^n = 1 - \left(\frac{2}{3}\right)^4$$

$$= 1 - \frac{16}{81} = \frac{65}{81}$$

Example: If x is a binomial Random Variable, then find the value of $\sum_{x=0}^{n} \frac{x}{n} {}^nC_x p^x q^{n-x}$

Solution: $\sum_{x=0}^{n} \frac{x}{n} {}^nC_x p^x q^{n-x} = \frac{1}{n} \sum_{x=0}^{n} x \, {}^nC_x p^x q^{n-x}$

In terms, of Binomial Random Variable,

$$\frac{1}{n} \sum_{x=0}^{n} x \, {}^nC_x p^x q^{n-x} = \frac{1}{n}\left[\sum_{x=0}^{n} x.P(x)\right] = \frac{1}{n} E(x) = \frac{1}{n} np = p$$

Example: If x is a binomial Random Variable and E(x) = 4, V(x) = $\frac{4}{3}$. Find P (x ≤ 2)

Solution:

$$E(x) = np = 4$$

$$V(x) = npq = 4q = \frac{4}{3}$$

$$\therefore q = \frac{1}{3}$$

Thus, Probability of Success, p = $\frac{2}{3}$

np = 4

Thus, the number of trials, $n = \frac{3}{2} \times 4 = 6$

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2) = \left(\frac{1}{3}\right)^6$$

$$+ {}^6C_1 \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right) + {}^6C_2 \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)2$$

$$P(x \leq 2) = \frac{1}{729}[1 + 12 + 60] = \frac{73}{729}$$

Example: There are 10 markers on a table, of which 6 are defective and 4 are not defective. If 3 are randomly taken from the above lot, what is the probability that exactly 1 of markers is defective?

Solution: Random Variable, X represents the number of defected markers from the selected lot. D is defective and ND is non defective.

$$P(x = 1) = \frac{6C_1 \times 4C_2}{10C_3} = 0.3$$

X can now take the values 0, 1, 2 or 3.

$$P(X = x) = \frac{6C_x \times 4C_{3-x}}{10C_3}$$

From the above formula, we can calculate the following:

$$P(X = 1) = \frac{6C_1 \times 4C_2}{10C_3}$$

$$P(x \geq 1) = P(x = 0) + P(x = 1) = \frac{6C_0 \times 4C_3}{10C_3}$$

$$+ \frac{6C_1 \times 4C_2}{10C_3}$$

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \left[\frac{6C_0 \times 4C_3}{10C_3}\right]$$

Poisson Distribution

If a Random Variable can take the values from the set of Natural Numbers, then it is modeled as a Poisson Random Variable. It can be used to model the,

- Arrival Rate
- Rare occurrence
- Defect probability
- Evolutionary process

If x is a Poisson Random Variable defined in the interval 0 →∞ with a parameter λ (>0) and its probability mass

$$\text{function is } P(x:\lambda>0)=P(x)=\begin{cases}\dfrac{e^{-\lambda x}}{x!} \\ 0\end{cases}$$

$\lambda > 0$

$0 \le x \le 0$ other wise

Conditions for Poisson Random Variable

- Observation are infinitely large $(n \to \infty)$
- Probability of success is very small $(p \to 0)$
- $np = \lambda \Rightarrow p = \dfrac{\lambda}{n}$
- $p(x:np) = \dfrac{e^{-np}(np)^x}{x!}$

Poisson Random Variable is an approximation of Binomial Random Variable as the number of observations or trials tend to infinite.

Properties

- E(x) = mean = λ
- V(x) $\mu_2=\lambda$
- $\mu_3=\lambda$
- Skewness, $\beta_1 = \dfrac{\mu_3^{\,2}}{\mu_2^{\,3}} = \dfrac{\lambda^2}{\lambda^3} = \dfrac{1}{\lambda}$
- Moment Generating Function, $m_x(t) = e^{\lambda(e^t-1)}$
- Characteristic Function, $\phi_x(t) = e^{\lambda(eit-1)}$

Note:

- In Poisson distribution mean = variance = parameter = λ
- It is always positively skewed.
- Sum of the independent Poisson Random Variable is also Poisson Random Variable
- The difference between the independent Poisson Random Variable is not a Poisson Random Variable.

Solved Examples

Example: A telephone switchboard receives 20 calls on an average during an hour. Find the probability that for a period of 5 minutes

i) no call is received

ii) exactly 3 calls are received

iii) Atleast 2 calls are received

Solution: Average number of calls in a period of 5 minutes,

$$\lambda = \frac{20\times 5}{60} = 1.67$$

i) $P(x=0 = \dfrac{e^{-1.67}(1.67)^0}{0!} = e^{-1.67})$

ii) $P(x=3) = \dfrac{e^{-1.67}(1.67)^3}{3!}$

iii) $P(x\ge 2) = 1 - P(x<2) = 1 -$

$[P(x=0)+P(x=1)]$

$$P(x\ge 2) = 1 - \left[e^{-1.65} + (1.65)e^{-1.65}\right] = 0.491$$

Example: If x and y are two independent Poisson Random Variable such that P(x=1)=P (x=2)=P & P (y=2)=P (y=3), Find V (3x-4y)?

Solution: It is given that, P(X = 1) = P (X = 2)

Thus, $\dfrac{e^{-\lambda 1}}{1!} = \dfrac{e^{-\lambda 2}}{2!}$

$1 = \dfrac{\lambda}{2} => \lambda = 2$

Also, $P(Y=2) = P(y=3)$

Thus, $\dfrac{e^{-\theta}\theta^2}{2!} = \dfrac{e^{-\theta}\theta^3}{3!}$

$1 = \dfrac{\theta}{3} => \theta = 3$

Thus, $E(X) = V(X) = 2$ and $E(Y) = 3$

$V(3x-4y) = 3^2V(x) + (-4)^2 V(y)$

$= 9\times 2 + 16\times 3 = 66$

Example: If x_1 and x_2 are two independent Random Variables with variances 1, 2. Find $P(x_1 + x_2 = 4)$

Solution:

$$P(x_1 + x_2 = k) = \frac{e^{-\lambda_1\lambda_2}\left(\lambda_1 + \lambda^{2k}\right)}{k!}$$

Since sum of Poisson Random Variable is a Poisson Random Variable

$\lambda_1 = 1, \lambda_2 = 2$

$$P(x_1 + x_2 = 4) = \frac{e^{-(1+2)}(1+2)^4}{4!} = \frac{e^{-3}3^4}{4!}$$

Example: If x is a Poisson Random Variable then find the value of $\sum_{x=0}^{\infty} \frac{x}{\lambda} \frac{e^{-\lambda}\lambda^x}{x!}$

Solution:

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

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

Example: If x is a Poisson Random Variable and $E(x^2) = 6$ find V(x).

Solution: Variance can be expressed as,

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

$$\lambda = 6 - \lambda^2 \qquad \because v(x) = E(x) = \lambda$$

$$\lambda^2 + \lambda - 6 = 0$$

$$(\lambda + 3)(\lambda - 2) = 0$$

$$\lambda = -3, 2$$

$$\lambda = 2 \Rightarrow v(x) = 2$$

(Because variance can never be a negative value)

Uniform Distribution

In general we say that X is a uniform random variable on the interval (a, b) if its probability density function is given by:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \text{ otherwise} \\ 0 & \end{cases}$$

Since f(x) is a constant, all values of x between α and β are equally likely (uniform).

$$\text{Mean} = E[X] = \frac{(\beta + \alpha)}{2}$$

$$\text{Variance} = v(X) = \frac{(\beta + \alpha)^2}{12}$$

Solved Examples

Example: If X is uniformly distributed over (0, 10), calculate the probability that:

a) X < 3

b) X > 6

c) 3 < X < 8

Solution: The value of Probability Density Function is given by,

(i) $P\{x < 3\} = \int_0^3 \frac{1}{10}dx = \frac{3}{10}$

(ii) $P\{x < 6\} = \int_6^{10} \frac{1}{10}dx = \frac{4}{10}$

(iii) $P\{3 < X < 8\} = \int_3^8 \frac{1}{10}dx = \frac{1}{2}$

Exponential Distribution

A continuous random variable whose probability density function is given for some $\lambda > 0$ by

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

Is said to be exponential random variable with parameter λ. The cumulative function F(a) of an exponential random variable is given by:

$$F(a) = P(x \leq a) = \int_0^a \lambda e^{-\lambda x}dx$$

$$= (e^{-\lambda x})_0^a = 1 - e^{-\lambda a} \quad a \geq 0$$

$$\text{Mean} = E[X] = \frac{1}{\lambda}$$

$$\text{Variance} = v(x) = \frac{1}{\lambda^2}$$

Solved Examples

Example: Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait.

(a) More than 10 times

(b) Between 10 and 20 minutes

Solution: Letting X denote the length of the call made by the person in the booth, we have the desired probabilities:

(a) $P\{x > 10\} = 1 - P(x < 10) = 1 - F(10)$

$$= 1 - \left(1 - e^{-\lambda \times 10}\right)$$

$$P\{X > 10\} = e^{-10\lambda} = e^{-1} = 0.368$$

(b) $P\{10 < X < 20\} = F(20) - F(10)$

$$= \left(1 - e^{-20\lambda}\right) - \left(1 - e^{-10\lambda}\right)$$

$$P\{10 < X < 20\} = e^{-1} - e^{-2} = 0.233$$

Normal Gaussian Distribution

It is called parent distribution because it covers the entire set of natural numbers.

If x is said to be a normal Random Variable defined in $(-\infty,\infty)$ with mean = μ and variance = σ^2, then its density function is

$$N(x\mu : \sigma^2,) = f(X) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}\, e} - \dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2 \\ 0 \end{cases}$$

$$-\infty < X < +\infty$$

$$-\infty < \mu < +\infty$$

$$0 < \sigma < \infty$$

otherwise

Standard Normal Random Variable

If x is a normal Random Variable with mean = 0 and variance = 1, then the Random Variable is known as standard normal Random Variable and its density function is

$$N(0,1) = f(X) = \frac{1}{\sqrt{2\pi^{e}}} - \frac{1}{2}x^2$$

Mathematically the standard normal variable is denoted by 'z' and is given by.

$$z = \frac{x - E(x)}{\sqrt{V(x)}}; -3 \le z \le +3$$

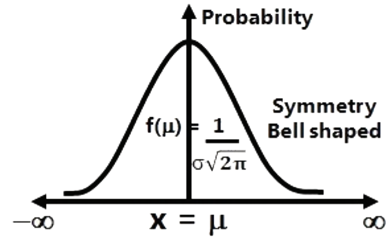

Fig. 5.2

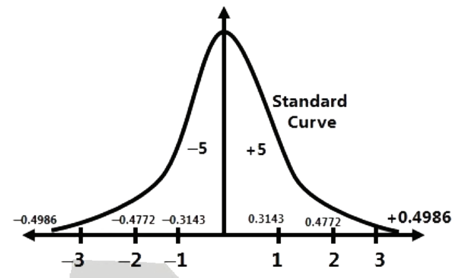

Fig. 5.3

Areas under Normal curve

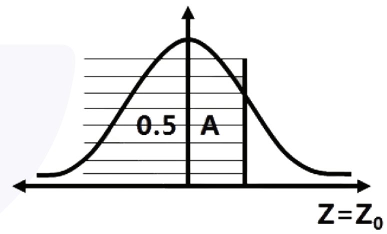

Fig. 5.4

$$P(Z \le Z_0) = 0.5 + A(0 < Z < Z_0)$$

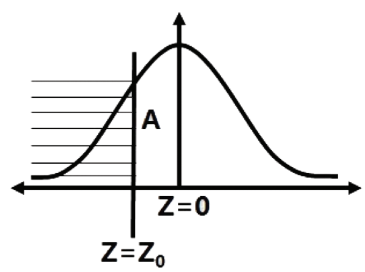

Fig. 5.5

$$P(Z \le Z_0) = 0.5 - A(Z_0 < Z < Z0)$$

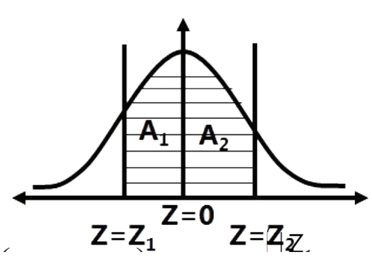

Fig. 5.6 Normal Probability distribution

$$P(Z_1 \le Z \le Z_0)^1 = A1 + A2 \begin{cases} Z_1 - ve \\ Z_2 + ve \end{cases}$$

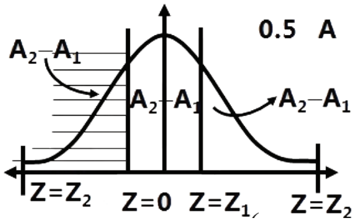

Fig. 5.7

$$P(Z_1 \le Z \le Z_0) = A_1 - A_2 \begin{cases} Z_1 \,\&\, Z_2 \\ (+\text{ve or} - v^2 e) \end{cases}$$

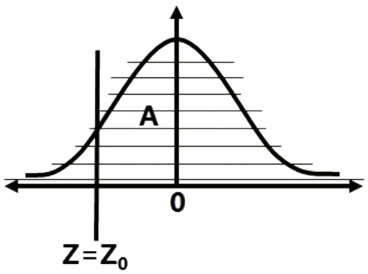

Fig. 5.8

$$P(Z \le Z_0) = 0.5 + A(Z_0 < Z < 0)$$

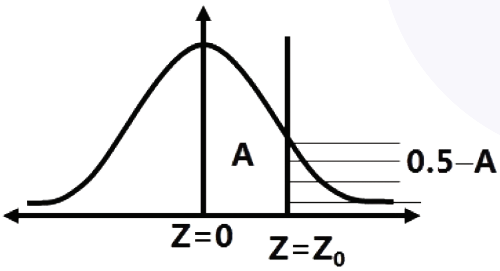

Fig. 5.9

$$P(Z \le Z_0) = 0.5 - A(0 < Z < Z_0)$$

Properties

- E(x) = mean = μ
- V(x) = $\mu 2$ = σ^2
- Third Central Moment, μ_3 = 0
- β_1 = 0, hence symmetric

Moment Generating Function: $m_x(t) = \dfrac{e^{tu+t^2\sigma^2}}{2}$

Characteristic function: $\phi_x(t) = \dfrac{e^{itu-t^2\sigma^2}}{2}$

If X is a standard Normal Variable,

$$m_x(t) = e\frac{t^2}{2} \text{ and } \phi x(t) = e - \frac{t^2}{2}$$

Sum of the independent Random Variable is also a normal Random Variable

The difference between normal Random Variable is also a normal Random Variable.

Solved Examples

Example: If 6 is distributed with S.D = 3.33, μ = 20, find the probability that 21.11 and 26.66. The area under the curve Z=0 to Z=0.33 is 0.1293 and The area under the curve Z=0 to Z=2 is 0.4772?

Solution: We need to determine P(21.11 ≤ x ≤ 26.66)

$z = 0 \text{ to } z = 0.33 \text{ is } 0.1293$

$z = 0 \text{ to } z = 2 \text{ is } 0.4772$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{21.11 - 20}{3.33} = \frac{1.11}{3.33} = 0.33$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{26.66 - 20}{3.33} = \frac{6.66}{3.33} = 2$$

$$P(21.11 \le x \le 26.66) = P(0.33 \le Z \le 2) =$$

$$P(0 \le x \le 2) - (0 \le Z \le 0.33)$$

$$P(21.11 \le x \le 26.66) = 0.4772 - 0.1293 = 0.3479$$

Example: A die is rolled 180 times, using the normal distribution find the probability that the face 4 will turn up at least 35 times. The area under the normal curve z = 0 to z = 1 is 0.3413

Solution: Number of Trials, n = 180

The probability of success, p = $\dfrac{1}{6}$

The probability of failure, q= $\dfrac{5}{6}$

Mean of the distribution = $np = \dfrac{1}{6} \times 180 = 30$

Variance of the distribution =

$$npq = \frac{1}{6} \times 180 \times \frac{5}{6} = 25$$

The variable z is,

$$= \frac{X - \mu}{\sigma} = \frac{35 - 30}{\sqrt{25}} = \frac{5}{5} = 1$$

$$P(X \ge 35) = P(Z \ge 1) = 0.5 - P(0 \ge Z \ge 1)$$

$$P(X \ge 35) = 0.5 - 0.3413 = 0.1587$$

Example: If x is normally distributed with mean = 30 and standard deviation = 5. Find P $(|x - 30| > 5)$ given that

P(0 ≤ z ≤ +1) = 0.3413?

$$P(|x-30|>5)=1-P(|x-30|\le 5)$$
$$P(|x-30|<5)=P(-5\le x-30\le 5)$$
$$=P(25\le x\le 35)$$
$$Z_1=\frac{x_1-\mu}{\sigma}=\frac{25-30}{5}=-1,$$
$$Z_2=\frac{35-30}{5}=1$$
$$P(25\le x\le 35)=P(-1\le z\le +1)$$
$$=0.3413+0.3413=0.6826$$
$$P(|x-30|>5)=1-P(|x-30|\le 5)$$
$$=1-0.6826=0.3174$$

Correlation and Regression

Correlation: The relation between the 2-D Random Variable in bivariable data is known as correlation. That means the changes in one variable is affecting the changes in other variable in parallel, then those variables are known as correlated variable.

Types of correlation

1. +ve correlation
2. -ve correlation
   a) If the changes in both variable are in same direction i.e. either both increasing or both decreasing then these variables are known as positively correlated variables.
   b) If the changes in the one variable are affecting the changes of the other variable parallel in the reverse direction i.e. increase in one variable causes decrease in the other then those variables are known as negatively correlated variables.

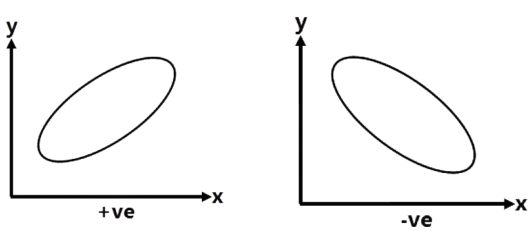

Fig. 5.10

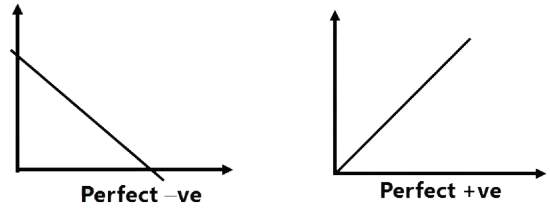

Fig. 5.11

Karl Pearson's Correlations

$$r(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}\ \text{Such that} -1\le r\le 1$$

Where $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$

Scatter Diagram

It is a graphical representation of correlation. If the points are very closer and very thick on the x-y plane, then their points are known as correlated points.

If the points are widely separated, then they are said to be uncorrelated.

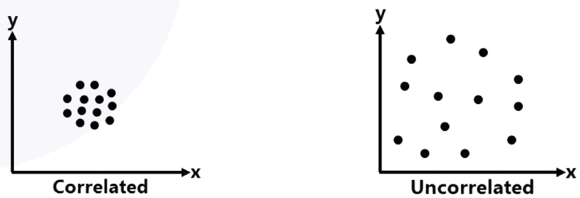

Fig. 5.12

If the two Random Variables are independent
$\Rightarrow$ cov(x, y) = 0 r(x, y) =0

They are highly uncorrelated.

Regression

The linear relationship between the 2-D Random Variables in bivariable data is known as regression. Lines of Regression

- Y on X

$$\left(y-\bar{y}\right)=r\times\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right)$$

- X on Y

$$\left(x-\bar{x}\right)=\gamma\times\frac{\sigma_x}{\sigma_y}\left(y-\bar{y}\right)=b_{xy}\left(x-\bar{x}\right)$$

Properties

Correlation coefficient is the geometric mean between

$$b_{yx}.b_{xy}=r\times\frac{\sigma_y}{\sigma_x}\times\gamma\frac{\sigma_x}{\sigma_y}=r^2 \qquad \therefore r=\pm\sqrt{b_{xy}\times b_{yx}}$$

Note:

- Both the regression coefficient must have the same sign.

  i.e if both are positive $\Rightarrow$ r is positive Both are negative $\Rightarrow$ r is negative
- If $b_{yx} > 1 \Rightarrow b_{xy} < 1$ & vice versa
- If the regression coefficient are equal $\Rightarrow$ variance also equal

$$b_{xy} = b_{yx} \Rightarrow r \times \frac{\sigma_y}{\sigma_x} = r\frac{\sigma_x}{\sigma_y}$$

$$\sigma_y{}^2 = \sigma_x{}^2$$

- Regression lines pass through the points

$$\bar{x}, \bar{y}\ \theta = \tan^{-1}\left[\frac{1-r^2 \quad \sigma_x \sigma y}{r\,\sigma_x{}^2 + \sigma_y{}^2}\right]$$

r = 0 $\Rightarrow \theta \frac{\pi}{2}$. Thus, both lines are perpendicular.

r = 1 $\Rightarrow \theta = \pi$. Thus, both lines are parallel.

Solved Examples

Example: The regression equations are

$x + 2y = 0 \,\&\, 2 + y = 1.$ Find i) r ii) $\bar{x}\,\bar{y}$

Solution: The equation for regression line of Y on X is, x + 2y = 0

$$2y = -x => y = -\frac{x}{2}$$

$$\text{Thus}, b_{yx} = -\frac{1}{2}$$

The equation for regression line of X on Y is, 2x + y = 1

$$2x = 1 - y => x = \frac{1}{2} - \frac{y}{2}$$

$$\text{Thus}, b_{xy} = -\frac{1}{2}$$

$$r = -\sqrt{\frac{1}{2} \times \frac{1}{2}} = -\frac{1}{2}$$

To determine the mean values, we will find the intersection points of both lines. Solving both equations we get,

$$\left(\bar{x}, \bar{y}\right) = \left(\frac{2}{3}, -\frac{1}{3}\right)$$

Example: The equation for the two regression lines are,

$x - 3y = 4\ (y \text{ on } x)$

$2x - y = 1\ (x \text{ on } y)$

Find i) r ii) $\bar{x}, \bar{y}$

Solution: The regression line for Y on X is, x - 3y = 4

$$-3y = 4 - x$$

$$y = \frac{-4}{3} + \frac{x}{3}$$

$$\text{Thus}, b_{yx} = \frac{1}{3}$$

The regression line for X on Y is, 2x - y = 1

$$2x = 1 + y$$

$$x = \frac{1}{2} + \frac{y}{2}$$

$$\text{Thus}, b_{xy} = \frac{1}{2}$$

$$r = \sqrt{\frac{1}{3} \times \frac{1}{2}} = \sqrt{\frac{1}{6}}$$

To determin the mean, we again find the intersection point of both lines,

$$\left(\bar{x}, \bar{y}\right) = \left(\frac{1}{5} - \frac{7}{5}\right)$$

Chebyshev's Inequality

In probability theory, Chebyshev's inequality says that the fraction of data in any distribution that lies within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$

Let x be a random variable with finite expected values μ and finite non – zero variance σ^2 . Then for any real number k > 0

$$P\left(\mu - k\sigma \le x \le \mu + k\sigma\right) \ge 1 - \frac{1}{k^2}$$

$$P\left(\left|x - \mu\right| \le k\sigma\right) \ge 1 - \frac{1}{k^2} \quad \text{for } k > 0$$

$$\text{This is also leads to}, P\left(\left|x - \mu\right| > k\sigma\right) \le \frac{1}{k^2}$$

Chapter Summary

Types of events

- **Complementary events**

  $\{E^C\} = \{S\} = \{E\}$

  The complement of an event E is set of all outcomes not in E.

- **Mutually Exclusive Events**

  Two events E & F are mutually exclusive iff $P(E \cap F) = 0$

- **Collectively exhaustive events**

  Two events E & F are mutually exhaustive $(E \cup F) = S$ Where S is sample space.

- **Independent events**

  If E & F are two independent events

  $P(E \cap F) = P(E)*P(F)$

De Morgan's Law

- $\left(\bigcup_{i-1}^{n} E_i\right)^c = \bigcap_{i-1}^{n} E^c{}_i$
- $\left(\bigcap_{i-1}^{n} E_i\right)^c = \bigcup_{i-1}^{n} E^c{}_i$

Axioms of Probability

$E_1 = E_2 = \ldots\ldots\ldots\ldots\ldots\ldots E_n$ are possible events & S is the sample space

a) $0 \le P(E) \le 1$

b) $P(S)=1$

c) $P \left(\bigcup_{i-1}^{n} E_1\right) = \sum_{i-1}^{n} P(E_i)$

for mutually exclusive events

Some important rules of probability

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

$P(A \cap B)=P(A)*P(B \mid A)=P(B)*P(A \mid B)$

$P(A \mid B)$ is conditional probability of A given B

If A & B are independent events

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

$P(A \mid B)=P(A)$

$P(B \mid A)=P(B)$

Total Probability Theorem

$P(A \cap B)=P(A \cap E)+P(B \cap E)$

$=P(A)*P(E \mid A)+P(B)*P(E \mid B)$

Baye's Theorem

$P(A \mid E)=P(A \cap E)+P(B \cap E)$

$= P(A)*P(E \mid A)+P(B)*P(E \mid B)$

Statistics

- Arithmetic Mean of Raw data

  $$\bar{x} = \frac{\sum x}{n}$$

  $\bar{x}$ = arithmetic mean; x = value of observation, n = number of observation

- Median of Raw data

  Arrange all the observation in ascending order

  $x_1 < x_2 < \ldots\ldots\ldots\ldots\ldots\ldots < xn$

  If n is odd, median

  If n is odd, median = $\frac{(n+1)}{2}$ th value

  If n is even, Median =

  $$\frac{\left(n/2\right)^{th} \text{ value} + \left(n/2 + 1\right) \text{th value}}{2}$$

Mode of Raw data.

Most frequently occurring observation in the data.

Properties of discrete distributions

a) $LP(x)=1$

b) $E(X) = LxP(x)$

c) $V(x) = E(x2) - E(x)\}$

Properties of continuous distributions

- $^{ID} f(x)dx = 1$
- $F'(x) = J f\{x\}dx \quad = \text{cumulative distributiobn}$
- $E(x) =: if(x)dx \quad = \text{expected value of } x$
- $V(x) = E(x^2) - [E(x)]^2 \quad = \text{variance of } x$

- **Properties Expectation & Variance**

  $E(ax+b) = aE(x)+b$

  $V(ax+b) = a^2V(x)$

  $E(ax_1+bx_2) = aE(x_1)+bE(X_2)$

  $V(ax_1+bx_2) = a^2V(x_1)+b^2V(x_2)$

  $Cov(x,y)=E(x\,y)- E(x)E(y)$

- **Binomial Distribution**

  No of trials =n

  Probability of failure = (1-P)

  $P(X=x)={}^nC_xP^x(1-P)^{n-x}$

  Mean = $E(X)=nP$

  Variance = $V[x]=nP(1-P)$

Poisson Distribution

A random variable x, having possible values 0, 1, 2, 3, ..., is Poisson Variable is

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

$$\text{Mean}=E(x)=\lambda$$

$$\text{Variance}=V(x)=\lambda$$

Continuous Distributions

Uniform Distribution

$$f(x)=\begin{cases}\dfrac{1}{b-a} \text{ if } a\le x\le b\\ 0 \text{ otherwise}\end{cases}$$

$$\text{Mean}=E(x)=\frac{b=a}{2}$$

$$\text{Variance}=V(x)=\frac{(b-a^2)}{12}$$

Exponential Distribution

$$f(x)=\begin{cases}e^{-\lambda x} \text{ if } x\ge 0\\ 0 \text{ if } x<0\end{cases}$$

$$\text{Mean}=E(x)=\frac{1}{\lambda}$$

$$\text{Variance}=V(x)=\frac{1}{\lambda^2}$$

Normal Distribution

$$f(x)=\frac{1}{\sqrt{2}}\exp\left(\frac{-(x-\mu)}{2\sigma^2}\right), -\infty<x<\infty$$

$$\text{Mean}=E(x)=\mu$$

$$\text{Variance}=V(x)=\sigma^2$$

Coefficient of Correlation

$$P=\frac{\text{cov}(x,y)}{\sqrt{\text{var}(x)\,\text{var}(y)}}$$

x & y are linearly related, if $p=\pm1$

x & y are un-correlated if $p=0$

Regression lines

- $(x-\bar{x})=b_{xy}(x-\bar{y})$
- $(x-\bar{y})=b_{xy}(x-\bar{x})$

Where $\bar{x}$ & $\bar{y}$ are mean values of x & y respectively

$$b_{xy}=\frac{\text{cov}(x,y)}{\text{var}(y)};\ b_{yx}=\frac{\text{cov}(x,y)}{\text{var}(x)}$$

$$P=\sqrt{b_{xy}b_{yx}}$$

Objective Questions of ESE (Prelims) EE

1. A fair dice is rolled twice. The probability that an odd number will follow an even number is

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

(B) $\frac{1}{4}$

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

(D) $\frac{1}{6}$

2. A urn contain 5 red ball and 5 black balls. In the first draw, one ball is picked at random and discarded with noticing its colour. The probability to get a red ball in the second draw is.

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

(B) $\frac{6}{9}$

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

(D) $\frac{4}{9}$

3. The probability that two friends share the same birth-month is

(A) $\frac{1}{6}$

(B) $\frac{1}{12}$

(C) $\frac{1}{144}$

(D) $\frac{1}{24}$

4. A box contains 5 black and 5 red balls. Two balls are randomly picked one after another form the box, without replacement. The probability for both balls being red is

(A) $\frac{1}{90}$

(B) $\frac{19}{90}$

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

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

5. An unbiased coin is tossed three times. The probability that the head turns up in exactly two cases is

(A) $\frac{1}{9}$

(B) $\frac{1}{8}$

(C) $\frac{3}{8}$

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

6. From a pack of regular playing cards, two cards are drawn at random. What is the probability that both cards will be Kings, if first card is NOT replaced?

(A) $\frac{1}{26}$

(B) $\frac{1}{52}$

(C) $\frac{1}{169}$

(D) $\frac{1}{221}$

7. A single die is thrown twice. what is the probability that the sum is neither 8 nor 9?

(A) $\frac{1}{9}$

(B) $\frac{1}{4}$

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

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

8. A box contain two washers, 3 nuts and 4 bolts. Items are drawn from the box at random one at a time without replacement. The probability of drawing 2 washers first followed by 3 nuts and subsequently the 4 bolts is

(A) $\frac{2}{315}$

(B) $\frac{1}{630}$

(C) $\frac{1}{1260}$

(D) $\frac{1}{2520}$

9. 9. An unbiased coin is tossed five times. The outcome of each toss is either a head or a tail. The probability of getting at least one head is:

(A) $\frac{1}{32}$

(B) $\frac{13}{32}$

(C) $\frac{16}{32}$

(D) $\frac{31}{32}$

10. Out of all 2-digit integers between 1 and 100, a 2-digit number has to be selected at random. What is the probability that the selected number is not divisible by 7?

(A) $\frac{13}{90}$

(B) $\frac{12}{90}$

(C) $\frac{78}{90}$

(D) $\frac{77}{90}$

Answer Key

1 – b	**2 – a**	**3 – b**	**4 – d**	**5 – c**
6 – d	**7 – d**	**8 – c**	**9 – d**	**10 – d**