Sampling Techniques

- Sampling
  - Probability
    - Random
    - Systematic
    - Stratified
  - Non-Probability
    - Snowball
    - Quota
    - Judgement
    - Convenience

Sampling is a statistical method that deals with the selection of individual observations within a population. It is performed to infer statistical knowledge about the population. There are two types of sampling techniques- Probability Sampling and Non-Probability Sampling.

Probability Sampling

This is a sampling technique in which samples from a large population are chosen using the theory of probability. There are three types of probability sampling - Random, Systematic and Stratified.

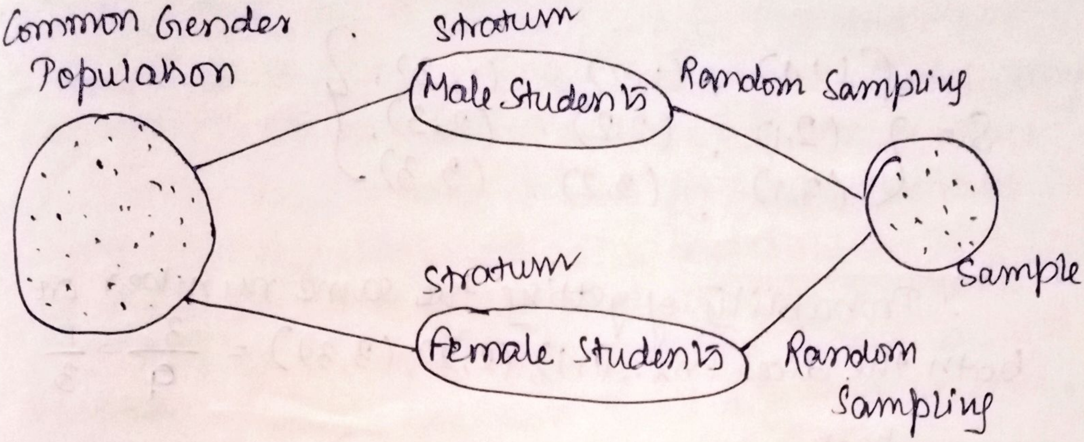
i) Random Sampling - In this method, each member of population has an equal chance of being selected in the sample.

ii) Systematic Sampling - In Systematic sampling, every nth record is chosen from the population to be a part of the sample(first create subgroups, then choose nth ~~rec~~ record)

iii) Stratified Sampling - In Stratified sampling, a ~~stradium~~ stratum is used to form ~~a large po~~ the samples from a large population.
A stratum is ~~a~~ the subset of population that shares atleast one common characteristic. After this, the random sampli

method is used to select a sufficient number of subjects from each stratum.

Example –



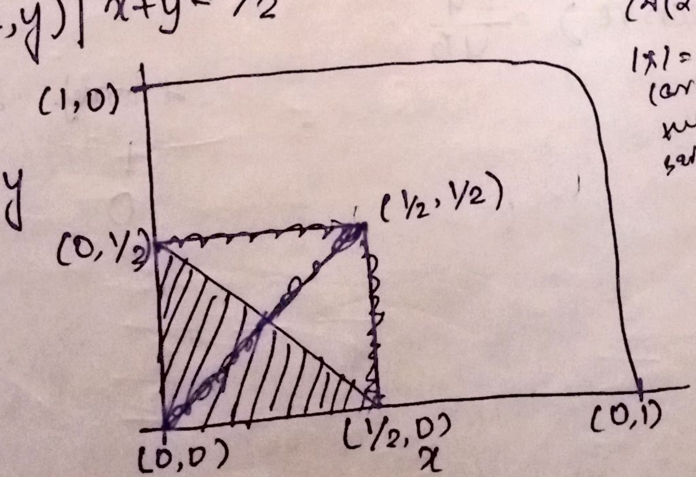
Laws of Probability

→ **Discrete Uniform Probability Law**: It states that if 'n' possible outcomes are equally similar, then probability of any event 'A' is – $P(A) = \dfrac{\text{no. of event A occurences likely}}{n}$

→ **Continuous Uniform Probability Law**: It states that it is going to measure the probability for a continuous outcome for a sub-interval $[a, b]$, within the probability range $[0, 1]$.

Probability of ~~showing~~ throwing a dart in the given graph –

$P(x, y) \mid x + y \leq 1/2$



$x \in [0, 1]$ → Prob. range

$|x| = 1$ (area of the square)

$\frac{1}{2} \times b \times h$

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

$= \frac{1}{8}$

$\therefore P(x, y) \mid x + y \leq 1/2$

$= \dfrac{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}}{1}$

$= \frac{1}{8}$.

What is the probability of getting the same number occurring in both dices when two three-sided dices are rolled at the same time.

$$S = \begin{Bmatrix} (1,1), & (1,2), & (1,3), \\ (2,1), & (2,2), & (2,3), \\ (3,1), & (3,2), & (3,3) \end{Bmatrix}$$

$\therefore$ Probability of getting the same number on both the dices (viz., (1,1), (2,2), (3,3)) $= \frac{3}{9} = \frac{1}{3}$

In a ~~bag~~ pack of 10 watches, 3 watches are known to be defective. If two watches are selected at random, what is the probability that atleast one watch is defective?

$({}^{7}C_1 \times {}^{3}C_1) + ({}^{3}C_2)$ $\qquad \frac{24}{45}$

$= 21 + 3 =$

Combinatorics

Combinatorics is all about number of ways of choosing some objects out of a collection of data.

For example, there are 5 members — A, B, C, D and E. One of them is to be chosen as the coordinator. Clearly, any one out of them can be chosen, so there are ~~ate~~ 5 ways.

Suppose 2 members are to be chosen for the positions of coordinator and secretary — A as the coordinator, and any one of the rest as secretary and likewise. So, there are $\underbrace{5}_{\text{coordinator}} \times \overbrace{4}^{\text{secretary}} = 20$ ways. Note that, choosing A and B, and then choosing B and A are considered different, i.e., the way of arrangement matters.

Suppose two coordinators are to be chosen. Here, A and B, and, B and A both are same. So, there will be 10 different ways of choosing.

Permutation of choosing 'r' distinct objects out of a collection of 'n' objects is calculated as $^nP_r = \frac{n!}{(n-r)!}$.

Combination of choosing 'r' distinct objects out of a collection of 'n' objects is ~~calle~~ calculated as $^nC_r = \frac{n!}{r!(n-r)!}$

Combinatorics Rules

~~A~~ For two sets A and B which consists of finite collection of elements, the following rules hold true:-

i) **Rule of Product**: The product rule states that if there are 'x' number of ways to choose one element from A and if there are 'y' number of ways to choose one element from B, then there will be $x * y$ ways to choose two elements — ~~from~~ one from A and one from B.

ii) **Rule of Sum**: The sum rule states that if there are 'x' number of ~~item~~ ways to choose one element from A and 'y' number of ways to choose one element from B, then there will be $x+y$ ways to choose one element either from A or from B.

Permutation with repetition

If we have 'n' objects out of which 'n_1' objects of type-1, 'n_2' objects of type-2, ..., n_k objects of type-k then no. of arrangements of these n objects can be calculated as –

$$\frac{n!}{n_1!\, n_2!\, n_3! \cdots n_k!}$$

Combination with repetition

If we have 'n' elements out of which we want to choose 'k' elements and it is allowed to choose one element more than once, then the number of ways can be calculated as –

$$^{n+k-1}C_k = \frac{(n+k-1)!}{k!\,(n-1)!}$$

/ How many distinct ways can ~~the~~ you order the letters of the word 'WATERFALL'?

$$\frac{9!}{2!\,2!}$$

$$= \frac{9!}{2\times2}$$

$$\cancel{\frac{9\times8\times7\times6\times5}{4}}$$

$$= 90720.$$

/ How many even numbers greater than 300 can be formed with digits 1, 2, 3, 4, 5 where no repetition is allowed?

1   3   2

5   2, 4

5   4   3

2, 4

2 6

$5\times4\times3\times2$

$4\times3\times2\times5$

~~3 2 2~~
~~4,5 2,4~~

~~$^4P_4$ 3 2 2,4~~

~~12~~

~~4! = 24~~

~~$^4P_3$~~

~~$^5P_4$~~

~~2,4~~

~~5! = 120~~

~~120~~

~~3 2 2 2,4~~

4 3 2 2
2,4

48

4 3 2 2 2
2,4

48

~~5 × 4 × 2~~

~~if 1 start~~ starting with 1

1 3 2
1 2,4

3 × 2
= 6

starting with 2

2 3 1
2 4

= 3

only even 3-digit

4 3 2
2,4

4 × 3 × 2 = 24

greater than 3 and even

= 24 − (6 + 3)

= 15

∴ Total = 15 + 48 + 48 = 111

Conditional Probability

If E and F are two events associated with the same sample space of a random experiment, then the conditional probability of the event E under the condition that event F has occurred is given by —

$$P(E/F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) \neq 0$$

Properties of Conditional probability

Let E and F be events associated with the sample space S of an experiment, then,

i) $P(S|F) = P(F|F) = 1$

ii) $P[(A \cup B)|F] = P(A/F) + P(B|F)$ ~~$P(A \cap B|F)$~~ $- P(A \cap B|F)$

iii) $P(E'|F) = 1 - P(E|F)$

~~iv) Multi~~

Multiplication Theorem on Probability

Let E and F be two events associated with the sample space of an experiment, then –

$$P(E \cap F) = P(E).P(F|E), \quad P(E) \neq 0$$

$$= P(F).P(E|F), \quad P(F) \neq 0$$

Random Variable

A random variable is a real-valued function whose domain is the sample space of a random experiment. The probability distribution of a random variable 'X' is the system of numbers

$$X \quad x_1 \quad x_2 \dots\dots x_n$$

$$P(x) \quad p_1 \quad p_2 \dots\dots p_n$$

where $p_i > 0$, $i = 1, 2, 3, \dots, n$

and, $\sum_{i=1}^{n} p_i = 1$

Mean and variance of a random variable

Let X be a random variable assuming values $x_1, x_2, \dots x_n$ with probabilities $p_1, p_2, \dots, p_n$ respectively such that $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$.

Mean of X, denoted by μ or expected value of X, denoted by $E(X)$ is defined as $\mu = E(X) = \sum_{i=1}^{n} x_i p_i$ and variance denoted by $\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i$

$$= E(X-\mu)^2$$

Standard deviation of a random variable 'X' is defined as $\sigma = \sqrt{\text{variance}} = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 p_i}$

Bernoulli's Trials

Trials of a random experiment are called Bernoulli Trials if they satisfy the following conditions -

i) There should be finite number of trials

ii) The trials should be independent

iii) Each trial has exactly two outcomes — success or failure

iv) The probability of success or failure remains the same in each trial.

Binomial Distribution

A random variable 'X' taking values 0, 1, 2, ..., n is said to have a binomial distribution with parameters n and p, if it's probability distribution is given by

$$P(X=r) = {}^nC_r\, p^r q^{n-r}$$

where $q = 1-p$ and $r = 0, 1, \ldots\ldots, n$

A and B are two candidates ~~sittin~~ seeking admission in a college. The probability that A is selected is 0.7 and the probability that exactly one of them is selected is 0.6. Find the probability that B is selected.

$P(A) = 0.7$

$P(A \cup B) = 0.6$

$\cancel{P(A\cup B) = P(A) + P(B) + P(A\cap B)}$

$\cancel{= 0.7 + 0.6 + 0.7 \times 0.6}$

$\cancel{P(A\cup B) - 0.7 = 0.7 \times 0.6 = P(B)}$

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

$P(A\cup B) = \cancel{\dfrac{P(A).P(B)}{P(A)}}$

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

$0.6 - 0.7 = P(B)(1 - P(A))$

$0.6 - 0.7 = P(B)\ 0.3$

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

$\cancel{P(A' \cup B) \cap P(A \cup B')}$

$P(A' \cap B) + P(A \cap B')$

$= 0.6$

$= P(A')\,P(B) + P(A)\,P(B')$

$= 0.6$

$\cancel{P(A') + P(B) + P(A')\,P(B)}$

$1 - 0.6 = 0.4 = 0.25$

$0.2b$
$-0.3b$

$(0.3)(b) + 0.7(1-b) = 0.6$

$= 0.3 + 0.7 - 0.7b = 0.6$

Let p be the probability that B gets selected.

P(Exactly one of A, B) is selected $= 0.6$

∴ P(A is selected, B is not selected) / B is selected, A is not selected)

$P(A \cap B') + P(A' \cap B) = 0.6$

$P(A).P(B') + P(A').P(B) = 0.6$

$0.7(1-p) + 0.3p = 0.6$

∴ $0.7 - 0.7p + 0.3p = 0.6$

∴ $0.7 - 0.4p = 0.6$

∴ $0.1 = 0.4p$

∴ $p = 0.25$

A bag contains 20 tickets marked with numbers 1, 2, ..., 20. 1 ticket is drawn at random. Find the probability that it will be — i) it will be a multiple of 2 or 5, ii) multiple of 3 or 5.

3, 6, 9, 12, 15, 18, 20

~~2, 4, 8,~~ 9/20 5, 10

1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
11, 12, 13, 14, 15, 16, 17, 18, 19, 20

3, 6, 9, 10,
12, 15, 18, 20

~~9/20~~

~~9 - $\frac{7}{20}$~~

$\frac{7}{20}$

$\frac{1}{4}$

2, 4, 6, 8, 10, 12, 14, 16, 18, 20

5, 10, 15, 20.

-10, 20

$\frac{12}{20} = \frac{3}{5}$

5 points are given in line one and 10 points are given on line two. Line one and line two are parallel lines. How many triangles can be formed by taking these points as vertices.

$$^5C_2 \times {}^{10}C_1 + {}^5C_1 \times {}^{10}C_2$$

$$= 325$$

Consider the experiment of tossing a coin 3 times in succession. Construct the sample space S. Write down the elements of two events E_1 & E_2 where E_1 is the event that no. of heads exceed no. of tails and E_2 is the event of getting head in the first trial find the prob. of E_1 & E_2 assuming that elements of S are equally likely.

$$S = \left\{ \begin{matrix} HHH, & HHT, & HTH, & HTT, \\ THH, & THT, & TTH, & TTT \end{matrix} \right\}$$

~~no. of tails~~

$P(E_1) = \frac{4}{8} = 1/2$

HHH,
HHT,
HTH,
THH

$P(E_2) = \frac{4}{8} = 1/2$

HHH, HHT,
HTH,
HTT

A committee of 4 persons is to be appointed for 3 officers of the p.d, 4 officers for pur. d., 2 officers of the sales dep. & one CA. find the prob. of forming the committee in the following manner:

i) there must be 1 from each category
ii) should have atleast 1 from pur. d.,
iii) CA must be in the comm'.

i) $^{3}C_1 \times {}^{4}C_1 \times {}^{2}C_1 \times {}^{1}C_1 \,/\, {}^{10}C_4$

$= 3\times4\times2 = 24 / 210 = \frac{12}{105}$

~~ii) $^{4}C_1 \times {}^{3}C_1 \times {}^{2}C_1 \times {}^{1}C_1 + {}^{4}C_2 \times {}^{3}C$~~

ii) $^{4}C_1 \times {}^{6}C_3 + {}^{4}C_2 \times {}^{6}C_2 + {}^{4}C_3 \times {}^{6}C_1 + {}^{4}C_4 \,/\, 210$

$= 80 + 90 + 24 + 1 = 195/210$

iii) $^{9}C_3 \times 1 = 84 / \cancel{210}\,5 = 2/5$

$\frac{20 \times 19^{9}}{20^{1}}$

~~Part~~

set-2

Q3)

1 2 3 4 5

1 2 3 4 5 6 7 8 9

$^{20}C_1 \times {}^{19}C_1 \times {}^{18}C_1 \times {}^{17}C_1 \times {}^{16}C_1 \times {}^{15}C_1 \times {}^{14}C_1 \times {}^{13}C_1 \times {}^{12}C_1 \times {}^{11}C_1$

$\cancel{20^{10}}$

$\left(\frac{19}{20}\right)^{9}$

1 → 1
2 → 2
3 → 3
4 → 4
5 → 5
6 → 6
7 → 7
8 → 8
9 → 9
10 → 10
11

i) originator no repeat - 9 times rumours spread

Total $= (20)(19)^{9}$ → 9 times rumours, ~~as originator~~ originator not ~~repeat~~ repeat-

ii) ~~repetition~~ total $= (20)^{10}$, as originator can be repeated

rem rem

20×19×18×...×11

i) Let us define the event E_1 — the rumour will be told 10 times without returning to the originator.

The originator can tell the rumour to any one of the remaining 20 persons in 20 ways.

And each of the $(10-1) = 9$ recepients of the rumour can tell it to any one of the remaining $(20-1) = 19$ persons without returning it to the originator in 19 ways.

$\therefore$ The required probability is given by $= \dfrac{20 \times 19^9}{20^{10}}$

$$= \left(\frac{19}{20}\right)^9$$

ii) Let us define the event E_2 — the rumour will be told 10 times without repeating to any person.

In this case, the first person (originator) can tell the rumour to any one of the available $(21-1) = 20$ persons; the second person can tell the rumour to any one of the available $(20-1) = 19$ persons; the third ~~persor~~ person can tell the rumour to any one of the remaining $(20-2) = 18$ persons; ......; the tenth person can tell the number to any one of the remaining $(20-9) = 11$ person.

Hence, the favourable number of cases for $E_2 =$ $20 \times 19 \times 18 \times \dots \times 11$.

$\therefore$ The required probability is given by $= \dfrac{20 \times 19 \times 18 \times \dots \times 11}{20^{10}}$

Monte Carlo analysis, or simulation is a computerized mathematical technique to generate random sample data based on some known distribution for numerical experiments. This method is applied to risk quantitative analysis and ~~pred~~ decision making problems. This method is used by professionals of various profiles such as finance, energy manufacturing, project management, engineering, research and development, insurance, oil and gas, transportation, etc.

This method was first used by scientists working on the atom bomb in 1940. This method can be used in those situations where we need to make an estimate and uncertain decisions such as weather forecasting.

Important characteristics of Monte Carlo simulation

Following are the important characteristics:

i) Its output must generate samples

ii) Its input distribution must be known

iii) Its result must be known while performing an experiment.

Advantages of Monte Carlo Simulation

i) Easy to implement

ii) Provides statistical sampling for numerical experiments using the computer

iii) Provides approximate solution to mathematical problems

iv) Can be used stochastic and deterministic problems

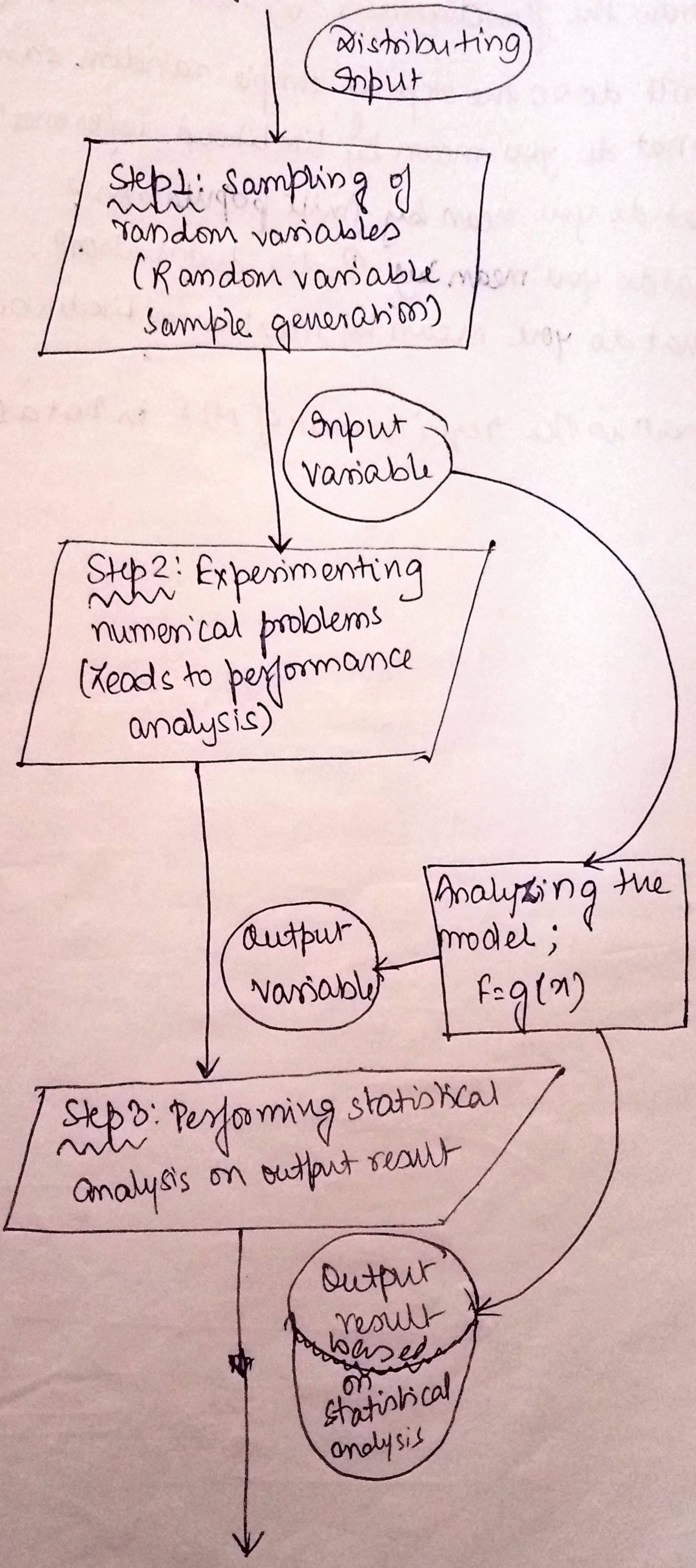
Disadvantages of Monte Carlo Simulation

i) Time Consuming

ii) There is a need to generate large number of sampling to get the desired output

iii) The results of these methods is only the approximation of the true values, but not the exact values.

Flowchart of Monte Carlo Simulation



Two cards are drawn—

a) successively with replacement

b) simultaneously (successively) without replacement

from a well-shuffled deck of 52 cards, find the probability distribution of the number of aces.

A A

$\cancel{\frac{4}{52} \times \frac{3}{51} = \frac{1}{221}}$

$\cancel{4C_1 \times 4C_1}$

$\cancel{4C_1 \times 3C_1}$

$\cancel{52 - 4}$

$\cancel{P(X=0) = 48C_2}$

$\cancel{P(X=1) = 4C_1}$

$\cancel{P(X=2) = 4C_1 \mid 3C_1}$

~~P(Ace) =~~

without

$\cancel{P(A) = \frac{4}{52}}$

$\cancel{P(AA) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{221}}$

$\cancel{P(A) = \frac{4}{52}}$

$\cancel{P(AA) = \frac{4}{52} \times \frac{4}{52}}$

Let 'X' denote the number of aces obtained in a draw of two cards. Obviously 'X' is a random variable which can take values 0, 1, 2.

(a) Probability of drawing Ace $= \frac{4}{52} = \frac{1}{13}$

Probability of not drawing an Ace $= 1 - \frac{1}{13} = \frac{12}{13}$

Since cards are drawn with replacement, all the draws are independent.

$P(X=2) = P(\text{Ace and Ace})$

$= \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$

$P(X=0) = P(\text{Non-Ace and Non-Ace})$

$= \frac{12}{13} \times \frac{12}{13} = \frac{144}{169}$

$P(X=1) = P(\text{Ace and Non-Ace}) + P(\text{Non-Ace and Ace})$

$= \frac{1}{13} \times \frac{12}{13} + \frac{12}{13} \times \frac{1}{13}$

$= \frac{24}{169}$

Probability Distribution of X —

X	0	1	2
$P(X)$	$\frac{144}{169}$	$\frac{24}{169}$	$\frac{1}{169}$

(b) If cards are drawn without replacement, then number of cases of drawing two cards out of 52 cards is $^{52}C_2$.

$P(X=2) = P(\text{Ace and Ace})$

$\cancel{= \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}}$

$= \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652}$

$= \frac{1}{221}$

$P(X=1) = P(\text{Ace and Non-Ace}) + P(\text{Non-Ace and Ace})$

$\cancel{\frac{1}{13} \times \frac{12}{13}}$

$= \frac{4}{52} \times \frac{48}{52} + \frac{48}{52} \times \frac{4}{52}$

$P(X=0) = \frac{12}{13} \times \frac{12}{13}$

$= \frac{144}{169}$

$P(X=0) = P(\text{Non-Ace and Non-Ace})$

$= \frac{^{48}C_2}{^{52}C_2} = \frac{188}{221}$

$P(X=1)$ ~~$P(\text{Ace and Non-Ace}) + P(\text{Non-Ace and Ace})$~~

$= P(\text{one Ace}) = \frac{^4C_1 \times {}^{48}C_1}{^5C_2}$ ~~$^{48}C_1 \times {}^4C_1$~~ $= \frac{32}{221}$

$P(X=2) = P(\text{both aces}) = \frac{^4C_2}{^{52}C_2} = \frac{6}{221}$

Probability Distribution of X

X	0	1	2
P(X)	$\frac{188}{221}$	$\frac{32}{221}$	$\frac{6}{221}$

/ Obtain the probability distribution of X, the number of heads in 3 tosses of a coin.

Let 'X' ~~be the probability~~ denote the number of heads in 3 tosses of a coin. Obviously, X is a random variable which can take ~~P(X=0) = P(no head~~ values - 0, 1, 2, 3. So, the sample space consists of 8 points.

S = {HHH HHT HTH HTT THH THT TTH TTT}

~~Probability of drawing~~

$P(X=0) = P(\text{no heads}) = \cancel{\frac{1}{2} \times \frac{2}{2} \times \frac{1}{21}} = \frac{1}{8}$

$P(X=1) = P(\text{one head}) = \frac{3}{8}$

$P(X=2) = \frac{3}{8}$

$P(X=3) = P(\text{three heads}) = \frac{1}{8}$

Probability Distribution of X -

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

When two ~~de~~ dice are rolled at random, obtain the probability distribution of the sum of the numbers on them.

Let 'X' denote the sum of the numbers on the two dice. Then X is a random variable which can take values 2, 3, ..., 12 ~~1,1 1,2, 1,3, 1,4,~~

$$S = \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

$P(X=2) = \frac{1}{36}$

$P(X=3) = P(1,2) + P(2,1)$

$= \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$

$P(X=4) = P(3,1) + P(1,3) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$

$P(X=5) = P(4,1), P(1,4) = \frac{4}{36} = \frac{1}{9}$

$P(X=6) = \frac{5}{36}$

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

$P(X=8) = \frac{5}{36}$

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

$P(X=10) = \frac{3}{36}$

$P(X=11) = \frac{2}{36}$

$P(X=12) = \frac{1}{36}$

X	2	3	4	5	6	7	8	9	10	11	12
p(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{26}$	$\frac{5}{36}$	$\frac{4}{26}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

~~Berr~~

Binomial Distribution or Bernoulli's Distribution

If two events are mutually exclusive, then the conditions for which we can use the distribution are as follows:

(a) If the random experiment is performed repeatedly and a finite and fixed number of times, then we can use Binomial or Bernoulli's distribution.

(b) each trial results in two mutual exclusive and exhaustive outcomes success ~~and~~ and failure.

(c) Trials have to be independent.

(d) If ~~are~~ 'p' is the probability of success then $q = 1-p$ is the probability of failure in any trial.

The formulae:

① $p(r) = P(X=r) = {}^nC_r\, p^r q^{n-r}$

Probability Mass Function (pmf) $\rightarrow p(r) = P(X=r)$

② $\sum_{r=0}^{n} p(r) = (p+q)^n = 1$

③ Mean of binomial distribution $= np \rightarrow$ Expected value

④ Variance of binomial distribution $= npq$

For Binomial distribution, variance is less than Mean

10 unbiased coins are tossed simultaneously. Find the probability of obtaining —

(a) exactly 6 heads

(b) no heads

~~p(r) = pt~~

(a) ${}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4$

$\cancel{210 \times \frac{1}{64} \times \frac{1}{16}} = \frac{210}{1024}$

(b) $\left(\frac{1}{2}\right)^{10}$

${}^{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10}$

A merchant's file of 20 accounts contain 6 expenses accounts and 14 non-expenses accounts. An auditor selects randomly 5 of these accounts for examination.

i) What is the probability that the auditor find exactly 2 expense account.

ii) Find the expected number of expense accounts in the sample selected.

$$\cancel{{}^{20}C_{2}\left(\frac{6}{20}\right)^{5}\left(\frac{14}{20}\right)^{15}}$$

(i) $^5C_2\left(\frac{6}{20}\right)^{2}\left(\frac{14}{20}\right)^{3}$

~~Expt~~

$$\cancel{{}^{20}C_{6}\left(\frac{6}{20}\right)^{5}\left(\frac{14}{20}\right)^{15}}$$

(ii) $5\times\frac{6}{20}$

4 bad apples are mixed accidentally with 20 good apples. Obtain the probability distribution of the number of bad apples in a draw of 2 apples at random.

Let X be the random variable which denotes the number of bad apples drawn. ~~There~~ $\therefore$ X can take values 0, 1, 2.

$\therefore$ Total cases $= {}^{24}C_2 = 276$

$\therefore P[X=0] = f(0) = {}^{20}C_2 \times {}^4C_0 = \cancel{190} = \frac{190}{276}$

$\therefore P[X=1] = f(1) = {}^{20}C_1 \times {}^4C_1 = \frac{80}{276}$

$\therefore P[X=2] = f(2) = {}^{20}C_0 \times {}^4C_2 = \frac{6}{276}$

$\therefore$ Probability Distribution of X —

X	0	1	2
f(x)	$\frac{\cancel{190}\ 95}{\cancel{276}\ 138}$	$\frac{\cancel{80}\ 40}{\cancel{276}\ 138}$	$\frac{\cancel{6}\ 3}{\cancel{276}\ 138}$

X	0	1	2
f(x)	$\frac{95}{138}$	$\frac{40}{138}$	$\frac{3}{138}$

Theory of Mathematical Expectation

If X is a random variable which can assume any one of the values $x_1, x_2, x_3, \ldots\ldots, x_n$ with respective probabilities $p_1, p_2, p_3, \ldots\ldots, p_n$, then the mathematical expectation of X, usually called the expected value of X and denoted by $E(X)$ is defined as —

$$E(X) = p_1 x_1 + p_2 x_2 + p_3 x_3 + \ldots\ldots + p_n x_n$$

$$\cancel{\sum_{i=1}^{n} x_i p_i \quad x_i = 1 \text{ to } n}$$

$$= \sum p \times x$$ where the Σ is taken on different values of x.

Imp

Physical Interpretion of $E(X)$

Let us consider the following frequency distribution of the random variable X —

X	x_1	x_2	x_3	x_4	......	x_n
f	f_1	f_2	f_3	f_4	......	f_n

The mean of the distribution,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i x_i$$

This can be written as,

$$\bar{x} = \frac{f_1}{N} x_1 + \frac{f_2}{N} x_2 + \ldots\ldots + \frac{f_n}{N} x_n$$

$$p_i = \frac{f_i}{N}$$

$$\text{or } \bar{x} = p_1 x_1 + p_2 x_2 + \ldots + p_3 x_3$$

$$\text{or } \bar{x} = E(X)$$

Hence, it is proved that mathematical expectation of a random variable X is nothing but its arithmetic mean.

Prove the following theorems -

i) $E(C) = C$ where C is a constant.

~~$E(C) = p_1 c_1 + p_1 c_2 + p_3 c_3 + \dots$~~

$$E(C) = \bar{C} = \frac{1}{N}(f_1 c + f_2 c + \dots + f_n c)$$

$$= \frac{1}{N} c (f_1 + f_2 + f_3 + \dots)$$

$$= \frac{N}{N} c$$

$$= C$$

$E(C) = $ Mean of C

$$= \frac{C + C + C + \dots + C}{N}$$

$$= \frac{\cancel{N}}{\cancel{N}} C$$

$$= C.$$

ii) $E(CX) = C E(X)$ where C is a constant

$E(CX) = $ Mean of CX

$$= \frac{c x_1 p_1 + c x_2 p_2 + c x_3 p_3 + \dots + c x_n p_n}{N}$$

$$= \frac{c(x_1 p_1 + x_2 p_2 + x_3 p_3 + \dots + x_n p_n)}{N}$$

$$\cancel{= c \bar{x}}$$

$$= C E(X).$$

iii) $E(ax+b) = aE(x) + b$ where a and b are constants.

also done by p.

$$E(ax+b) = \frac{(ax_1 + b) + (ax_2 + b) + \dots + (ax_n + b)}{N}$$

$$= \frac{a(x_1 + x_2 + \dots + x_n)}{N} + \frac{b\cancel{N}}{\cancel{N}}$$

$$= aE(x) + b$$

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

If X and Y are random variables, then.

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

i.e., Expected value of the sum of two random variable is equal to the sum of their expected values.

Generalized

$$E(X_1+X_2+\cdots+X_n) = E(X_1)+E(X_2)+\cdots+E(X_n)$$

or

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Generating function

The corresponding generating function, $f(x) = a_0 + a_1x + \cdots + a_nx^n$

a_n is the n^{th} term of $f(x)$ and the coefficient of x^n

The Generating function is not a function which returns the n^{th} term as the output, instead, it is a function whose power series displays the terms of the sequence.

For example, the power series $2+3x+5x^2+8x^3+12x^4+\cdots$ displays the sequence $2,3,5,8,12,\ldots$ as coefficients.

An infinite power series is simply an infinite sum of terms of the form C_nx^n where C_n is some constant. So we can an

a power series $\sum_{k=0}^{\infty} C_kx^k$

What sequence is represented by the generating series—
$3+8x^2+8x^3+x^5/7+100x^6+\dots$

$f(x) = 3+8x^2+x^3+x^5/7+100x^6+\dots$

~~$= \frac{1}{x^2}(3x^2+$~~

~~$= \frac{1}{x}(3x+8x^3$~~

~~$= \frac{1}{x}(3x+8x^3$~~

Here, $a_0 = 3$ since the coefficient of x^0 is 3.
$a_1 = 0$ since ~~the coefficient of x is 0.~~ we can't find any term with x^1

$a_2 = 8$
$a_3 = 1$
$a_4 = 0$
$a_5 = 1/7$
$a_6 = 100$ and so on

Therefore the sequence is $3, 0, 8, 1, 0, 1/7, 100, \dots$

What is the generating function for the series
$1, 1, 1, 1, 1, \dots$

Generating series— $1+x+x^2+x^3+x^4+\dots$

The series is in GP with common ratio x.

$S = 1+x+x^2+x^3+x^4+\dots$
$xS = x+x^2+x^3+x^4+x^5+\dots$

$(S-xS) = 1$
$S = \frac{1}{1-x}$

Generating function of the series $= \frac{1}{1-x}$

Find out the generating function for the series
1, -1, 1, -1, ...

~~Sol~~ $1-x+x^2-x^3+x^4 - \ldots\ldots$

Gen function $= \cancel{2x}\,(1+x)^{-1}$
$= \frac{1}{1+x}$

$S = 1-x+x^2-x^3+x^4 - \ldots$
$+\ xS = x-x^2+x^3-x^4+ \ldots$
$\cancel{S(1-x)}\ S(1+x) = 1$
$\therefore S = \frac{1}{1+x}$

1, 3, 9, 27, ...

$1+3x+9x^2+27x^3+ \ldots$

~~$3(1+3x+9x^2$~~

$S = 1+3x+9x^2+27x^3+ \ldots$
$3xS = 3x+9x^2+27x^3+ \ldots$
$S(1-3x) = 1$
$\therefore S = \frac{1}{1-3x}$

~~$S = 3+9x+27x^2$~~

2,2,2,2,2, ...

$2+2x+2x^2+2x^3+2x^4+ \ldots$

$S = 2+2x+2x^2+2x^3+2x^4+ \ldots$
$-\ xS = 2x+2x^2+2x^3+ \ldots$
$S = \frac{2}{1-x}$

/ 3, 9, 27, 81

$3 + 9x + 27x^2 + 81x^3 + \cdots$

$$S = 3 + \cancel{9x} + \cancel{27x^2} + \cancel{81x^3} + \cdots$$
$$- \; 3xS = \cancel{9x} + \cancel{27x^2} + \cancel{81x^3} + \cdots$$

$$S(1-3x) = 3$$

$$S = \frac{3}{1-3x}$$

/ 2, 4, 10, 28, 82, ...

~~2(1+2+5+14~~

$$S = 2 + 4x + 10x^2 + 28x^3 + 82x^4 + \cdots$$
$$= (1+1) + (1+3)x + (1+9)x^2 + (1+27)x^3 + (1+81)x^4 + \cdots$$

$$= (1 + x + x^2 + x^3 + x^4 + \cdots) + (1 + 3x + 9x^2 + 27x^3 + 81x^4 + \cdots)$$

$$= \frac{1}{1-x} + \frac{1}{1-3x}$$