Page No:
Date: 22 07

# Probability & Statistics

Text Book - Introduction to statistics and Data Analysis.
- By Michael Schanaker Shalath.

RB - Probability and Statistics for Engineering Scientists.
- By Walpole.

∴ Data Set - collected Data. (Population) $\Omega$

∴ Unit - Entity of Data (also called as observation) $\omega$

∴ Sample Space - Selection of observations.

$\{\omega_1, \omega_2, \ldots, \omega_n\} \subseteq \Omega \qquad \omega \in \Omega$

∴ Variable - A particular feature of any observation, denoted by $X$

$X : \Omega \rightarrow S$

Variable Types:

A) Qualitative and Quantitative Variable

| Qualitative | Quantitative |
|---|---|
| Variables which can take Values that cannot be ordered in a logical or natural way<br>eg. Name, Colour, etc. | Variables which represents measurable Quantity<br>eg. Numeric values, |

Page No:
Date:

B) Discrete and continuous Variable

| | |
|---|---|
| Variables that can take finite no. of values<br>eg. Distance from clg to hostel. | Variables that can take infinite no. of values.<br>eg. Height. |

C) * Scales 24-07.

Used to classify the consideration of variables.

1) Nominal Variables

Values cannot be ordered.

Ex.

2) Ordinal Variables

Values can be ordered, but differente between the ~~co~~ values can't be interpreted in a meaningful way.

Ex.

3) Continuous Variables

Values can be ~~intepretted~~ ordered and difference can be interpreted in a meaningful way

Ex.

Types of Continuous Variable

a) Interval scale - In this scale difference between the values can be interpreted but not the ratios eg Ice cube $-10^\circ$, $-15^\circ$

b) Ratio Scale - In this scale difference and ratios both can be interpreted. eg speed of vectors.

Page No:
Date: 24 07

c) Absolute scale - In this scale values are measured in natural units

Ex. Natural Numbers.

D) * Grouped / Categorical Data —

In group data, data is available in summarized form.

Note - Any grouped variable which can take only two values are called binary variables.

Q. Can qualitative data be continuous ? → No but it can be Nominal

Q. Quantitative data is alway discrete ?

Note Yes, Qualitative data is alway discrete.

• Frequency

1) Absolute frequency 2) Relative frequency.

Absolute frequency - The number of observation in a particular category (denoted by ni)

Relative Frequency - It gives the comparision between the number of times a number has been repeated to the total frequencies of all the numbers. $\left( f_i = \frac{n_i}{n} \right)$

It lies between 0 & 1

Ex. 1

M, F, M, F, M, M, M, F, M, M

$n = 10$

Ex. 2

28, 35, 42, 90, 70, 56, 75, 66, 30, 89, 75, 64, 81, 69, 55, 83, 72, 68, 73, 16

$n = 20$

Ex 1 Frequency Distribution Table $\quad f_1 + f_2 + f_3 \dots f_n = 1$

| Category | Absolute $n_i$ | Relative $f_i$ |
|---|---|---|
| M | 7 | $f_i = \frac{7}{10}$ |
| F | 3 | $f_i = \frac{3}{10}$ |
| | 10 | 1 |

Ex 2

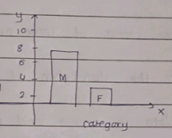
| Class intervals | $n_i$ | $f_i$ |
|---|---|---|
| 0 - 20 | 1 | 1/20 |
| 21 - 40 | 3 | 3/20 |
| 41 - 60 | 3 | 3/20 |
| 61 - 80 | 9 | 9/20 |
| 81 - 100 | 4 | 4/20 |
| | 20 | 1 |

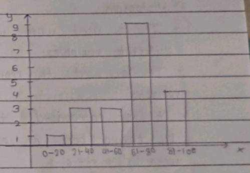
## Graphical Representation

1) Bar chart

- It consist of one bar for each category.
- Height of the bar is determined by absolute frequency or relative frequency of the respective category on Y-axis.
- It is used for Nominal and ordinal variables.
- It is used when number of categories are not too large.

Ex. 1



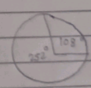




2) Pie chart

- Pie chart is represented by a circle partioned into segments where each segment represents a category.
- It is used to visualized the absolute frequency or relative frequency of nominal & ordinal values.
- The size of segment depends upon the relative frequency & is determined by the angle given by $f_i \times 360°$

eg 1)

| Category | Absolute | Relative $f_i$ | Angle |
|---|---|---|---|
| M | 7 | 7/10 | $\frac{7}{10} \times 360° = 252$ |
| F | 3 | 3/10 | $\frac{3}{10} \times 360° = 108$ |
| | 10 | 1 | |



2 eg.

| class inter | $n_i$ | $f_i$ | Angle |
|---|---|---|---|
| 0-20 | 1 | 1/20 | $1/20 \times 360^\circ = 18^\circ$ |
| 21-40 | 3 | 3/20 | $3/20 \times 360^\circ = 54^\circ$ |
| 41-60 | 3 | 3/20 | $3/20 \times 360^\circ = 54^\circ$ |
| 61-80 | 9 | 9/20 | $9/20 \times 360^\circ = 162^\circ$ |
| 81-100 | 4 | 4/20 | $4/20 \times 360^\circ = 72^\circ$ |
| | 20 | 1 | |

3) Histogram

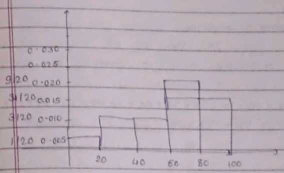
- It is used when variables consist of large number of different variables
- It is used to represent the distribution of continuous variables.
- It is categorised the data in groups and plot the graph by taking bars of each category with height $h_j = \frac{f_j}{d_j}$
where $d_j = e_j - e_{j-1}$

(width of the class)

eg.

| class interval | Af | RF $f_i$ | height $h_j$ |
|---|---|---|---|
| 0-20 | 1 | 1/20 | $h_j = \frac{f_j}{d_j} = \frac{1}{20} \times \frac{1}{20} = \frac{1}{400} = 0.0025$ |
| 20-40 | 3 | 3/20 | $3/20 \times 1/20 = 0.0075$ |
| 40-60 | 3 | 3/20 | $3/20 \times 1/20 = 0.0075$ |
| 60-80 | 9 | 9/20 | $9/20 \times 1/20 = 0.0225$ |
| 80-100 | 4 | 4/20 | $4/20 \times 1/20 = 0.01$ |

20 $d_j$



Page No:
Date:

## Arithmatic Mean

i) Measures of central tendancy

Let $x_1, x_2, \ldots, x_n$ be the set of values of a variable then the arithmetic mean is given by $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

eg. 2) $\frac{28 + 35 + 42 + 90 + 70 + 56 + 75 + 66 + 30 + 19 + 75 + 64 + 81 + 69 + 55 + 83 + 72 + 68 + 73 + 16}{20}$

$= \frac{1237}{20}$

$= 61.85$

## Weighted Mean

It is used to find the mean of grouped data.

$\frac{0+20}{2}$

ex. 2)

| class interval | f $n_i$ | rf $f_i$ | mid value $m_i$ | |
|---|---|---|---|---|
| 0-20 | 1 | 1/20 | 10 | 1/2 |
| 20-40 | 3 | 3/20 | 30 | 9/2 |
| 40-60 | 3 | 3/20 | 50 | 15/2 |
| 60-80 | 9 | 9/20 | 70 | 63/2 |
| 80-100 | 4 | 4/20 | 90 | 36/2 |
| | 20 | 1 | | $\frac{124}{2} = 62$ |

$\bar{x} = \frac{n_1 \times m_1 + n_2 \times m_2 + \cdots n_n \times m_n}{n}$

$= \sum^{k} \frac{n_i m_i}{n} \quad \sum^{k} \frac{n_i}{n} m_i = \sum^{k} f_i m_i$

Weighted mean $= 62$.

* Note- The results of mean & weighted mean differ as we use middle class as an aproximation of the mean within the scale.

Page No:
Date:

- The sum of deviations of each variable around arithmetic mean is zero

$x_1, x_2, x_3, \ldots x_n \quad \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$

Deviation $\{\bar{x} - x_1, \bar{x} - x_2, \bar{x} - x_3, \bar{x} - x_n\}$

$(\bar{x} - x_1) + (\bar{x} - x_2) + (\bar{x} - x_3) + \cdots + (\bar{x} - x_n)$

$= \bar{x} - x_1 + \bar{x} - x_2 + \bar{x} - x_3 + \cdots + \bar{x} - x_n$

$= n\bar{x} - x_1 - x_2 - x_3 \cdots x_n = \cancel{n}\left(\frac{x_1 + x_2 + x_3 + x_n}{\cancel{n}}\right) = \cancel{x_1} + \cancel{x_2} + \cancel{x_3} + \cancel{x_n} - \cancel{x_1} - \cancel{x_2} - \cancel{x_3} - \cancel{x_n} = 0.$

If $y_i = a + bx_i$; $a, b$ are constants then $\bar{y} = a + b\bar{x}$

$y_1 = 2 + 3(5)$

$y_2 = 2 + 3(10) \quad \bar{y} = 2 + 3\left[\frac{5 + 10 + 15 + \cdots + 50}{10}\right]$

$y_3 = 2 + 3(15)$

$\vdots$

$y_{10} = 2 + 3(50)$

$\rightarrow x_1, x_2, \ldots, x_n \quad \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

$y_1 = a + bx_1$

$y_2 = a + bx_2$

$\vdots$

$y_n = a + bx_n$

$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}$

$= \frac{(a + bx_1) + (a + bx_2) + (a + bx_n)}{n}$

$= \frac{a + bx_1 + a + bx_2 + a + bx_n}{n}$

Page No.
Date

$$= \frac{na + bx_1 + bx_2 + \cdots + bx_n}{n}$$

$$= \frac{na + b(x_1 + x_2 + \cdots + x_n)}{n}$$

$$= \frac{\cancel{n}a}{\cancel{n}} + b(x_1 + x_2 + \cdots + x_n)$$

$$\bar{y} = a + b\bar{x}$$

- Median

Median defers to the middle value after arranging data in ascending order. It is denoted by $\tilde{x}_{0.5}$

n odd = $\tilde{x}_{0.5}$ = middle value

n even = $\tilde{x}_{0.5} = (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})\frac{1}{2}$

eg 2) 16, 28, 30, 35, 42, 55, 56, 64, 66, 68, 69, 70, 72, 73, 75, 81, 83, 88, 90

= Median 68.5

31/07

- Quantile

1) Quanties partitioned the data into different proportion.

Let

$\alpha$ be a number between 0 & 1 then the quantile is given by $\tilde{x}_\alpha = (\alpha \times 100\%)$ which means that quantile divides the data into $\alpha \times 100\%$ and $(1-\alpha) \times 100\%$.

2) If $\alpha = 0.1, 0.2, \ldots, 0.9$ then the quantile is called "Deciles."

3) If $\alpha = 0.2, 0.4, 0.6, \ldots$ then the quantile is called "Quintile".

Page No.
Date

4) If $\alpha = 0.25, 0.50, 0.75$ then is called "Quartile".

5) If $\alpha \times 100$ is integer then quantile are called "percentile".

6) Quantile can be defined as $\tilde{x}_\alpha = \{x(k)$, If $n\alpha$ is not integer then choose k as smallest integer strictly greater than $n\alpha$

$$\tilde{x}_\alpha = \begin{cases} x(k) \\ \frac{1}{2}(x(n\alpha) + x(n\alpha+1)), & \text{if } \alpha \text{ is an integer} \end{cases}$$

- ~~Mode~~ Mode

The mode of $n$ observations $x_1, x_2, \ldots, x_n$ is the value which occur the most when compared with other values. Denoted by $\bar{x}_M$

- Absolute Deviation

Consider deviations of $n$ observations around a venue A then the arithmatic mean of all the deviations

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)$$

Ex $A = 20$ $x_1 = 19$ $x_2 = 19.5$ find deviations

$$D = \frac{1}{2}\left((19-20) + (19.5-20)\right)$$

$$= \frac{1}{2}(-1-0.5) = -ve$$

The problem with deviations is it can be positive or negative and hence sum can be very very small or zero & therefore we use modulus. And Hence

$$D = \frac{1}{n}\sum_{i=1}^{n}|x_i - A|.$$

Page No.
Date.

- **Absolute Mean Deviation**

Formula - $D(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$

- **Absolute median Deviation**

Formula - $D(\tilde{x}) = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}_{0.5}|$

Note - Absolute deviation is minimum when A = median.

1. **Mean Squared Error**

$S^2(A) = \frac{1}{n} \sum_{i=1}^{n} (x_i - A)^2$

2. **Variance** - $\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

3. **Standard Deviation** - $\tilde{s} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Q.

| week | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| A | 0 | 0 | 0 | 0 | 0 |
| B | -10 | +10 | -10 | +10 | -10 |
| C | 3 | 5 | 6 | 2 | 4 |

Consider 3 students A, B & C arrivies at different time in the class to attain their lectures for 5 weeks as shown in the given data. Calculate the variance, standard deviation & absolute deviation of three students.

$\bar{x}$ - mean

Page No.
Date.

→ Variance of A, B & C

1) $\tilde{s}^2 = \frac{1}{5} \sum_{i=1}^{n} (0-0) = \frac{1}{5} \times 0 = 0$

2) $\tilde{s}^2 \quad \frac{1}{5} (-10+2)^2 + (10+2)^2 + (-10+2)^2 + (10+2)^2 + (-10+2)^2$

$\frac{1}{5}(-8^2 + 12^2 - 8^2 + 12^2 - 8^2) = 64 + 144 + 64 + 144 + 64$

$= \frac{1}{5}(480) = 96$

3) $\tilde{s}^2 \quad \frac{1}{5} (3-4)^2 + (5-4)^2 + (6-4)^2 + (2-4)^2 + (4-4)^2$

$= \frac{1}{5}(1+1+4+4+0) = \frac{1}{5} 10 = \underline{\underline{2}}$

Standard Deviation of A, B & C

1) $\tilde{s} = \sqrt{\frac{1}{5} (0-0)^2 + (0-0)^2 \cdots + (0-0)^2} = \underline{0}$

3) $\tilde{s} = \sqrt{\frac{1}{5} (3-4)^2 + (5-4)^2 + (6-4)^2 + (2-4)^2 + (4-4)^2}$

$= \sqrt{\frac{1}{5}(\cancel{480}\ 10)} = \cancel{4\sqrt{6}}\ \underline{\underline{\sqrt{2}}}$

2) $\tilde{s} = \sqrt{\frac{1}{5}(480)} = \underline{\underline{4\sqrt{6}}}$

Absolute Deviation

1) $D = \frac{1}{5} \sum_{i=1}^{n} |0 - A| = 0$

2) $8 + 12 + 8 + 12 + 8 = \frac{1}{5} \times 48 = 9.6$

3) $\frac{1}{5} \times 6 = \frac{6}{5} = \underline{1.2}$

Page No:
Date.

- **Permutation & Combination**

If a operation is performed in any one way and if for each way a second operation can be performed in n2 ways then the two operations can be performed together in $n_1 \times n_2$

Q If 22 member club meet to elect a chairmen and a leader then how many different ways are possible.

→ $\underset{ch}{22} \times \underset{le}{21}$

Q. • If an operation can be performed in $n_1$ ways and if for each of this a second operation can be performed in $n_2$ ways and if for each of these a third operation can be performed in $n_3$ ways & so on, then all the operations can be performed in $n_1 \times n_2 \times n_3 \cdots \times n_k$.

Q How many even 4 digit numbers can be formed from the digit 0, 1, 2, 5, 6, 7. If each of the digit can be used only once.

→ 0, 1, 2, 5, 6, 7.

1st case → take 0 as even at 4 digit  $\underset{5}{\_} \times \underset{4}{\_} \times \underset{3}{\_} \times \underset{1}{0} = 60$

2nd case take 2 | 6 as 4 digit even number.  $\underset{4}{\_} \times \underset{4}{\_} \times \underset{3}{\_} \times \underset{2}{2|6} = 96.$

$\therefore n_1 + n_2 = 60 + 96 = 156$

$60 \times 96 = 5760$

Page No:
Date.

2 eg Twelve students compete in a race. In how many ways first three prizes be given?

→ $\underset{12}{\overset{1}{\_}} \times \underset{11}{\overset{2}{\_}} \times \underset{10}{\overset{3}{\_}}$

$12 \times 11 \times 10 = 12 \times 11 \times 10 = \underline{\underline{1320}}$ ways

05/08/24

Definition

A permutation is an arrangement of all or some part of the set of objects.

Theorem 1:

The Number of permutation of n objects is n factorial

Theorem 2:

The number of permutation of n distinct objects taken at r time is given by $\boxed{{}^nP_r = \frac{n!}{(n-r)!}}$

Q. A president and a hunter are to be choosen from a club of 50 members. How many different choises are possible if:
a) There is no restriction
b) S will serve only if he is president.
c) B and D serves together or not at all

→ a) ${}^nP_r = \frac{50!}{(50-2)!} = \frac{50 \times 49 \times 48!}{(48)!} = \underline{2450}$

Page No:
Date:

b) case 1 → S serve

$\underline{S}$, $\underline{49}$

case 2 : S do not serve

${}^{49}P_2$

$49 + 2352 = \underline{\underline{2401}}$

c) case 1 B & D serve togethe

$\underline{B}$ $\underline{D}$

$\underline{D}$ $\underline{B}$

case 2 → Both don't serve

${}^{48}P_2$

$2 + {}^{48}P_2 = \underline{\underline{2258}}$

Theorem 3:

The no. of permutations of n objects arranged in a circle is $\boxed{(n-1)!}$

Theorem 4:

The no. of distinct permutation of n objects out of which $n_1$ are of one type $n_2$ are of $2^{nd}$ type and so on $n_k$ $^{th}$ are of $k^{th}$ type. then it is given by:

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

Theorem 5:

The no. of ways of partitioning a set of n objects into r cells with $n_1$ elements in the $1^{st}$ cell $n_2$ in $2^{nd}$ cell and so on is given by:

$$\boxed{\frac{n!}{n_1!\,n_2!\cdots n_k!} \quad \text{where } n_1+n_2+\cdots+n_k = n}$$

Page No:
Date:

Q. How many different letter arrangements can be made from the letters in the word "statistics"

→ $\frac{10!}{3!\,3!\,2!} = 50,400$

COMBINATORICS

# Probability

1) <u>Random Experiment</u> -

A experiment that can be repeated any no. of times under the same set of conditions and it's outcome is known only after the complition of experiment is called a random experiment.

eg. Tosing of coin, rolling of dice

2) <u>Simple Events</u> -

A possible outcome of a random experiment is called a simple event. It is also called as an elementary event. It is denoted by ( $w_i$ ) {1} or {2}

3) <u>Sample Space</u> - ( $\Omega$ )

The set of all possible outcomes of a random experiment is called a sample space. It is denoted by ( $\overset{\Omega}{\cancel{W}i}$ ) { $w_1$, $w_2$, $w_k$ }. eg. { {1}, {2}, {3}, .... {6} }

4) <u>Events</u> -

Subsets of sample space $\overset{\Omega}{\cancel{W}}$ are events. They are denoted by capital letters. { {1}, {3}, {5} }

5) <u>Composite / complementary event</u>

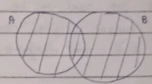
It refers to the non occuring of the event. It is denoted by $\bar{A}$.

Note ( $\Omega$ is a event which always occurs & so it is called as sure events / certain event. On the other hand $\phi$ is an event called as impossible event ).

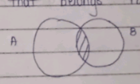
6) <u>Venn Diagram</u>

In venn diagram two or more sets are visualized by circles. Overlapping circles implies that both the events have one or more identical simple events. Seperated circles shows that none of the simple events of event A are conent in a sample space of event B.

a) $A \cup B$ - The union of events $A \cup B$ is the set of all simple events of A & B. Which occurs if atleast 1 of the simple events of A or B occurs.

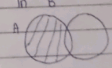



b) $A \cap B$ - The insersection of events is the set of all simple events of A or B which occurs when a simple events occurs that belongs to both A and B.



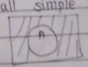

c) $A \setminus B$ (A but not B) -

The event A minus B contains all the simple events of A which are not contain in B




d) $\bar{A}$ (not A) -

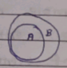
It contains all simple events of $\Omega$ which are not contain in A



Page No:
Date.

e) $A \subset B$ ( subset )

It contain all simple events of A which are also the part of sample space of event B.




7) Disjoint Events -

Two events A & B are said to be disjoint if $A \cap B = \phi$ holds i.e both events cannot occurs simultaneously.
Note - The event A & $\bar{A}$ are disjoint events.

8) Mutually Disjoint Events -

The events $A_1, A_2, \cdots, A_k$ are said to be mutually disjoint events if $A_i \cap A_j = \phi$ for all $i \neq j$.

9) Complete Decomposition -

The events $A_1, A_2, \cdots, A_k$ forms a complete decomposition of $\Omega$ if and only if $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$ and $A_i \cap A_j \neq \phi$ for all $i \neq j$

Page No:
Date.

Q. $P = \{a \mid a$ is an odd prime less than $7\}$ $\{3, 5\}$
$Q = \{b \mid b \in \mathbb{N}, 0 \leq b < 5\}$ $\{1, 2, 3, 4\}$
a) find no. of proper subsets of P and Q
b) $P \cup Q$ $\{0, 1, 2, 3, 4, 5\}$
c) $P \setminus Q$ $\{5\}$

Q. $x = \{x \mid x = 3n - 1, n \in \mathbb{N}\}$ $\{2, 5\}$
$y = \{y \mid y$ is prime no $< 7\}$ $\{2, 3, 5\}$
find $x \cap y$ $\{2, 5\}$

- Axiomatic Definition Probability

Axiom 1 :- Every random event A has the probability in close interval zero to one i.e $0 \leq P(A) \leq 1$

Axiom 2 :- The sure event has probability 1 i.e $P(\Omega) = 1$.

Axiom 3 :- If $A_1$ & $A_2$ are disjoint events then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.
Note :- Theorem of additivity of disjoint events-, if $A_1, A_2 \ldots A_k$ are disjoint events then $P(A_1 \cup A_2 \cup \ldots A_k) = P(A_1) + P(A_2) \cdots + P(A_k)$

Q. Suppose an event is defined as the number of coins observed on the ~~upper~~ upper surface of die with rolling of decire

→ $P\{\{2\}\} + P\{\{3\}\} \cdot P\{\{5\}\} - 1/6$

Page No:
Date:

**Corally 1 :-** $P(\bar{A}) = 1 - P(A)$.

proof - By (Axiom 2) = $P(\Omega) = 1$

$$\Omega = A \cup \bar{A} \quad , \quad A \cap \bar{A} = \phi$$

$$P(\Omega) = P(A \cup \bar{A})$$

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

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

Q. The probability of occurance of impossible events in zero.

→ **Corally 2 :-** $P(\phi) = P(\bar{\Omega}) = 0$.

proof :-

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

$$P(\bar{\Omega}) = \cancel{1-P(\Omega)}\ 1 - P(\Omega)$$

$$= 1 - 1 = 0.$$

Q Let $A_1$ and $A_2$ be not necessarily disjoint events then the probability of occurance of $A_1$ / $A_u$ is

→ **Corally 3 -** $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

**Corally 4 -** If $A \subset B$ then $P(A) \leq P(B)$.

Page No:
Date: 07 08 24

Ex:] Rolling of a dice to get an even no.

1) Random Exp :- ✓
2) Sample space :- $\{\{1\}, \{2\}, \{3\}, \ldots, \{6\}\}$
3) Simple event :- $\{1\}$ or $\{2\}$
4) Event :- $\{\{2\}, \{4\}, \{6\}\}$
5) Complementry events :- $\{\{1\}, \{3\}, \{5\}\}$.

- **Aximatic Defifitian of Probability.**

**Axiom 1**

Every random event A has the probability $[0, 1]$ i.e probability of $0 \leq P(A) \leq 1$.

**Axiom 2**

The sure event has a probability 1 i.e $P(\Omega) = 1$

**Axiom 3**

$A_1$ & $A_2$ are disjoint event then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

**Note :** Theorem of Additivity of disjoint events

If $A_1, A_2, \ldots, A_k$ are disjoint event then

$$P(A_1 \cup A_2 \ldots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$$

$$\text{Probability} = \frac{\text{no of favarable outcomes}}{\text{Total no. of outcomes}}$$

Page No:
Date:

Suppose an event is defined as the number of coins observed on the upper surface of the dise when rolling it what is the probability of 5 prime number.

$$P\text{ (Prime numbers)} = P\{2\} \cup \{3\} \cup \{5\}$$
$$= P\{2\} + P\{3\} + P\{5\}$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$
$$= \frac{3}{6} = \frac{1}{2}$$

Write down the probability of complementary event is A.

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

proof : By (Axioms 2); $P(\Omega) = 1$

$$\Omega = A \cup \bar{A}, \quad A \cap \bar{A} = \phi$$
$$P(\Omega) = P(A \cup \bar{A})$$
$$1 = P(A) + P(\bar{A})$$
$$1 - P(A) = P(\bar{A})$$ hence proved

The probability occurance of an impossible events is 0.

Corollary 2 :- $P(\phi) = P(\bar{\Omega}) = 0$

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

$$P(\bar{\Omega}) = 1 - P(\Omega) \quad \cdots \text{(corollary 1)}$$
$$= 1 - 1$$
$$= 0$$ hence proved

Let $A_1$ and $A_2$ be not necessarily disjoint events then the probability of occurance of $A_1$ or $A_2$ is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

If $A \subseteq B$ then $P(A) \le P(B)$

Page No:
Date:

Note :

1. $0 \le P(A) \le 1$
2. $P(\Omega) = 1$
3. $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ if $A_1$ & $A_2$ are disjoint events.
4. $P(\phi) = P(\bar{\Omega}) = \phi$
5. $P(\bar{A}) = 1 - P(A)$
6. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
7. If $A \subseteq B$ then $P(A) \le P(B)$

→ A and $\bar{A} \cap B$ are disjoint set

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

By Axiom 3

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

Imp

## Condilianal Probability

tossing of 3 coins

HHH THH $E$ : getting atleast two heads $P(E) : \frac{4}{8}$

HHT THT

HTH TTH $F$ : getting atleast two tails $P(F) : \frac{4}{8}$

HTT TTT

$$P(E \cap F) = \frac{1}{8}$$

$$P(E/F) = \frac{1}{4}$$

$$P(E/F) = \frac{P(E \cap F)}{P(F)}$$

- Let A & B event such that $P(A) > 0$ then conditional probability of event B given that event A has already occured is that

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

Page No:
Date:

## * Multiplication of Probability:

- For Any two arbitary events A & B the following eq$^n$

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

## * Law of Total Probability:

Assume that $A_1, A_2 \cdots\cdots A_k$ are given such that $A_1 \cup A_2 \cdots \cup A_k = \Omega$ and $A_i \cap A_j = \phi$ for all $i \neq j$ $P(A_i) > 0$ for all $A_1, A_2, \cdots, A_k$ forms a complex decomposition of $\Omega$ in pairwise Disjoint events. then the probability of an event B

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + \cdots + P(B|A_k) \cdot P(A_k)$$

$$= \sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)$$

Q. In a class 35% of students studied science and history. 65% of students study science. What is the probability of a student studing history even that he or she already studying science.

$\rightarrow$ $$P(H|S) = \frac{P(H \cap S)}{P(S)} = \frac{0.35}{0.65} = \underline{\underline{0.538}}$$

Page No:
Date:

- ## Bayes' Theorem

It gives relation between probability $(A|B)$ and $P(B|A)$.

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

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

$$= P(B|A) \times \frac{P(A)}{P(B)}$$

$$P(A|B) = P(B|A) \times \frac{P(A)}{P(B)}.$$

Note :- let $A_1, A_2, \cdots, A_k$ be the events such that $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$ and $A_i \cap A_j = \phi$ $\forall i \neq j$.

$P(A_i) > 0$ $\forall i$

B is another event, then $$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{j=1}^{k} P(B|A_j) \cdot P(A_j)}$$

$P(A_j)$ = Prior Probability
$P(B|A_j)$ = Model Probability
$P(A_j|B)$ = Posterior Probability.

Page No:
Date:

- **Idenpendent Events** – also called stochastically Independent Events

Two random events A and B are called independent events if $P(A \cap B) = P(A) \cdot P(B)$ i.e if the probability of simultaneous occurings of both the events A and B is taken then it is the product of individual probablities of A and B.

- **Mutually Independent Events**

The events $A_1, A_2, \ldots A_k$ are mutually independent events if for any m events $A_{i_1}, A_{i_2}, \ldots A_{i_k}$ $(m \le k)$

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_m}) = P(A_{i_1})(A_{i_2}) \cdots P(A_{i_m})$$

Q Consider a bag with 4 balls the following combination of 0's + 1's are printed on the balls.

110, 101, 011, 000

One ball is drawn from the bag now consider the following events

$A_1$ : The $1^{st}$ digit is 1

$A_2$ : The $2^{nd}$ digit is 1

$A_3$ : $3^{rd}$ — 1 —

Find whether the events are independent or not.

→ $P(A_1) = 1/2$

$P(A_2) = 1/2$

$P(A_3) = 1/2$

$(PA_1 \cap PA_2 \cap PA_3) = \frac{0}{8} = 0$

Show that these events are pair wise independent.

Page No:
Date:

- **Random Variables**

Let $\Omega$ represents the sample space of a random experiment. Let R denotes the set of real numbers then a random variable is a function denoted by 'X' which assigns each element of sample space one and only one real number i.e. $X : \Omega \longrightarrow R$

$\Omega$ Tossing of 3 coins

HHH → 3

HHT → 2

HTH → 2

HTT → 1

TTT → 0

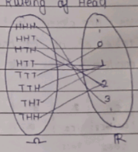
TTH → 1

THT → 1

THH → 2

Ruleing of Head

X :




- Cumulative Distribution Function (CDF) –

The CDF of a Random Variable, X is defined as

$$f(x) = P(X \le x)$$

f(x) satisfies the following properties.

1) $f(x)$ is a monotonically non-decreasing function.
2) $\lim_{x \to -\infty} f(x) = 0$ ( lower limit is 0 )
3) $\lim_{x \to +\infty} f(x) = 1$ ( upper limit is 1 )
4) $f(x)$ is continuous from right.
5) $0 \le f(x) \le 1 \ \forall x \in \mathbb{R}$.

Distribution table :-

eg.

| X (no. of ) | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| P(x) | 1/8 | 3/8 | 3/8 | 1/8 |

$$f(x) = \begin{cases} 0 & x < 0 \\ 1/8 & x \le 0 \\ 4/8 & x \le 1 \\ 7/8 & x \le 2 \\ 1 & x \le 3 \\ 1 & x \le f\infty \end{cases}$$

$-\infty$ ↓ $+\infty$

Q. 4 bag apples are placed with 16 good apples find probability distribution table of no. of bag apples when 2 apples are drawn at a random. Also finds its CDF.

→

| x | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| P(x) | $\frac{16C2}{20C2}$ | $\frac{4C1\,16C2}{20C2}$ | $\frac{1C2}{20C2}$ | 0 | 0 |

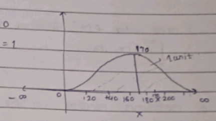
$$F(x) = \begin{cases} \frac{16C2}{20C2} & x < 0 \\ \frac{4C1 \cdot 16C2}{C2} + \frac{16C2}{20C2} & x \le 1 \\ \frac{16C2}{20C2} + \frac{4C1 \cdot 16C2}{20C2} \cdot \frac{1C2}{20C2} & x \le 2 \\ 1 + 0 & x \le 3 \\ 1 + 0 & x \le 4 \end{cases}$$

Page No:
Date: 21 08 24

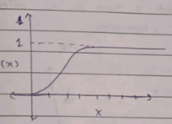
Ex. Consider the height of students of 20 years of age. Graphs for P.D.F and C.D.F

PDF i) $f(x) \ge 0$

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




CDF




F(x)

PMF

| x | 0 | 1 | 3 |
|---|---|---|---|
| P(x) | 0.1 | 0.2 | 0.7 |

f(x)

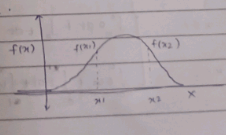
- Theorem

let x be a continuous Random Variable with CDF F(x) If $x_1 < x_2$, where $x_1, x_2$ are unknown constants then

$$P(x_1 \le x \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$$

Proof

Consider the graph of PDF of a continuous random variable $x$.



$P(x_1 < x \le x_2) = F(x_2) - F(x_1)$

$= $ (area under $f(x)$ from $-\infty$ to $x_2$) $-$ (area under $f(x)$ from $-\infty$ to $x_1$)

$=$ area under $f(x)$ from $x_1$ to $x_2$

$= \int_{x_1}^{x_2} f(x)\,dx$.

Q.1 Consider a continuous random variable X : " waiting time of the Train ". Assume that a train arrives after every 20 min. Find the PDF of given information.

→ PDF

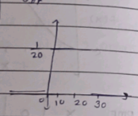
$$f(x) = \begin{cases} k & x \in [0, 20] \\ 0 & \text{else} \end{cases}$$

i) $f(x) \ge 0$

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

$k = 1/20$

CDF



Q.2 If $f(x) = \begin{cases} x & 0 \le x < 1 \\ 2-x & 1 \le x < 2 \\ 0 & x \ge 2 \end{cases}$ is a PDF,

$\int_{-\infty}^{0}$, $\int_0^1 x$, $\int_1^2 x-2$

Then find its CDF.

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

$$F(x) = \begin{cases} \int_{-\infty}^{x} f(t)\,dt & x \le 0 \\ \int_{-\infty}^{x} f(t)\,dt & x \le 1 \\ \int_{-\infty}^{x} f(t)\,dt & x \le 2 \end{cases} = \begin{cases} \int_{-\infty}^{x} 0\,dt = 0 & x \le 0 \\ \int_{-\infty}^{0} 0\,dt + \int_0^x t\,dt & x \le 1 \\ \int_{-\infty}^{0} 0\,dt + \int_0^1 t\,dt + \int_1^x (2-t)\,dt & x \le 2 \end{cases}$$

$$\begin{cases} 0 & x \le 0 \\ \left[t^2/2\right]_0^x = \frac{x^2}{2} - 0 = \frac{x^2}{2} & x \le 1 \\ x^2/2 + \left[2t - \frac{t^2}{2}\right]_1^x = 2x - \frac{x^2}{2} - 2 - \frac{1}{2} = 0 + \frac{x^2}{2} + 2x - \frac{x^2}{2} - 2 - \frac{1}{2} = -\frac{1}{2} & x \le 2 \end{cases}$$

Q.1] CDF

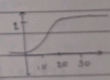
$F(x) = \int_{-\infty}^{x} f(t)\,dt$

$$F(x) = \begin{cases} \int_{-\infty}^{x} f(t)\,dt & x \le 0 \\ \int_{-\infty}^{x} f(t)\,dt & x \le 20 \end{cases}$$

$\int_{-\infty}^{0} 0$, $\int_0^{20} 1/20$, $\int_{20}^{\infty} 0$

$$f(x) \begin{cases} \int_{-\infty}^{x} 0\,dt = 0 & x \le 0 \\ \int_{-\infty}^{0} 0\,dt + \int_0^x 1/20\,dt & x \le 20 \end{cases}$$

$$f(x) \begin{cases} 0 & x \le 0 \\ x/20 & x \le 20 \\ 1 & x > 20 \end{cases}$$



Q.3] Find pdf where cdf is : $F(x) = \begin{cases} 0 & x \le 1 \\ k(x-1)^4 & 1 < x \le 3 \\ 1 & x > 3 \end{cases}$

Also find k.

→ pdf i) $f(x) \ge 0$

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

$f(x) = \Big\{$ X : is continuous Random Variable, F(x) will be continuous

Page No:
Date:

At $x = 1$

$LHS = 0$

$RHS = K(1-1)^4 = 0$

$F(1) = 0$

At $x = 3$

$LHS = K(3-1)^4$

$RHS = 1$

$f(3) = K(3-1)^4$

$K(3-1)^4 = 1$

$\Rightarrow K \cdot 2^4 \Rightarrow K \cdot 16 = K = \frac{1}{16}$.

CDF $F(x) = \; -\infty \underbrace{\quad}_{0} 1 \underbrace{\quad}_{\frac{1}{16}(x-1)^4} 3 \underbrace{\quad}_{1} \infty$

To get pdf i.e $f(x)$ we will differentiate.

$f(x) \; -\infty \underbrace{\quad}_{0} 1 \underbrace{\quad}_{\frac{4}{16}(x-1)^3} 3 \underbrace{\quad}_{0} \infty$

pdf. $f(x) = \begin{cases} 0 & x < 1 \\ \frac{4}{16}(x-1)^3 & 1 \le x \le 3 \\ 0 & x > 3 \end{cases}$

$$f(x) = \begin{cases} 4/16\,(x-1)^3 & 1 \le x \le 3 \\ 0 & \text{else} \end{cases}$$

Page No:
Date: 26 08 24

* Expectation of a Random Variable.

1) Expectation of a Discrete Random Variable -

For a discrete random variable, X which takes the value $x_1, x_2, \ldots, x_n$ with respective probabilities $p_1, p_2, \ldots p_k$, the expectation of X is

$$E(X) = \sum_i x_i p_i$$

$$= x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$

2) Expectation of a Continuous Random Variable -

let X be a continuous Random Variable having PDF $f(x)$ with $\int |x| \cdot f(x)\,dx < \infty$,

Expectation is

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

Note: Expectation is also denoted by $\mu$. i.e $E(x) = \mu$.

* Variance :-

1. It describes the variability of a Random Variable
2. It tells about the concentration or Dispersion of values around the Arithmetic mean of Distribution.
3. Variance is given by

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

4. Variance for Discrete -

$$Var(x) = \sum_i (x_i - E(x))^2 \cdot p_i$$

5. Variance for Continuous

$$Var(x) = \int_{-\infty}^{\infty} (x - E(x)^2) \cdot f(x)\,dx.$$

Page No:
Date:

* Standard Deviation.

The positive square root of variance is called as standard Deviation denoted by $\sigma$.

i.e. $\sigma^2 = Var(x)$.

1) $\sigma$ measure how the value of Random Variable are dispressed.
2) A low value of $\sigma$ indicates that the values are highly concentrated around the mean.
3) A high value of $\sigma$ indicates lower concentration of data values around the mean.

| Discrete | Continuous. |
|---|---|
| 1) $E(x) = \sum x_i p_i$ | $E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$ |
| 2) $Var(x) = \sum (x_i - E(x))^2 p_i$ | $Var(x) = \int_{-\infty}^{\infty} (x - E(x))^2 \cdot f(x)\, dx$. |
| 3) $\sigma = \sqrt{var(x)}$ | |

Q.1] Experiment : Rolling of a dice.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p(x)$ | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6. |

find $E(x)$, $var(x)$, $\sigma$.

$\rightarrow$ $E(x) = \sum x_i p_i = \sum 21 \times \frac{1}{6} = \frac{21}{6} = 3.5$

$Var(x) = \sum (x_i - E(x)^2) p_i = \sum \left(1 - \frac{21}{6}\right)^2 \cdot \frac{1}{6} + (2 - 21/6)^2 + (3 - 21/6)^2 = 2.91$

$\sigma = \sqrt{var(x)} = 1.707$

Page No:
Date:

Q 2] $x$ : " Waiting time of the Train " where train arives after every 20 min.

PDF

$$f(x) = \begin{cases} k & 0 \le x \le 20 \\ 0 & \text{else} \end{cases} \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

find $E(x)$, $Var(x)$, $\sigma$.

$\rightarrow$ $E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{0}^{20} 1 \cdot 1 = \int_{0}^{20} 1 = \frac{20^2}{2} = \frac{400}{2} = 10$

$Var(x) = \int_{-\infty}^{\infty} (x - E(x))^2 \cdot f(x)\, dx = \int_{0}^{20} (1 - \quad = 33.34$

$\sigma = 5.77$

Theorem -

The Variance of a Random Variable $x$ is expressed as

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

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

Consider $E(x) = \mu$

$Var(x) = E(x - \mu)^2$

$\rightarrow$ $Var(x) = E(x^2 - 2\mu x + \mu^2)$

$= E(x^2) - E(2\mu x) + E(\mu^2)$

$= E(x^2) - 2\mu E(x) + E(\mu^2)$

$= E(x^2) - 2\mu \cdot \mu + E(\mu^2)$

$= E(x^2) - 2\mu^2 + \mu^2$

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

Q. 1) Var(x) using formula from proof.

$\mathrm{Var}(x) = E(x^2) - (E(x))^2$

$= E(1^2) + E(2^2) + E(3^2) + E(4^2) + E(5^2) + E(6^2) - (E(3.5))^2$

$= E(1) + E(4) + E(9) + E(16) + E(25) + E(36) - (E(3.5))^2$

$= E(91) - E(12.25)$

$= 78.75$

31-08-24

## * Calculation Rules for Expectations

For any constants a and b and any Random Variable x and y the following holds -

1) $E(a) = a$
2) $E(bx) = bE(x)$
3) $E(a+bx) = a + bE(x)$
4) $E(x+y) = E(x) + E(y)$.

| Discrete | Continuous |
|---|---|
| $E(x) = \sum x_i p_i$ | $E(x) = \int_{-\infty}^{\infty} x(f(x))dx$ |
| $\mathrm{var}(x) = \sum (x - E(x))^2 p_i$ | $\mathrm{var}(x) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)$ |

Proof.

3) $E(a+bx) = a + bE(x)$

case 1 : x is discrete Random Variable

$E(a+bx) = \sum (a + bx_i) p_i$

$= \sum (ap_i + bx_i p_i)$

$= \sum ap_i + \sum bx_i p_i$

$= a \sum p_i + b \sum x_i p_i$

$= a \cdot 1 + bE(x)$ .... $\because \sum p_i = 1$ & $\sum x_i p_i = E(x)$

$= a + bE(x)$

Case 2 : x is continuous.

$E(a+bx) = \int_{-\infty}^{\infty} (a + bx) f_x \, dx$

$= \int_{-\infty}^{\infty} (af_x + bx_i f_x) \, dx$

$= \int_{-\infty}^{\infty} af_x \, dx + \int_{-\infty}^{\infty} bx_i f_x \, dx$

$= a \int_{-\infty}^{\infty} f_x \, dx + b \int_{-\infty}^{\infty} x_i f_x \, dx$

$= a \cdot 1 + bE(x)$

$= a + bE(x)$

1) $E(a) = a$

case 1 :- $E(a) = \sum a x_i p_i$

$= a \sum x_i p_i$

$= a E(x)$.

case 2 :- $E(a) = \int_{-\infty}^{\infty} a x_i f_x dx$

$= a \int_{-\infty}^{\infty} x_i f_x dx$

$= a E(x)$.

4) $E(x+y) = E(x) + E(y)$.

case 1 :- $E(x+y) = \sum (x+y) p_i$

$= \sum (x_i p_i + y_i p_i)$

$= \sum x_i p_i + \sum y_i p_i$

$= E(x) + E(y)$.

case 2 :- $= \int_{-\infty}^{\infty} (x+y) f_x dx$

$= \int_{-\infty}^{\infty} ((x_i + y_i) f_x) \, dx$

$= \int_{-\infty}^{\infty} (x_i f_x + y_i f_x) dx$

$= \int_{-\infty}^{\infty} x_i f_x dx + \int_{-\infty}^{\infty} y_i f_x dx$

$= E(x) + E(y)$

2) $E(bx) = bE(x)$

case 1 :- $E(bx) = \sum b x_i p_i$

$= b \sum x_i p_i$

$= bE(x)$

case 2 :- $= \int_{-\infty}^{\infty} b x_i f_x dx$

$= b \int_{-\infty}^{\infty} x_i f_x dx$

$= bE(x)$

$E(a+bx) = a + bE(x)$

Page No: Date:

* Rules for Variance for any constants a and b and for any Random Variables X and Y, following holds -

1) $var(a) = 0$
2) $var(bx) = b^2 var(x)$
3) $var(a+bx) = b^2 var(x)$

Proof -

3) $Var(a+bx) = b^2 var(x)$ (Theorem $var(x) = (E(x^2) - (E(x))^2)$)

$\rightarrow var(a+bx) = E((a+bx)^2) - (E(a+bx))^2$

$= E(a^2 + 2ab + b^2x^2) - (a^2 + 2abE(x))^2$

$= E(a^2 + 2ab + b^2x^2) - (a + bE(x))^2$

$= a^2 + 2abE(x) + E(b^2x^2) - a^2 + 2abE(x) + b^2(E(x))^2$

$= \cancel{a^2} + 2ab\cancel{E(x)} + b^2E(x^2) - \cancel{a^2} - 2ab\cancel{E(x)} - b^2(E(x))^2$

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

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

$= b^2 var(x)$

Ex. X = "Waiting time of the Train", where train arrives after every 20 min.

Y = "Waiting time of the Bus to go to the Train Satation", where bus arrives after every 60 min

$\rightarrow$ $E(x_1)$, $Var(x)$ $\quad F(x) = \begin{cases} k_1 & 0 < x < 20 \\ 0 & else \end{cases}$ $\quad k_1 = \frac{1}{20}$

$E(y)$, $Var(y)$ $\quad F(y) = \begin{cases} k_2 & 0 < y < 60 \\ 0 & else \end{cases}$ $\quad k_2 = \frac{1}{60}$

$E(x) = 10$

$var(x) = 33.33$

$y = 3x$

$E(y) = E(3x) = 3 \cdot 10 = 30$

$var(y) = var(3x) = 3^2 \cdot var(x)$

$= 9 \cdot (33.33)$

Page No: Date:

Definition - Independently Identically Distributed (IId) -

Consider the Random Variable $x_1, x_2, \ldots, x_n$

If all the $x_i$ is follows the same Distribution and are stochastically independent, then they're s.t.b in i.i.d

* Expectation & Variance for the Arithmetic Mean -

let $x_1, x_2, \ldots, x_n$ be in i.id

let $E(x_i) = \mu$, $var(x_i) = \sigma^2$ for each $i = 1, 2, \ldots, n$

We know that Arithmetic Mean -

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

1) $E(x) =$

$E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = E\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right)$

$\therefore (E(bx) = bE(x))$ $\quad \because E(x+y) = E(x) + E(y)$

$= \frac{1}{n} E(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}(E(x_1) + E(x_2) + \cdots + E(x_n))$

$E(\bar{x}) = \frac{1}{n}(\mu + \mu + \mu + \cdots + \mu) = \frac{1}{n}(n\mu) = \mu$

$$\boxed{E(\bar{x}) = \mu}$$

2) $var(\bar{x})$

$var(\bar{x}) = Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = Var\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right)$

$(var(bx) = b^2 var x)$

$= \frac{1}{n^2}(var(x_1 + x_2 + \cdots + x_n)) = \frac{1}{n^2}(var\, x_1 + var\, x_2 + \cdots + var\, x_n)$

$= \frac{1}{n^2}(\underbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}_{n\, times}) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} = \frac{var(x_i)}{n}$

$$\boxed{var(\bar{x}) = \frac{var(x_i)}{n}}$$

Page No:
Date:

Ex. Tossing of a coin
coin 1 outcomes : H / T
0 / 1

$x_1$ :

pmF

| $x_1$ | 0 | 1 |
|---|---|---|
| $P(x_1)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

$P(1) = 1/2$
$P(0) = 1/2$
$E(x_1) = \frac{1}{2}$

$var(x_1) = \frac{1}{4}$

$x_2$ :

pmF

| $x_2$ | 0 | 1 |
|---|---|---|
| $P(x_2)$ | 1/2 | 1/2 |

$E(x_2) = 1/2$
$var(x_2) = \frac{1}{4}$

⋮

$x_n$

pmF

| $x_n$ | 0 | 1 |
|---|---|---|
| $P(x_n)$ | 1/2 | 1/2 |

$$\bar{x} = \frac{x_1 + x_2 + \quad + x_n}{n}$$

$$E(\bar{x}) = \frac{1}{2} = E(x_i)$$

$$var(\bar{x}) = \frac{var(x_i)}{n} = \frac{1}{4n}$$

Note : Statistic - A function of Random Variable is called "Statistic".

02/09/24

Page No:
Date: 20

- Types of Discrete Distribution

1) Bernoullis Distribution -
It is used only when there are 2 possible outcomes and our interest lies in any one of the 2 outcomes.

Definition - A Random Variable, X has a Bernollis distribution if the PMF of X is given as

$$f(x) = \begin{cases} p^{\text{ expection}} & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

| x | 1 | 0 |
|---|---|---|
| P | p | 1-p |

where p is the probability of success.

ii) The CDF of x is

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x \le 1 \\ (P)1 & x \ge 1 \end{cases}$$

iii) The Mean or expectation of x is imp

$$E(x) = p$$

$E(x) = \sum x_i p_i$

iv) The Variance of x is imp

$$Var(x) = p(1-p) = p - p^2$$

Q. A company organizes a raffle at an end of the year function there are three hundreed lottery tickets in total and 50 of them are marked as winning tickets. The event A of intrest is
A : " Ticket wins " (coded as x=1)
and if the complementary event occurs then it is coded as x=0. Find :-(i) pmf ii) Expected value of x
iii) Variance of x.

$\rightarrow$ $P(X=1) = P(\text{Ticket wins}) = \dfrac{\text{no. of winning Ticket}}{\text{Total no. of Ticket}}$

$= \dfrac{50}{300} = \dfrac{1}{6}$

$P(X=0) = 1 - \dfrac{1}{6} = \dfrac{5}{6}$

i) PMF.

| $X$ | 1 | 0 |
|---|---|---|
| $P(X)$ | $\frac{1}{6}$ | $5/6$ |

ii) $E(X) = \dfrac{1}{6}$

$Var(X) = \dfrac{1}{6}\left(1 - \dfrac{1}{6}\right) = \dfrac{1}{6} \cdot \dfrac{5}{6} = \dfrac{5}{36}$

2) Binomial Distribution –

It is used when number of trials are more but there are only two possible outcomes.

Definition – A Random Variable, X has a Binomial Distribution if the PMF of X is given as

$$P(X=r) = {}^nC_r (P)^r (1-P)^{n-r}$$

where p is the probability of success.

ii) We can also write

$X \sim B(n, p)$; $n$ = Total no of Trials

$p$ = proof of success.

iii) The Mean or Expectation of X is or Expected value of X

$E(X) = np$.

iv) The Variance of X is

$Var(X) = np(1-p)$

fair coin
$P(H) = \frac{1}{2}$, $P(T) = \frac{1}{2}$ for unfair coin $P(H) \neq \frac{1}{2}$

Q.1 A fair coin is tossed 10 times find the probability of getting 2 tails. $\rightarrow \frac{1}{2}$

$\rightarrow$ $n = 10$, $r = 2$

$P(X=2) = {}^nC_r p^r (1-p)^{n-r}$ $\therefore$ p → probability of tails

$\downarrow$
no. of tails

$= \cancel{{}^{10}C_2 p^2}\ {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{10-2}$

$= {}^{10}C_2 \frac{1}{4}\left(1 - \frac{1}{2}\right)^8 = {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^8$

$= {}^{10}C_2 \frac{1}{4}\left(\frac{1}{2}\right)^8 = \frac{10 \times 9}{2} \times \frac{1}{1024}$

$= \dfrac{45}{1024}$

Q.2 The probability of a man hiting a target is $\frac{1}{4}$. He fires 7 times. What is the probability of his hitting the target atleast twice.

$P(X \geq 2) = P(X=2) + P(X=3) + \cdots + P(X=7)$

$\rightarrow$ $n = 7$ $r = 2$

$P(X \geq 2) = {}^7C_2 \left(\frac{1}{4}\right)^2 \left(1 - \frac{1}{4}\right)^{7-2}$

$= {}^7C_2 \left(\frac{1}{4}\right)^2 \left(1 - \frac{1}{4}\right)^5$

$= {}^7C_2 \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^5$

OR

$\boxed{P(X \geq a) = 1 - P(X < a)}$

$P(X \geq 2) = 1 - P(X < 2)$

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

$=$