5/8/21

Program: Master of Computer Applications

Curriculum Scheme: MCA 2 year Course

Examination: MCA 1 SEMESTER II

Course Code: MCA21 and Course Name: Mathematical Foundation for Computer Science 2

Time: 2 HRS

Max. Marks: 80

Section I - MCQS (40 Marks) – 40 Minutes Section II – Subjective (40 Marks) – 80 Minutes

The timings If the Examination Time is 10:00 am to 12:00 noon

Section I – 11:00 am – 11:40 am Section II – 11:40 am – 1:00 pm

==

SECTION II

Q2.Solve any two out of three (10 Marks each)

a. Use a graphical method to solve the following LP problem.

Maximize $Z = 3X_1 + 2X_2$

Subject to constraints

$X_1 - X_2 >= 1$

$X_1 + X_2 >= 3$

$X_1, X_2 >= 0$

b. Use VAM to solve the following transportation problem

Distribution Centre

		D1	D2	D3	D4
Plant	P1	2	3	11	7
	P2	1	0	6	1
	P3	5	8	15	9

c. In a supermarket four salesmen A, B, C, D are available for four counters W, X, Y, Z. Each salesman can handle one counter at a time. Service time in (hrs) by the salesmen for each counter is given below. Assign the salesman to the counters so that the service time is minimized.

		Salesman			
		A	B	C	D
Counters	W	5	3	2	8
	X	7	9	2	6
	Y	6	4	5	7
	Z	5	7	7	8

Q3. Solve any two out of three (10 Marks each)

a. Use Simplex method to solve

Maximize $Z = 16 X_1 + 17 X_2 + 10 X_3$

Subject to constraints

$X_1 + X_2 + 4X_3 <= 200$

$2X_1 + X_2 + X_3 <= 360$

$X_1 + 2X_2 + 2X_3 <= 240$

$X_1, X_2, X_3 >= 0$

b. A garage mechanic finds the time spent on his jobs has exponential distribution with mean 45 minutes. If he repairs cars in the order in which they come in, which follows poisson distribution with a mean of 5 per 8 hour day. What is the mechanic's idle time each day. How many jobs are ahead of the average number of cars which came in?

c. Find the optimum strategy for Player A & B and value of the game where payoff matrix is given as follows

		Player B		
		B1	B2	B3
Player A	A1	7	3	1
	A2	1	7	3
	A3	0	1	7

Program: Master of Computer Applications

Curriculum Scheme: MCA 2 YEAR COURSE

Examination: MCA SECOND YEAR SEMESTER-2 JANUARY-2022 (ATKT)

Course Code: MCA21 and Course Name: Mathematical Foundation for Computer Science 2

Time: 2:00 pm to 4:00 pm (2 Hrs) Max. Marks: 80

Section I - MCQS (40 Marks) – 40 Minutes (2:00 pm to 2:40 am)

Section II – Subjective (40 Marks) – 80 Minutes (2:40 am to 4:00 pm)

==

SECTION II

Q.2 Solve any **TWO** questions out of three questions. **[20 Marks]**

1. Use simplex method to solve the following LPP

Maximize Z = 30x1+40x2

subject to $60x_1 + 120x_2 \leq 12000$

$8x_1 + 5x_2 \leq 600$

$3x_1 + 4x_2 \leq 500$

and $x_1, x_2 \geq 0$

2. Reduce the game by dominance property and solve using algebraic method

	Player B			
Player A	1	7	2	4
	0	3	7	8
	5	2	6	10

3. A company has factories F1, F2 and F3 which supply warehouses W1, W2, W3. Weekly factory capacities are 200, 160 and 90 units respectively. Weekly warehouse requirements are 180, 120 and 150 units respectively.

Find the initital basic feasible solution (IBFS) using Vogel's Approximation Method. Unit shipping costs are

	W1	W2	W3
F1	16	20	12
F2	14	8	18
F3	26	24	16

Q.3 Solve any **TWO** questions out of three questions. **[20 Marks]**

1. Using two-phase method solve the following LPP

Maximize Z = 2x1 + 3x2 - 5x3

subject to x1 + x2 + x3 = 7

2x1 - 5x2 + x3 ≥ 10

and x1, x2, x3 ≥ 0

2. An automobile dealer wishes to put four repairmen to four jobs. The repairmen have somewhat different kinds of skills and they exhibit different levels of efficiency from one job to another. The dealer has estimated the number of man hours that would be required for each man job combination. This is given in the matrix form in the following table.

		Jobs			
		A	B	C	D
Men	1	5	3	2	8
	2	7	9	2	6
	3	6	4	5	7
	4	5	7	7	8

Find the optimum assignment that will result in minimum man hours required.

3. At a railway reservation booking window, customers arrive randomly at the average rate of 16 per hour approximated to Poisson's distribution. If service time is exponentially distributed with a mean of 20 per hour, determine:

a) Percentage utilization capacity

b) Probability that there are at least 3 customers in the queue

c) Average time spent in the system

d) Average number of customers waiting in the line

e) Probability that there are 5 customers in the system

MATHEMATICAL FOUNDATION FOR COMPUTER SCIENCE 1

Subject Incharge - Ruchi Rautela
ruchi.rautela@ves.ac.in

Semester I

<table>
<tr><th>Course</th><th>Course Name</th><th colspan="3">Teaching Scheme</th><th colspan="3" rowspan="2">Credits Assigned</th></tr>
<tr><td rowspan="7">MCA11</td><td rowspan="7">Mathematical Foundation for Computer Science 1</td><td colspan="3">Contact Hours</td></tr>
<tr><td colspan="2">Theory</td><td>Tutorial</td><td>Theory</td><td>Tutorial</td><td>Total</td></tr>
<tr><td colspan="2">3</td><td>--</td><td>3</td><td>--</td><td>3</td></tr>
<tr><td colspan="6">Examination Scheme</td></tr>
<tr><td colspan="3">Theory</td><td rowspan="2">Term Work</td><td rowspan="2">End Sem Exam</td><td rowspan="2">Total</td></tr>
<tr><td>CA</td><td>Test</td><td>AVG</td></tr>
<tr><td>20</td><td>20</td><td>20</td><td>25</td><td>80</td><td>125</td></tr>
</table>

PRE-REQUISITE

- Measures Of Central Tendency And Dispersion
- Set Theory
- Basic Principles Of Counting

Module 1- Skewness

- Karl Pearson’s coefficient of skewness,
- Bowley's coefficient of skewness.

Self Learning Topics: Determining skewness of data related to real system and its graphical representation.

Course Objectives: Learner/ student will learn & perform.

S. No.	Course Objective
1	Statistical measures on various types of data

Course Outcomes: On successful completion of course learner/student will be able to -

Sr.No.	Outcome	Bloom Level
1	Apply different statistical measures on various types of data	Applying

Reference Books:

Sr. No.	Reference Name
1	S C Gupta, Fundamentals of Statistics, Himalaya Publishing house, Seventh edition.
2	S.C.Gupta, V.K.Kapoor , S Chand , Fundamentals of Mathematical Statistics, Sultam and Chand sons publication, First Edition

Measures Of Central Tendency (Pre-requisite)

- Mean
- Median
- Mode

Mean

Mean is the average of a set of data. To calculate mean find the sum of data and then divide it by number of data.

12, 15, 11, 11, 7, 13

First, find the sum of the data.

12 + 15 +11 + 11 + 7 + 13 = 69

Then divide by the number of data.

69 / 6 = 11.5

The mean is 11.5

Solve

An electronics store sells CD players at the following prices: \$350, \$275, \$500, \$325, \$100, \$375, and \$300. What is the mean price?

Find your answer before clicking!

\$350 + \$275 + \$500 + \$325 + \$100 +\$375 + \$300 = \$2225

\$2225 / 7 = \$317.86

The mean or average price of a CD player is \$317.86.

Median

Median is middle number in a set of data when the data is arranged in numerical order.

12, 15, 11, 11, 7, 13

First, arrange the data in numerical order.

7, 11, 11, 12, 13, 15

Then find the number in the middle or the average of the two numbers in the middle.

11 + 12 = 23 23 / 2 = 11.5

The median is 11.5

Solve

An electronics store sells CD players at the following prices: \$350, \$275, \$500, \$325, \$100, \$375, and \$300. What is the median price?

Find your answer before clicking!

First place the prices in numerical order.

\$100, \$275, \$300, \$325, \$350, \$375, \$500

The price in the middle is the median price.

The median price is \$325.

Mode

Mode is the number that occurs most in the set of data.

12, 15, 11, 11, 7, 13

The mode is 11.

Sometimes a set of data will have more than one mode.

For example, in the following set the numbers both the numbers 5 and 7 appear twice.

2, 9, 5, 7, 8, 6, 4, 7, 5

5 and 7 are both the mode and this set is said to be bimodal.

Mode

Sometimes there is no mode in a set of data.

3, 8, 7, 6, 12, 11, 2, 1

All the numbers in this set occur only once therefore there is no mode in this set.

Solve

$100, $275, $300, $325, $350, $375, $500

What is the mode ?

Find your answer before clicking!

There is no mode!

Summary

Mean → The average

Median → The number or average of the numbers in the middle

Mode → The number that occurs most

PROBLEMS

1. Find the mean of the following data.

(a) 9, 7, 11, 13, 2, 4, 5, 5

(b) 16, 18, 19, 21, 23, 23, 27, 29, 29, 35

(c) 2.2, 10.2, 14.7, 5.9, 4.9, 11.1, 10.5

(d) $1\,{}^{1}/_{4}$, $2\,{}^{1}/_{2}$, $5\,{}^{1}/_{2}$, $3\,{}^{1}/_{4}$, $2\,{}^{1}/_{2}$

PROBLEMS

2. Find the mode of the following data.

(a) 12, 8, 4, 8, 1, 8, 9, 11, 9, 10, 12, 8

(b) 15, 22, 17, 19, 22, 17, 29, 24, 17, 15

(c) 0, 3, 2, 1, 3, 5, 4, 3, 42, 1, 2, 0

(d) 1, 7, 2, 4, 5, 9, 8, 3

PROBLEMS

3. Find the median of the following data.

(a) 27, 39, 49, 20, 21, 28, 38

(b) 10, 19, 54, 80, 15, 16

(c) 47, 41, 52, 43, 56, 35, 49, 55, 42

(d) 12, 17, 3, 14, 5, 8, 7, 15

MEAN

MEAN = $\sum(X_i \times f_i) / \sum f_i$

Here X_i is data and f_i is frequency.

e.g

X_i : 2, 5 , 7, 1, 3, 9, 6, 4

f_i : 4, 3, 1, 2, 1, 2, 1, 2

MEAN = {(2 x 4)+(5 x 3)+(7 x 1)+(1 x 2)+(3 x 1)+(9 x 2)+(6 x 1)+(4 x 2)}/(4+3+1+2+1+2+1+2)

MEAN =67/16= 4.1875

MEAN-For Grouped Data

Grouped Frequency Table

Seconds	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4

Midpoint	Frequency	Midpoint × Frequency fx
53	2	106
58	7	406
63	8	504
68	4	272
		1288

$$\text{MEAN} = \sum(X_i \times f_i) / \sum f_i$$

$$\textbf{Estimated Mean} = \frac{1288}{21} = \mathbf{61.333}$$

$$\sum f_i = N$$

MEDIAN-Grouped Data

Seconds	Frequency	C.F.
51 - 55	2	2
56 - 60	7	9
61 - 65	8	17
66 - 70	4	21

F

N - SUM OF FREQUENCY

L1- LOWER LIMIT OF MEDIAN CLASS

L2- UPPER LIMIT OF MEDIAN CLASS

F- C.F OF PRE MEDIAN CLASS

$$N/2 = \sum f_i / 2 = 10.5$$

$$\text{MEDIAN} = L1 + \left[\frac{N}{2} - F\right] * \left[\frac{L2-L1}{f}\right] = 61 + \left[\frac{21}{2} - 9\right] * \left[\frac{65-61}{8}\right]$$

$$= 61.75$$

MODE-Grouped Data

Seconds	Frequency
51 - 55	2
56 - 60	7 ← f_1
61 - 65	8 ← f
66 - 70	4 ← f_2

MODAL CLASS

L1- LOWER LIMIT OF MODAL CLASS
L2- UPPER LIMIT OF MODAL CLASS
d1,d2 - DIFFERENCE B/W THE FREQUENCY OF MODAL CLASS AND PREMODAL ,POSTMODAL CLASS

$d1=f - f_1$

$d2=f -f_2$

$$MODE = L1 + \left[\frac{d1}{d1 + d2}\right] * (L2 - L1)$$

MODE = 61.8

PROBLEMS

1. Find the mean,median,mode of the following data.

Age Groups	Frequency f_i
0 - 10	40
10 - 20	53
20 - 30	58
30 - 40	64
40 - 50	72
50 - 60	49
60 - 70	36
70 - 80	25
Total	**397**

Mean = 36.9
Median = 37.4
Mode = 42.6

Practice Problem 1

Find the group mean of the following data.

Scores	**Frequency**
1-20	5
21 - 40	20
41 - 60	47
61 - 80	15
81 - 100	3

Practice Problem 2

Find the group mode of the following data.

Students	**Frequency**
1-3	135
4-6	457
7-9	549
10-12	392

Relation between mean,median and mode

$$\textbf{Mode} = 3(\text{Median}) - 2(\text{Mean})$$

Measures of Dispersion

The measure of **dispersion** shows how the data is spread or scattered around the mean.

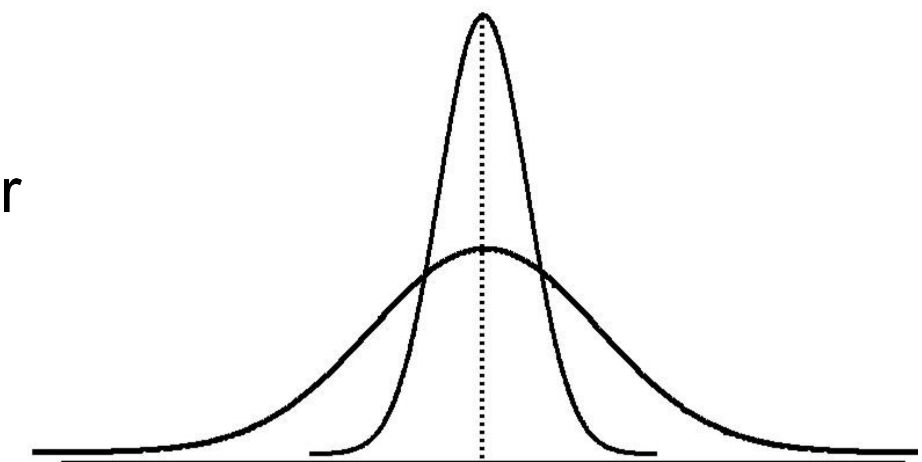
- Range
- Quartiles
- Variance
- Standard Deviation
-
- Coefficient of Variation

Measures of Dispersion

Variation

- **Range**
- **Variance**
- **Standard Deviation**
- **Coefficient of Variation**

Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.



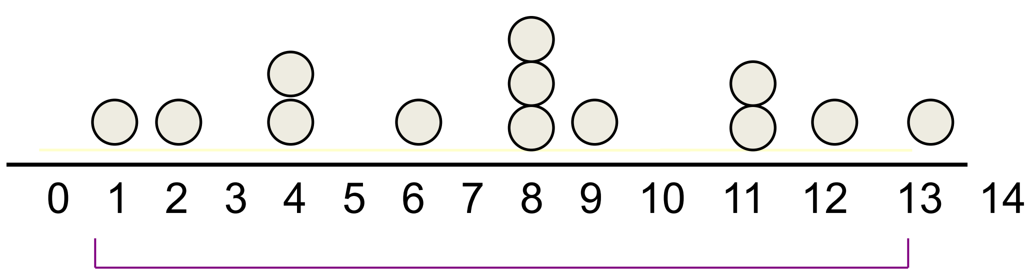
Same centre, different variation

Measures of Dispersion: The Range

- Simplest measure of dispersion
- Difference between the largest and the smallest values:

$$\text{Range} = X_{\text{largest}} - X_{\text{smallest}}$$

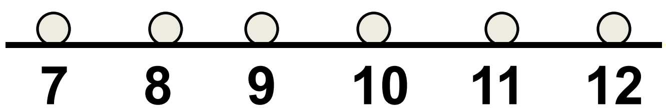
Example:



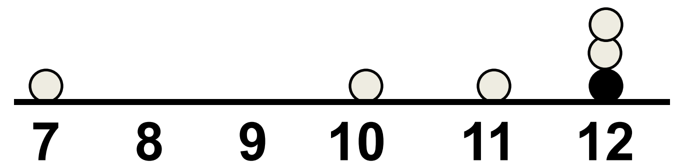
Range = 13 - 1 = 12

Measures of Dispersion: Why The Range Can Be Misleading

- Ignores the way in which data are distributed



Range = 12 - 7 = 5



Range = 12 - 7 = 5

- Sensitive to outliers

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**5**

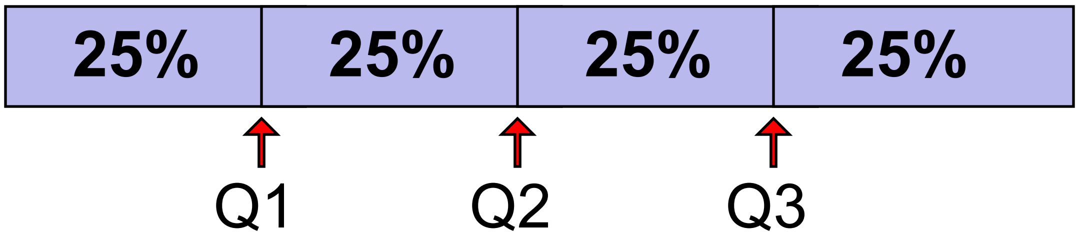
Range = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**120**

Range = 120 - 1 = 119

Quartile Measures

- Quartiles split the ranked data into 4 segments with an equal number of values per segment



- The first quartile, Q_1, is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% of the observations are smaller and 50% are larger)
- Only 25% of the observations are greater than the third quartile

Quartiles

$$Q1 = L1 + \left[\frac{N}{4} - F\right] * \left[\frac{L2-L1}{f}\right]$$

$$N = \sum f_i$$

$$Q2 = L1 + \left[\frac{N}{2} - F\right] * \left[\frac{L2-L1}{f}\right]$$

$$Q3 = L1 + \left[\frac{3N}{4} - F\right] * \left[\frac{L2-L1}{f}\right]$$

QUARTILES

Calculate all the three quartiles for the following data.

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
No. of students	65	45	120	25	90	80	120	60

Marks (Class)	No. of students (frequency)	Cumulative frequency
0–10	65	65
10–20	45	110
20–30	120	230
30–40	25	255
40–50	90	345
50–60	80	425
60–70	120	545
70–80	60	605
Total	**605**	

$\therefore$ **First Quartile (Q_1):**

We know that number of observations below 1st quartile are

$$\frac{N}{4} = \frac{605}{4} = 151.25$$

$\therefore$ Quartile class $(a - b) = 20 - 30$

$\therefore$ Frequency of Quartile class $= 120$

$\therefore$ Cumulative frequency of just previous to Q_1 class $= 110$.

$\therefore$ We have $a = 20$, $b = 30$, $f = 120$, $F = 110$.

$\therefore$ By definition

$$Q_1 = a + \frac{(b - a)}{f}\left(\frac{N}{4} - F\right)$$

$$= 20 + \frac{(30 - 20)}{120}(151.25 - 110)$$

$$= 23.43$$

Third quartile (Q_3):

We know that, the number of observations below third quartile are

$$\frac{3N}{4} = \frac{3\,(605)}{4} = 453.75$$

$\therefore$ Quartile class $(a - b) = 60 - 70$

∴ Frequency of quartile class = 120

∴ Cumulative frequency of just previous to Q_3 class = 425.

∴ We have a = 60, b = 70, f = 120, F = 425

∴ By definition

$$Q_3 = a + \frac{(b-a)}{f}\left(\frac{3N}{4} - F\right)$$

$$\therefore \quad Q_3 = 60 + \frac{(70-60)}{120}(453.75 - 425)$$

$$= 62.39$$

$$Q2 = a + \left[\frac{N}{2} - F\right] * \left[\frac{b-a}{f}\right]$$

$$Q2 = 40 + (302.5 - 255)*10/90$$

$$= 45.28$$

Measures of Dispersion

Mean Deviation about mean , median or mode

$$\text{Mean Deviation} = \frac{\sum | x_i - \bar{x} |}{n}$$

$\bar{x}$ = mean , median or mode

n = Sum of frequency

Example 1

Find the mean deviation about the mean for the following data:

6, 7, 10, 12, 13, 4, 8, 12

Mean of the given data = $\frac{Sum\ of\ all\ terms}{Total\ number\ of\ terms}$

$\bar{x} = \frac{6 + 7 + 10 + 12 + 13 + 4 + 8 + 12}{8}$

$= \frac{72}{8}$

$= 9$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
6	$6 - 9 = -3$	$\lvert -3 \rvert = 3$
7	$7 - 9 = -2$	$\lvert -2 \rvert = 2$
10	$10 - 9 = 1$	$\lvert 1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
13	$13 - 9 = 4$	$\lvert 4 \rvert = 4$
4	$4 - 9 = -5$	$\lvert -5 \rvert = 5$
8	$8 - 9 = -1$	$\lvert -1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
		$\sum_{1}^{8} \lvert x_i - \bar{x} \rvert = 22$

Mean deviation about mean

$= \frac{\sum \lvert x_i - \bar{x} \rvert}{8}$

$= \frac{22}{8}$

= 2.75

PROBLEM

Find mean deviation about mean , median and mode for the following data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	12	8	5	10	7	3	2

Measures of Dispersion

Standard deviation = $\sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}}$

Variance = $(S.D)^2$

$$= \frac{\sum f_i x_i^2}{N} - \left[\frac{\sum f_i x_i}{N} \right]^2$$

Example: 9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4

The mean is:

$$\frac{(9+2+5+4+12+7+8+11+9+3+7+4+12+5+4+10+9+6+9+4)}{20}$$

MEAN = 140/20 = **7**

$(x_i - \bar{x})^2$

$(9 - 7)^2 = (2)^2 =$ **4**

$(2 - 7)^2 = (-5)^2 =$ **25**

$(5 - 7)^2 = (-2)^2 =$ **4**

$(4 - 7)^2 = (-3)^2 =$ **9**

$(12 - 7)^2 = (5)^2 =$ **25**

$(7 - 7)^2 = (0)^2 =$ **0**

$(8 - 7)^2 = (1)^2 =$ **1**

SO ON...

And we get these results:

4, 25, 4, 9, 25, 0, 1, 16, 4, 16, 0, 9, 25, 4, 9,9, 4, 1, 4, 9

$\sum (x_i - \bar{x})^2$

=4+25+4+9+25+0+1+16+4+16+0+9+25+4+9 +9+4+1+4+9 = **178**

VARIANCE $= \frac{\sum (x_i - \bar{x})^2}{n}$ =(1/20) × 178 = **8.9**

S.D =√8.9 = 2.983

PROBLEM

Find S.D and Variance.

Rainfall	70-80	80-90	90-100	100-110	110-120	120-130
No. of Places	12	7	6	21	19	18

Relative Measures of Dispersion

Coefficient of Range = $\frac{M - m}{M + m}$

Coefficient of Quartile deviation = $\frac{Q3 - Q1}{Q3 + Q1}$

Coefficient of Mean Deviation =

$$\frac{\text{M.D about mean , median or mode}}{\text{mean/median/mode}}$$

Coefficient of Variation = $\frac{S.D \times 100}{Mean}$ %

Skewness

It is the *degree of distortion* from the symmetrical bell curve . It measures the lack of symmetry in data distribution.

It differentiates extreme values in one versus the other tail. A symmetrical distribution will have a skewness of 0.

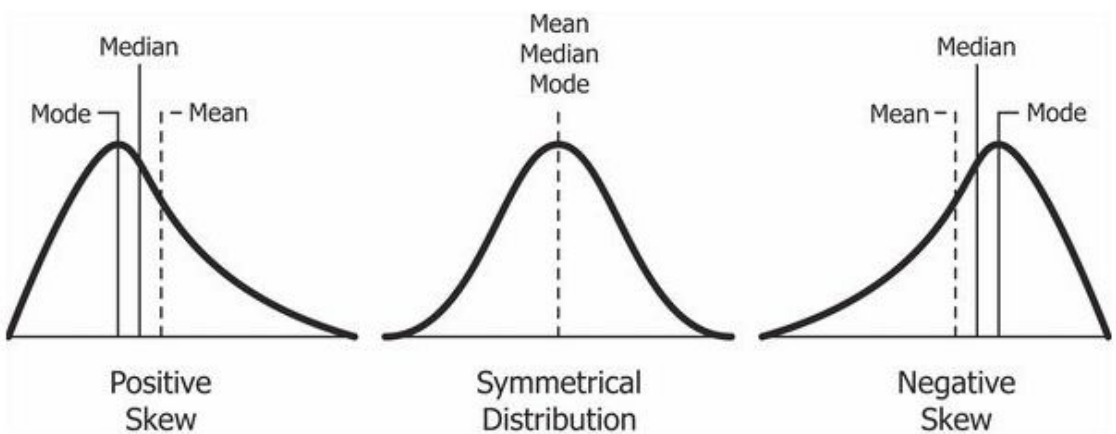
Skewness

There are two types of Skewness: Positive and Negative.

Positive Skewness means when the tail on the right side of the distribution is longer or fatter. The mean and median will be greater than the mode.

Negative Skewness is when the tail of the left side of the distribution is longer or fatter than the tail on the right side. The mean and median will be less than the mode

.



Skewness

- If the skewness is between -0.5 and 0.5, the data are fairly symmetrical.
- If the skewness is between -1 and -0.5(negatively skewed) or between 0.5 and 1(positively skewed), the data are moderately skewed.
- If the skewness is less than -1(negatively skewed) or greater than 1(positively skewed), the data are highly skewed.

Karl Pearson's coefficient of skewness

$$\text{K.P coefficient of skewness} = \frac{\text{Mean - Mode}}{\text{S.D}}$$

Note: if there is more than one mode or no mode in distribution then it is defined by-

$$\text{K.P coefficient of skewness} = \frac{3(\text{Mean - Median})}{\text{S.D}}$$

PROBLEM

Compute K.P coefficient of skewness for the following data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	15	15	23	22	25	10	5	10

Bowley's Coefficient of Skewness

Bowley's Coefficient of Skewness

$$= \frac{(Q_1 - 2Q_2 + Q_3)}{Q_3 - Q_1}$$

Q_1, Q_2, Q_3 are quartiles.

PROBLEM

Compute Bowley's coefficient of skewness for the following data

C.I	30-35	35-40	40-45	45-50	50-55	55-60
Frequency	5	10	30	35	15	5

Skewness in real life

https://youtu.be/XSSRrVMOqlQ

Regression and correlation

Module 2

Syllabus

- Correlation: Karl Pearson's coefficient of correlation, Spearman's rank correlation coefficient.
- Regression: Linear and Non-linear regression (quadratic and cubic), Estimation using linear regression.

Self Learning Topics: Apply correlation and regression on real world data and its graphical representation

Correlation and Regression

Correlation Analysis

Correlation analysis is applied in quantifying the association between two continuous variables, for example, an dependent and independent variable or among two independent variables.

Regression Analysis

Regression analysis refers to assessing the relationship between the outcome variable and one or more variables. The outcome variable is known as the dependent or response variable and the risk elements, and cofounders are known as predictors or independent variables. The dependent variable is shown by “y” and independent variables are shown by “x” in regression analysis.

The sample of a correlation coefficient is estimated in the correlation analysis. It ranges between -1 and +1, denoted by r and quantifies the strength and direction of the linear association among two variables. The correlation among two variables can either be positive, i.e. a higher level of one variable is related to a higher level of another or negative, i.e. a higher level of one variable is related to a lower level of the other.

The sign of the coefficient of correlation shows the direction of the association. The magnitude of the coefficient shows the strength of the association.

For example, a correlation of $r = 0.8$ indicates a positive and strong association among two variables, while a correlation of $r = -0.3$ shows a negative and weak association. A correlation near to zero shows the non-existence of linear association among two continuous variables

Karl Pearson’s Coefficient of Correlation

Pearson's Correlation Coefficient is a linear **correlation coefficient** that returns a value of between -1 and +1. ,-1 means there is a strong negative **correlation** and +1 means that there is a strong positive **correlation**. A 0 means that there is no **correlation** (this is also called zero **correlation**).

e.g the relationship between the age of a consumer and the color of shirt they might purchase or the level of education of a consumer and the delivery mechanism.

The coefficient of correlation rxy between two variables x and y, for the bivariate dataset (xi,yi) where i = 1,2,3…..N; is given by –

$$r_{(x,y)} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

$$cov(x,y) = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{n}$$

cov(x,y): the covariance between x and y and *σx and σy* are the standard deviations of the distributions x and y.

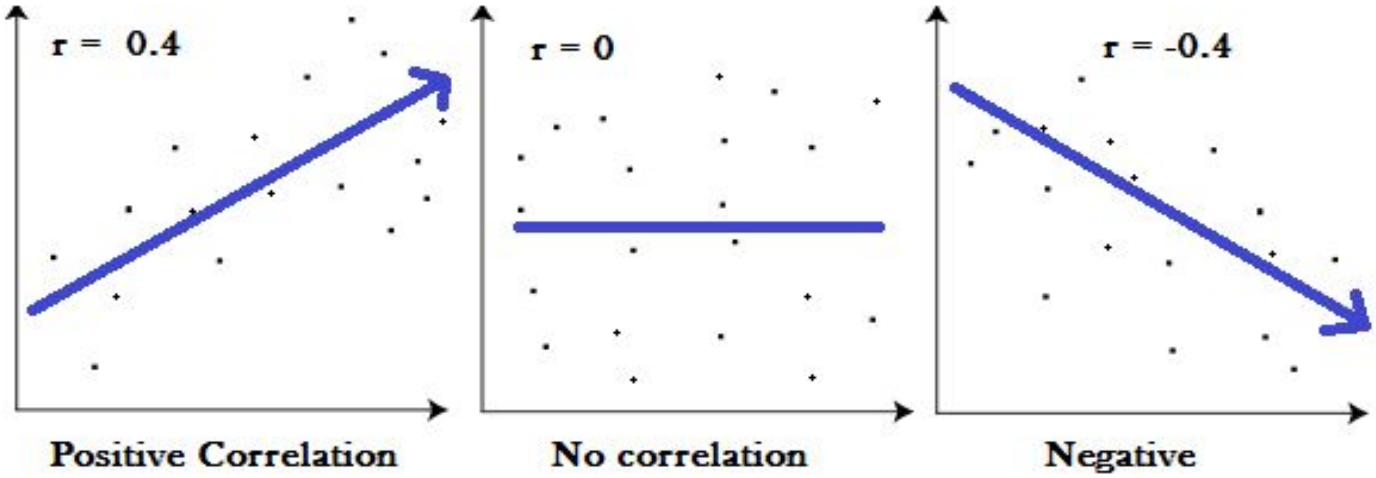
Karl Pearson's Coefficient of Correlation

A Single Formula for Discrete Datasets –

$$r_{xy} = \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{N\Sigma x_i^2 - (\Sigma x_i)^2}\sqrt{N\Sigma y_i^2 - (\Sigma y_i)^2}}$$

r is unit-less. Thus, we may use it to compare association between totally different bivariate distributions as well.

Graphs showing a correlation of -1, 0 and +1



Scatter diagram

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.

Example:

Example question: Find the value of the correlation coefficient from the following table:

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

From our table:

- $\Sigma x = 247$
- $\Sigma y = 486$
- $\Sigma xy = 20{,}485$
- $\Sigma x^2 = 11{,}409$
- $\Sigma y^2 = 40{,}022$
- n is the sample size, in our case = 6

The correlation coefficient =

- $6(20{,}485) - (247 \times 486) / [\sqrt{[[6(11{,}409) - (247^2)] \times [6(40{,}022) - 486^2]]}$
  $= 0.5298$

The range of the correlation coefficient is from -1 to 1. Our result is 0.5298 or 52.98%, which means the variables have a moderate positive correlation.

PROBLEM

Calculate Karl Pearson's coefficient of correlation between x and y for the following data and interpret the result:

(1,6),(2,5),(3,7),(4,9),(5,8),(6,10),(7,11),(8,13),(9,12)

ANSWER: 0.95

Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient, r_s, is the nonparametric version of the Pearson correlation coefficient. Your data must be ordinal, interval or ratio.

Spearman's returns a value from -1 to 1, where:

+1 = a perfect positive correlation between ranks

-1 = a perfect negative correlation between ranks

0 = no correlation between ranks.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

The scores for nine students in physics and math are as follows:

- Physics: 35, 23, 47, 17, 10, 43, 9, 6, 28
- Mathematics: 30, 33, 45, 23, 8, 49, 12, 4, 31

Compute the student's ranks in the two subjects and compute the Spearman rank correlation

Step 1: Find the ranks for each individual subject. Find the ranks by ordering the scores from greatest to smallest; assign the rank 1 to the highest score, 2 to the next highest and so on:.

Physics	Rank	Math	Rank
35	3	30	5
23	5	33	3
47	1	45	2
17	6	23	6
10	7	8	8
43	2	49	1
9	8	12	7
6	9	4	9
28	4	31	4

Step 2: Add a third column, d, to your data. The d is the difference between ranks. For example, the first student's physics rank is 3 and math rank is 5, so the difference is 3 points. In a fourth column, square your d values.

Physics	Rank	Math	Rank	d	d squared
35	3	30	5	2	4
23	5	33	3	2	4
47	1	45	2	1	1
17	6	23	6	0	0
10	7	8	8	1	1
43	2	49	1	1	1
9	8	12	7	1	1
6	9	4	9	0	0
28	4	31	4	0	0

Step 3: Sum (add up) all of your d-squared values. 4 + 4 + 1 + 0 + 1 + 1 + 1 + 0 + 0 = 12. You'll need this for the formula (the Σ d2 is just "the sum of d-squared values").

Step 4: Insert the values into the formula. These ranks are not tied, so use the first formula:

= 1 – (6*12)/(9(81-1))

= 1 – 72/720

= 1-0.1

= 0.9

The Spearman Rank Correlation for this set of data is 0.9.

Spearman Rank Correlation: What to do with Tied Ranks

Tied ranks are where two items in a column have the same rank. They are given the mean rank.

E.g

Xi	Rank(Xi)	Yi	Rank(Yi)	d_i^2
20	8	18	8	0
23	6.5	22	7	0.25
23	6.5	24	6	0.25
25	5	29	5	0
27	3.5	33	4	0.25
27	3.5	36	2	2.25
32	2	36	2	0
45	1	36	2	1
Total				**4**

To calculate ranks

For Xi

Rank 3,4 are tied hence

(3+4)/2= 7/2=3.5

Rank 6,7 are tied hence

(6+7)/2= 13/2=6.5

For Yi

Rank 1,2,3 are tied hence

(1+2+3)/3=6/3=2

Correction in formula

m-> number of times rank repeated

$(m^3-m)/12$ is added to $\sum d_i^2$

$$\rho = 1 - 6\left[\frac{\sum D_i^2 + \frac{1}{12}\left(m_1^3 - m_1\right) + \frac{1}{12}\left(m_2^3 - m_2\right) + \dots}{n\left(n^2 - 1\right)}\right]$$

Spearman Rank Correlation: What to do with Tied Ranks

Rank 3.5 is repeated m=2 times, hence correction $(m^3-m)/12=(2^3-2)/12=0.5$

Rank 6.5 is repeated m=2 times, hence correction $(m^3-m)/12=(2^3-2)/12=0.5$

Rank 2 is repeated m=3 times, hence correction $(m^3-m)/12=(3^3-3)/12=2$

Correction $\sum d_i^2 = 4 + 0.5 + 0.5 + 2 = 7$

Hence ρ is given by -

$$\rho = 1 - 6\left[\frac{\sum D_i^2 + \frac{1}{12}\left(m_1^3 - m_1\right) + \frac{1}{12}\left(m_2^3 - m_2\right) + \ldots}{n\left(n^2 - 1\right)}\right]$$

$$\rho = 1 - \frac{6 \times 7}{8(8^2-1)} = 0.92$$

PROBLEM

1. Two referees in a flower beauty competition rank the 10 types of flowers as follows:

Referee A	1	6	5	10	3	2	4	9	7	8
Referee B	6	4	9	8	1	2	3	10	5	7

2. Compute the rank correlation coefficient for the following data of the marks obtained by 8 students in the Commerce and Mathematics.

Marks in Commerce	15	20	28	12	40	60	20	80
Marks in Mathematics	40	30	50	30	20	10	30	60

Linear Regression

Linear regression attempts to model the relationship between two variables by fitting a **linear** equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

When a **correlation coefficient** shows that data is likely to be able to predict future outcomes and a scatter plot of the data appears to form a straight line, you can use simple linear regression to find a predictive function.

Lines of regression - We measure the deviations of the points vertically and horizontally along y axis and x axis. Thus we have two lines of regression.

a. Lines of regression Y on X
b. Lines of regression X on Y

Least Square Method

It is the mathematical tool useful to obtain lines of regression of X on Y and Y on X.

i) Lines of regression of Y on X

$$Y = a + bX$$

ii) Lines of regression of X on Y

$$X = a + bY$$

Coefficients of regression

i) Coefficient of regression of Y on X

It is the slope of the line of regression of Y on X i.e **b** from the equation y=a +bx, it is given by

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

r = coefficient of correlation

σ_x = standard deviation of x

σ_y = standard deviation of y

$$b_{yx} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

line of regression of Y on X is $(y - \bar{y}) = b_{yx}(x - \bar{x})$

Coefficients of regression

ii) Coefficients of regression of X on Y

It is the slope of the line of regression of X on Y i.e **b** from the equation x=a +by, it is given by

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

r = coefficient of correlation

σ_x = standard deviation of x

σ_y = standard deviation of y

$$b_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(y_i - \bar{y})^2}$$

line of regression of X on Y is $(x - \bar{x}) = b_{xy}(y - \bar{y})$

PROBLEM

1. Find the equation of line of regression of y on x ,line of regression of x on y and estimate the most probable value of y when x=8.

X	3	4	5	6	4	5	6	7
Y	3	5	3	2	3	4	6	6

Solution

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
3	3	-2	4	-1	1	2
4	5	-1	1	1	1	-1
5	3	0	0	-1	1	0
6	2	1	1	-2	4	-2
4	3	-1	1	-1	1	1
5	4	0	0	0	0	0
6	6	1	1	2	4	2
7	6	2	4	2	4	4
Total:40	**32**		**12**		**16**	**6**

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

$= 40/8=5$

$\bar{y} = \sum y / n$

$= 32/8 = 4$

Coefficient of regression of Y on X is

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

$$= 6/12=0.5$$

Line of regression of y on x is $(y - \bar{y})=b_{yx}(x - \bar{x})$

$$y- 4 = 0.5 (x - 5)$$

$$y = 0.5 x + 1.5$$

When x=8

$$y=(0.5)(8) +1.5 =5.5 \simeq 6$$

Line of regression for x on y

$$b_{xy} = 6/16 = 0.325$$

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

$$x = 0.375y +3.5$$

PROBLEM

1. Find the equation of lines of regression and estimate the most probable value of y when x=15.5.

X	10	12	13	16	17	20	25
Y	19	22	24	27	29	33	37

Non- Linear Regression

Non- Linear Regression

Nonlinear regression is a form of **regression** analysis in which observational data are modeled by a function which is a **nonlinear** combination of the **model** parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations.

Nonlinear models are more complicated than linear models to develop because the function is created through a series of approximations (iterations) that may stem from trial-and-error.

- Quadratic regression
- Cubic regression

Quadratic Regression

Quadratic regression is finding the best fit equation for a set of data shaped like a parabola.

The first step in regression is to **make a scatter plot**. If your scatter plot is in a “U” shape, either concave up (like the letter U) or concave down (∩), you’re probably looking at some type of quadratic equation as the best fit for your data. A quadratic doesn’t have to be a full “U” shape; you can have part of a it (say, a quarter or 3/4).

Quadratic regression is an **extension of simple linear regression.** While linear regression can be performed with as few as two points (i.e. enough points to draw a straight line), quadratic regression come with the disadvantage that it requires more data points to be certain your data falls into the “U” shape.

Quadratic regression is a way to model a relationship between two sets of variables. The result is a regression equation that can be used to make predictions about the data. The equation has the form:

$y = ax2 + bx + c$,

where $a \neq 0$.

Quadratic Regression

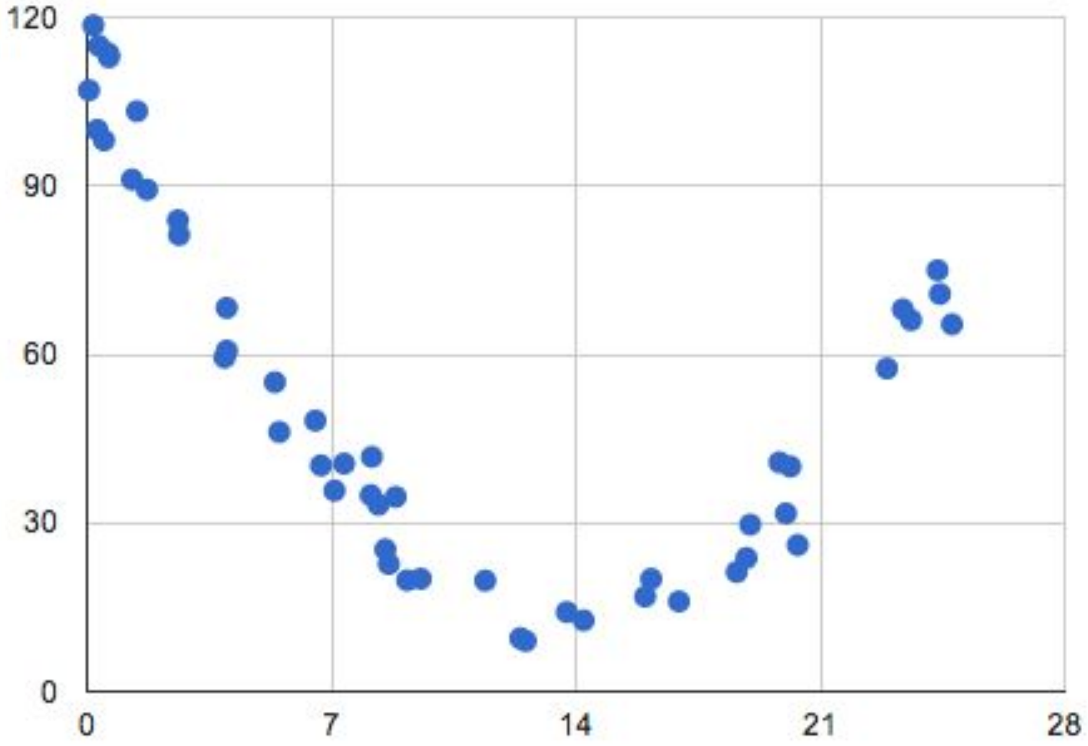
In order to find the quadratic regression by hand, you have to solve the following **system of equations.**This set of equations is sometimes called *normal equations*.

$$a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i$$

$$a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i$$

$$a\sum x_i^2 + b\sum x_i + cn_i = \sum y_i$$

Data points that suggest quadratic regression would be a good fit.



PROBLEM

1. Find the quadratic equation for the following set of data

   x: 1, 3, 5, 7, 9

   y: 32.5, 37.3, 36.4, 32.4, 28.5

SOLUTION

Make a table . Input your x-values in the first column and your y-values in the second column. Add 5 more columns labeled x2, x3, x4 xy, and x2y:

x	y	x^2	x^3	x^4	xy	x^2 * y
1	32.5					
3	37.3					
5	36.4					
7	32.4					
9	28.5					

Calculate each column. For example, the x2 column is simply the squares of the first column; the last column is the third column multiplied by the second column (the y-values)

	x	y	x^2	x^3	x^4	xy	x^2 * y
	1	32.5	1	1	1	32.5	32.5
	3	37.3	9	27	81	111.9	335.7
	5	36.4	25	125	625	182	910
	7	32.4	49	343	2401	226.8	1587.6
	9	28.5	81	729	6561	256.5	2308.5
TOTAL	25	167.1	165	1225	9669	809.7	5174.3

Use the last row (the summations) to fill in the values. All you're doing is transferring the numbers to the normal equation (*n* is the number of items in the set, which is 5 in our QUESTION)

$$a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i$$

$$9669a + 1225b + 165c = 5174.3$$

$$a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i$$

$$1225a + 165b + 25c = 809.7$$

$$a\sum x_i^2 + b\sum x_i + cn_i = \sum y_i$$

$$165a + 25b + 5c = 167.1$$

Solve the system of equations to get the values.

$a = -0.3660714$

$b = 3.015714$

$c = 30.42179$

Insert the values into the quadratic equation (I'm rounding to 3 decimal places).

$y = ax^2 + bx + c$

$y = -0.366x^2 + 3.016x + 30.422$

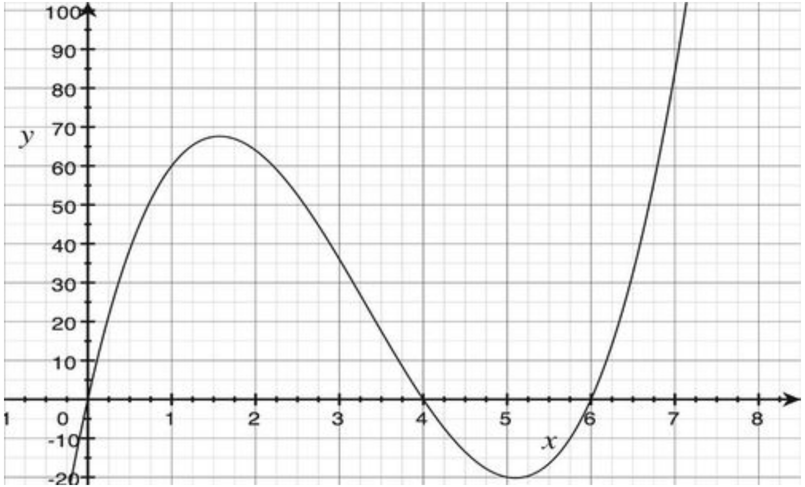
Find the quadratic regression with a Calculator

https://www.youtube.com/watch?v=FWSjjyWrcKs

https://www.youtube.com/watch?v=1b4I-N1vH_0

Cubic Regression

A **Cubic regression model** uses a cubic functions (of the form ax^3+bx^2+cx+d) to model real-world situations. They can be used to model three-dimensional objects to allow you to identify a missing dimension or explore the result of changes to one or more dimensions. A cubic curve has two humps–one facing upward and the other down. The curve goes down, back up, then back down again (or vice-versa).



Cubic Regression

$$\begin{bmatrix} \sum x^6 & \sum x^5 & \sum x^4 & \sum x^3 \\ \sum x^5 & \sum x^4 & \sum x^3 & \sum x^2 \\ \sum x^4 & \sum x^3 & \sum x^2 & \sum x \\ \sum x^3 & \sum x^2 & \sum x & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \sum x^3 y \\ \sum x^2 y \\ \sum xy \\ \sum y \end{bmatrix}$$

Alternatively solve using calculator.

PROBLEM

1. Find the cubic equation for the following set of data

   x: -3, -2, -1, 0, 1, 2 , 3

   y: 3 , -8, -7, 0, 7 ,8, -3

a= -1 , b=0, c=8, d=0

Introduction to probability & conditional probability

Module 3

Syllabus

Introduction to probability, Random experiment, Sample space, Events, Axiomatic Probability, Algebra of events. Conditional Probability, Multiplication theorem of Probability, Independent events, Bayes' Theorem

Introduction To Probability

Randomness and uncertainty exist in our daily lives as well as in every discipline in science, engineering, and technology. Probability theory, is a mathematical framework that allows us to describe and analyze random phenomena in the world around us. By random phenomena, we mean events or experiments whose outcomes we can't predict with certainty.

"**Probability** is the branch of mathematics concerning numerical descriptions of how likely an event is to occur, or how likely it is that a proposition is true. The **probability** of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty."

For **example**, when a coin is tossed in the air, the possible outcomes are Head and Tail.

Random Experiment

An experiment whose all possible outcomes are known in advance but the outcome of any specific performance cannot be predicted with certainty before the completion of the experiment is called a random experiment.

e.g

a) Throwing a dice
b) Tossing a coin
c) Drawing a ball from a box

In particular, a random experiment is a process by which we observe something uncertain. After the experiment, the result of the random experiment is known. An **outcome** is a result of a random experiment. When we repeat a random experiment several times, we call each one of them a **trial**. Thus, a trial is a particular performance of a random experiment.

Sample Space

A **sample space** is the set of all possible outcomes in an experiment.

Example:
Two coins are tossed. Represent the sample space for this experiment by making a list, a table, and a tree diagram.
(H – Head, T – Tail)

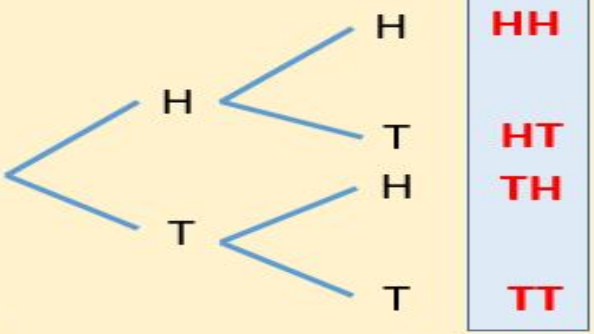
List:

HH HT TH TT

Table:

	H	T
H	**HH**	**HT**
T	**TH**	**TT**

Tree Diagram:



The sample space is {HH, HT, TH, TT}

Denoted with S = { HH, HT, TH, TT }

Events

An event is a subset of the sample space.

The outcomes of a random experiment are called events connected with the experiment. Denoted by A, B, E… etc

For example; ‘head’ and ‘tail’ are the outcomes of the random experiment of throwing a coin and hence are events connected with it.

Algebra of Events

Types of events.

1. Simple Event
2. Compound Event
3. Certain Events / Sure Events
4. Impossible Event
5. Equivalent Events / Identical Events
6. Equally Likely Events
7. Exhaustive Events
8. Favorable Events
9. Mutually Exclusive Events
10. Complementary Event

Algebra of Events

Simple Event

If an event E consists of only one outcome of the experiment then it is called an elementary event.

For example:

In tossing a coin, E = event of getting a head, F = event of getting a tail are both elementary events.

In throwing a die,

A = event of getting 5, is an elementary event while

B = event of getting an even number, is not an elementary event because its favourable outcomes are 2, 4, 6 (three outcomes).

Compound Event

If there are more than one element of the sample space in the set representing an event, then this event is called a compound event.

For example; if we throw a die, having S = {1, 2, 3, 4, 5, 6}, the event of a odd number being shown is given by E = {1, 3, 5}.

Algebra of Events

Certain Events / Sure Events

An event which is sure to occur at every performance of an experiment is called a certain event connected with the experiment.

For example, “Head or Tail’ is a certain event connected with tossing a coin.

Face-1 or face-2, face-3, ……, face-6 is a certain event connected with throwing a die.

Certain Events also known as Sure Event.

Impossible Event

An event which cannot occur at any performance of the experiment is called an impossible event.

Following are such examples ----

(i) ‘Seven’ in case of throwing a dice.

(ii) ‘Sum-13’ in case of throwing a pair of dice.

Algebra of Events

Equivalent Events / Identical Events

Two events are said to be equivalent or identical if one of them implies and implied by other. That is, the occurrence of one event implies the occurrence of the other and vice versa.

For example, “even face” and “face-2” or “face-4” or “face-6” are two identical events.

Equally Likely Events

When there is no reason to expect the happening of one event in preference to the other, then the events are known equally likely events.

For example; when an unbiased coin is tossed the chances of getting a head or a tail are the same.

Algebra of Events

Exhaustive Events

All the possible outcomes of the experiments are known as exhaustive events.

For example; in throwing a die there are 6 exhaustive events in a trial.

Favorable Events

The outcomes which make necessary the happening of an event in a trial are called favorable events.

For example; if two dice are thrown, the number of favorable events of getting a sum 5 is four, i.e., (1, 4), (2, 3), (3, 2) and (4, 1).

Algebra of Events

Mutually Exclusive Events:

If there be no element common between two or more events, i.e., between two or more subsets of the sample space, then these events are called mutually exclusive events.

If E_1 and E_2 are two mutually exclusive events, then $E_1 \cap E_2 = \varnothing$

For example, in connection with throw a die "even face" and "odd face" are mutually exclusive.

Complementary Event:

An event which consists in the negation of another event is called complementary event of other event. In case of throwing a dice, 'even face' and 'odd face' are complementary to each other. "Multiple of 3" ant "Not multiple of 3" are complementary events of each other.

For example: In the throw of a dice if

E = event of getting an odd number

Then $\overline{E}$ = event of not getting an odd number, that is, event of getting an even number.

Axiomatic Probability

Let an experiment result in any one of n mutually exclusive and equally likely events and m of them are favourable to an event A then probability of A denoted as P(A) is given by m/n.

$P(A) = m/n$

$P(A) = n(A)/n(S)$

Throwing a dice and getting number 3.

$P(A) = 1/6$

Axiomatic Probability

Let S be the sample space and A be the event associated with the random experiment. Then the probability of the event A , denoted by P(A) is defined as a real number satisfying following axioms.

1) 0<= P(A) <= 1
2) P(S)= 1
3) If A and B are mutually exclusive events, then P(AUB) = P(A)+P(B)
4) If A1, A2,.........An are set of mutually exclusive events , then

   P(A1 U A2 U A3 U………...UAn)= P(A1) + P(A2) + P(A3) +............P(An)

Axioms of Probability

1) $P(\phi) = 0$ { Probability of impossible event is 0)

2) $P(\overline{A}) = 1 - P(A)$ { Probability of complementary event}

3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ {Addition Theorem of probability}

Problems of Probability

Example 1: A coin is thrown 3 times .what is the probability that at least one head is obtained?

Sol: Sample space = [HHH, HHT, HTH, THH, TTH, THT, HTT, TTT]

Total number of ways = 2 × 2 × 2 = 8. Fav. Cases = 7

P (A) = 7/8

Example 2: Find the probability of getting a numbered card when a card is drawn from the pack of 52 cards.

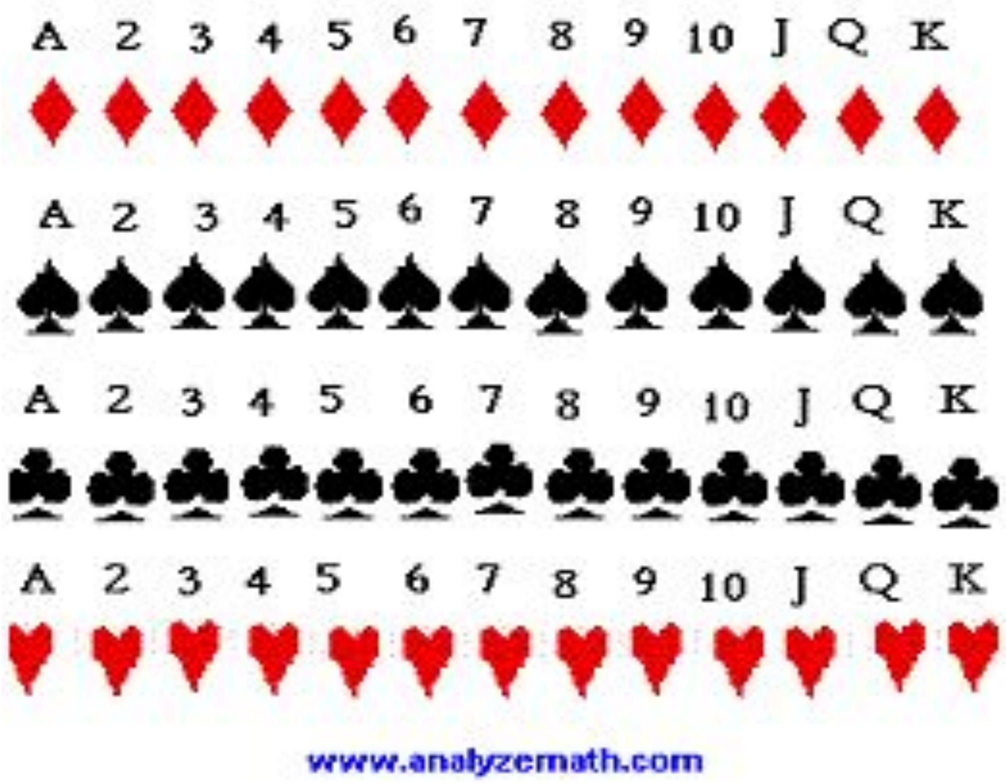
Sol: Total Cards = 52. Numbered Cards = (2, 3, 4, 5, 6, 7, 8, 9, 10) 9 from each suit 4 × 9 = 36

P (E) = 36/52 = 9/13

Example 3: What is the probability of getting a sum of 7 when two dice are thrown?

Sol: Probability math - Total number of ways = 6 × 6 = 36 ways. Favorable cases = (1, 6) (6, 1) (2, 5) (5, 2) (3, 4) (4, 3) --- 6 ways. P (A) = 6/36 = 1/6

DECK of Cards



clubs (♣), **diamonds** (♦), **hearts** (♥), and **spades** (♠)

honor cards = (A, J, Q, K)

face cards = (J,Q,K)

Problems of Probability

1. Two fair coins are tossed . Find the probability of :
   a. Getting both heads
   b. Getting exactly one head
2. If a dice is tossed once , find the probability of numbers 2 to 5
3. 1 card is drawn at random from the pack of 52 cards.

   (i) Find the Probability that it is an honor card.

   (ii) It is a face card.
4. What is the probability of the occurrence of a number that is odd or less than 5 when a fair die is rolled.
5. Two dice are rolled, find the probability that the sum is
   a) equal to 1
   b) equal to 4
   c) less than 13

Probability of Mutually Exclusive Events

Example 1: What is the probability of getting a 2 <u>or</u> a 5 when a die is rolled?

Example 2: Consider the example of finding the probability of selecting a black card or a 6 from a deck of 52 cards.

Conditional Probability

Conditional probability is calculating the probability of an event given that another event has already occured .

The formula for conditional probability P(A|B), read as P(A given B) is

$P(A|B) = P(A \text{ and } B) / P(B)$

Example 1: In a class, 40% of the students study math and science. 60% of the students study math. What is the probability of a student studying science given he/she is already studying math?

Solution

$P(M \text{ and } S) = P(M \cap S) = 0.40$

$P(M) = 0.60$

$P(S|M) = P(M \text{ and } S)/P(M) = P(M \cap S)/P(M) = 0.40/0.60 = 2/3 = 0.67$

Conditional Probability

Example 2

In a group of kids, if one is selected at random the probability that he/she likes oranges is 0.6, the probability that he/she likes oranges AND apples is 0.3. If a kid, who likes oranges, is selected at random, what is the probability that he/she also likes apples?

Solution to Example 2

Let event **O:** kid likes oranges , event **A:** kid likes apples

Given $P(O)=0.6$

Given $P(A \cap O)=0.3$

We are asked to find the conditional probability $P(A|O)$ that the kid likes apples given that he likes oranges.

Conditional Probability Problems

1. Two dies are thrown simultaneously and the sum of the numbers obtained is found to be 7. What is the probability that the number 3 has appeared at least once?
2. There are 100 students in a class of which 38 are boys studying statistics and 13 girls are not studying statistics, if there are 55 girls in all. Find the probability that a student selected at random is not studying statistics given that the student is boy?
3. A card is drawn from a pack of cards. Given that it is a diamond , what is the probability that it is not a face card?

Multiplication Theorem of Probability

If ***A*** and ***B*** are any two events of a sample space such that $P(A) \neq 0$ and $P(B) \neq 0$,

then

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

Example: If $P(A) = 1/5$ $P(B|A) = 1/3$ then what is $P(A \cap B)$?

Solution: $P(A \cap B) = P(A) * P(B|A) = 1/5 * 1/3 = 1/15$

Independent Events

Two or more events are said to be independent if the occurrence of any one does not depend on the occurrence or non-occurrence of the other.

i.e. Two events ***A*** and ***B*** are said to be independent if
$P(A|B) = P(A)$ where $P(B) \neq 0$.
$P(B|A) = P(B)$ where $P(A) \neq 0$.

i.e. Two events ***A*** and ***B*** are said to be independent if
$P(A \cap B) = P(A) * P(B)$.

For n independent events, the multiplication theorem reduces to

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n).$$

Theorems on independent events

1. If A and B are independent events then $P(A \cup B) = 1 - P(\overline{A}).P(\overline{B})$
2. If A and B are independent events then $P(A \cap \overline{B}) = P(A).\ P(\overline{B})$
3. If A and B are independent events then $P(\overline{A} \mid \overline{B}) = P(\overline{A})$
4. If 3 events A,B and C are independent the $P(A \cap B \cap C) = P(A)*P(B)*P(C)$.

Combination

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

Permutation

$$P(n,r) = \frac{n!}{(n-r)!}$$

PROBLEMS

A basket consist of 4 bad and 6 good apples. Two are drawn out of the basket at a time. One of them is found to be good, what is the probability that the other one is also good?

Let A is event that one of the apple drawn is good, and B be the event that the other apple is good.

$\therefore$ $P(A \cap B)$ = Event that both the apples drawn are good.

Since 2 good apples can be selected from 6 apples in ${}^{6}C_{2}$ ways. $\therefore$ $n(A \cap B) = {}^{6}C_{2}$

As 2 apples can be selected from 10 apples in ${}^{10}C_{2}$ ways $\therefore$ $n(S) = {}^{10}C_{2}$

$\therefore P(A \cap B) = n(A \cap B) / n(S) = {}^{6}C_{2} / {}^{10}C_{2} = 6 \times 5/10 \times 9 = \frac{1}{3}$

Knowing that one apple is good , the conditional probability that the other apple is also good is required.

i.e $P(B|A) = P(A \cap B) / P(A) = (\frac{1}{3}) / (6/10) = 5/9$

PROBLEMS

Probability that a student A can solve a problem is $\frac{1}{3}$, B can solve is $\frac{1}{2}$ and C can solve is $\frac{1}{4}$. If all of them try it independently , what is the probability that the problem is solved.

$P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$ and $P(C) = \frac{1}{4}$

$P(\bar{A}) = 2/3$, $P(\bar{B}) = \frac{1}{2}$ and $P(\bar{C}) = \frac{3}{4}$ { A, B, C cannot solve the problem}

$(A \cup B \cup C)$ = Event that the problem is solved.

$(\overline{A \cup B \cup C}) = \bar{A} \cap \bar{B} \cap \bar{C}$ = Event that the problem is not solved

$\therefore$ Probability that the problem is solved = 1- Probability that the problem is not solved

$= 1 - P(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - P(\bar{A}).\,P(\bar{B}).P(\bar{C}) = 1 - \frac{2}{3} \times \frac{1}{2} \times \frac{3}{4} = \frac{3}{4}$

PROBLEMS

1. An engineer applied for job in 2 firms A and B. Probability of being selected in firm A is 0.7 and being rejected in B i s 0.5. Probability that at least one of the firm rejects is 0.6. What is the probability that he will be selected in one of the firms.
2. A box contains tickets marked 1, 2,3, 4…….n. Two tickets are drawn at random without replacement, find the probability that the tickets bear consecutive numbers.
3. A box contains 3 Red , 5 white and 7 blue balls. 3 balls are drawn , what is the probability that one is blue and the other two are red.
4. Find the probability of constructing 2-digit even number using digits 1,2,3,4,5,6,7,8,9. if
   a. Repetition of digits is allowed
   b. Repetition of digits not allowed

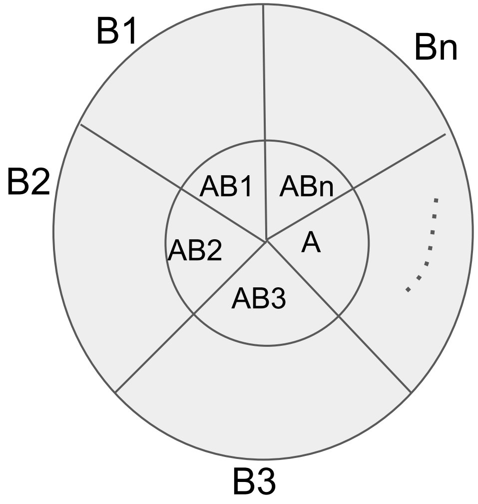
PROBLEMS

5. If the letters of word ‘LOGARITHM’ are arranged at random. Find the probability that the arrangement starts and ends with vowels.
6. Probability that A wins a race is $\frac{3}{8}$ and B wins is $\frac{1}{6}$. If both runs , then find the probability that one of them will win the race, assuming both cannot win together.
7. Probability that a certain drama gets award for its story is 0.23, it will get award for dialogues is 0.15, and for both is 0.07, what is the probability that the drama will get award for:
   a. At least one of the two.
   b. Exactly one of the two.
8. If A and B toss a coin alternatively. Find their probability of winning the toss is A toss first.
9. Two cards are drawn from a pack of cards. Find the probability that both the cards are black.
10. Two dice are thrown together. What is the probability that the number obtained on one of the dice is multiple of number obtained on the other dice?

Total Probability Theorem

If B1, B2 ………...Bn be a set of exhaustive and mutually exclusive events, and A is another event associated with Bi then

$$P(A) = \sum_{i=1}^{n} P(Bi)\, P(A/Bi)$$



The inner circle represents the event A, which have occurred along with and due to B1, B2,B3…….Bn that are exhaustive and mutually exclusive.

A∩B1, A∩B2…. A∩Bn are also exclusive So

$$A = (A \cap B1) + (A \cap B2) + \ldots\ldots\ldots\ldots + (A \cap Bn)$$

$$A = \sum_{i=1}^{n} (A \cap Bi) \quad \Rightarrow \quad P(A) = P\left(\sum_{i=1}^{n} (A \cap Bi)\right)$$

$$P(A) = \sum_{i=1}^{n} P(Bi)\, P(A/Bi) \quad \{ \text{ by conditional probability theorem}\}$$

Bayes' Theorem

If B1, B2 ………...Bn be a set of exhaustive and mutually exclusive events, and A is another event associated with Bi then

$$P(B_i/A) = \frac{P(B_i) \times P(A/B_i)}{\sum_{i=1}^{n} P(B_i) \times P(A/B_i)} \qquad i = 1,2,3,\ldots\ldots\ldots$$

By conditional probability , we have

$P(B_i \cap A) = P(B_i) \times P(A/B_i)$

$P(B_i \cap A) = P(A) \times P(B_i/A)$

$\therefore P(B_i) \times P(A/B_i) = P(A) \times P(B_i/A)$

$$\therefore P(B_i/A) = \frac{P(B_i) \times P(A/B_i)}{P(A)}$$

$$\therefore P(B_i/A) = \frac{P(B_i) \times P(A/B_i)}{\sum_{i=1}^{n} P(B_i) \times P(A/B_i)} \qquad i = 1,2,3,\ldots\ldots\ldots$$

Bayes' Theorem Problems

A bag I contain 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

Solution:

Let E_1 be the event of choosing the bag I, E_2 the event of choosing the bag II, and A be the event of drawing a black ball.

Then, $P(E_1) = P(E_2) = \frac{1}{2}$

Also, $P(A|E_1) = P(\text{drawing a black ball from Bag I}) = \frac{6}{10} = \frac{3}{5}$

$P(A|E_2) = P(\text{drawing a black ball from Bag II}) = \frac{3}{7}$

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{3}{7}} = \frac{7}{12}$$

Bayes' Theorem Problems

A man is known to speak truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Find the probability that the number obtained is actually a four.

Solution:

Let A be the event that the man reports that number four is obtained.

Let E_1 be the event that four is obtained and E_2 be its complementary event.

Then, $P(E_1)$ = Probability that four occurs = $\frac{1}{6}$

$P(E_2)$ = Probability that four does not occurs = $1 - P(E_1) = 1 - \frac{1}{6} = \frac{5}{6}$

Also, $P(A|E_1)$ = Probability that man reports four and it is actually a four = $\frac{2}{3}$

$P(A|E_2)$ = Probability that man reports four and it is not a four = $\frac{1}{3}$

By using Bayes' theorem, probability that number obtained is actually a four,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1) + P(E_2)P(A|E_2)} = \frac{\frac{1}{6} \times \frac{2}{3}}{\frac{1}{6} \times \frac{2}{3} + \frac{5}{6} \times \frac{1}{3}} = \frac{2}{7}$$

Problem

A product manufactured by 3 machines A,B,C

Number of products manufactured are A=2B, B=C

2% of the product by A and B are defective

4% of product by C are defective

1 product is chosen at random from the stock

what is the prob that it is defective?

Bernoulli's trials

Many random experiments that we carry have only two outcomes that are either failure or success. For example, a product can be defective or non-defective, etc. These types of independent trials which have only two possible outcomes are known as Bernoulli trials. For the trials to be categorized as Bernoulli trials it must satisfy these conditions:

- A number of trials should be finite.
- The trials must be independent.
- Each trial should have exactly two outcomes: success or failure.
- The probability of success or failure remains does not change for each trial.

Generalised Bernoulli trials

Let P1 denotes probability that object is of first type, P2 denotes probability that object is of second type……. and Pn denotes that the probability is of nth type… then

$$P[n1, n2, n3 \ldots\ldots\ldots\ldots nk] = \frac{n! \times P1^{n1} \times P2^{n2} \ldots\ldots\ldots Pk^{nk}}{n1!\, n2!\, n3! \ldots\ldots\ldots\ldots nk!}$$

Also

$$P[n1,n2,0] = (P1 + P2)^{n}$$

Problems

Out of 100 jobs received in a call center , 50 are class 1, 30 are class 2 and 20 are class 3. A sample of 30 jobs is taken with replacement.
Find the probability that the sample will contain
i) 10 jobs of each class
ii) No jobs of class 3.

Solution : Let P1, P2 and P3 are the probability that the jobs are of class 1 , 2 and 3 respectively.

P1=0.50 P2=0.30 and P3= 0.20

By Bernoulli's trials

i) $P[10,10,10] = 30! \times P1^{10} \times P2^{10} \times P3^{10} / 10! \times 10! \times 10! = 0.00325$

ii) $P[n1,n2,0] = (P1 + P2)^{n} = (0.50 + 0.30)^{30} = 0.00124$

Problems

An inspection plan calls for inspection of 5 chips, It either accepts, rejects or submit for re-inspection with probabilities 0.7, 0.2 and 0.1 respectively. What is the probability that:

i) all the 5 chips must be re-inspected

ii) none of the chips must be re-inspected

Solution : Let P1, P2 and P3 are the probability for accepting , rejecting and re-inspecting respectively.

P1=0.7 P2=0.2 and P3= 0.1

By Bernoulli's trials

i) $P[0,0,5] = 5! \times P1^0 \times P2^0 \times P3^5 / 0! \times 0! \times 5! = 0.00001$

ii) $P[n1,n2,0] = (P1 + P2)^n = (0.70 + 0.20)^5 = 0.59049$

Problems

Roy plays 12 games with computer , he wins 6 games , computer wins 4 games and 2 games were draw. Roy decides to play 3 more games. Find the probability that:

i) Roy wins all 3 games

ii) 2 games draw

iii)Roy and computer wins alternatively

iv) computer wins at least one game

Random Variable

Module 4

Syllabus

Discrete random variable, Continuous random variable, Two-dimensional random variable, Joint probability distribution, Stochastic independence, Properties of Expectation and Variance, Covariance.

Random Variable

A variable whose exact value is not known before the experiment but probability is known, is called a random variable.

It is of two types:-

i) Discrete random variable - If a random variable takes values at definite places then it is called as discrete random variable. e.g $x = 0, 1, 2$ ...so on

ii) Continuous random variable - If a random variable takes all the values in an interval then it is called as continuous random variable. e.g $0 \leq x \leq 100$

Difference discrete & Continuous random variables

Discrete

1. The probability function of discrete random variable is called probability mass function (pmf).
2. pmf is denoted by P(x)
3. Total probability is given by

$$\sum_{x=-\infty}^{x=\infty} P(x) = 1$$

4. Cumulative probability distribution function is given by (CDF)

$$F(x) = P(x \leq X) = \sum_{i} P(x_i)$$

Continuous

1. The probability function of discrete random variable is called probability density function (pdf).
2. pdf is denoted by f(x)
3. Total probability is given by

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

4. Cumulative probability distribution function is given by (CDF)

$$F(x) = P(x \leq X) = \int_{x=-\infty}^{x} f(x)dx \qquad -\infty < x < \infty$$

Example - discrete random variable

1. Consider an experiment of tossing 3 coins simultaneously. Let random variable x denotes number of heads. Find the
   i) Value of x
   ii) P(x) for all values of x
   iii)CDF

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

i) $x = \{0,1,2,3\}$

ii) $P(x=0) = \frac{1}{8}$, $P(x=1) = \frac{3}{8}$, $P(x=2) = \frac{3}{8}$, $P(x=3) = \frac{1}{8}$

iii) $F(x=0) = \sum P(x \leq 0) = \frac{1}{8}$

$F(x=1) = \sum P(x \leq 1) = P(x=0) + P(x=1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$

$F(x=2) = \sum P(x \leq 2) = P(x=0) + P(x=1) + P(x=2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$

$F(x=3) = \sum P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$

Example

P(x) & F(x) for all values of x

x	0	1	2	3
P(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
F(x)	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{7}{8}$	1

Example

2. A random variable x takes values 1, 2, 3, 4 such that

   $4P(x=1)=2P(x=2)=3P(x=3)=P(x=4)$. Find pmf and CDF of x.

   Solution :

   $4P(x=1)=2P(x=2)=3P(x=3)=P(x=4) = k$

   $P(x=1) = k/4$ $\quad$ $P(x=2)=k/2$ $\quad$ $P(x=3)= k/3$ $\quad$ $P(x=4)=k$

   $\sum P(x) =1$

   $k/4 +k/2 +k/3+k = 1$

   $3k+6k+4k+12k=12 \Rightarrow 25k=12 \Rightarrow k=12/25$

x	1	2	3	4
P(x)	3/25	6/25	4/25	12/25
F(x)	3/25	9/25	13/25	1

Problems

1. Random variable x has following pmf

x	0	1	2	3	4	5	6	7
P(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

   Find k ,pmf and cdf

2. A random variable x is 0 except at points 0,1,2 such that $P(0)= 3c^2$,$P(1)= 4c-10c^2$, $P(2)= 5c-1$. Find

   i) c

   ii) $P(x<2)$ and $P(1<x\leq 2)$

   iii) Find the biggest x such that $F(x) <½$

   iv) Find the smallest x such that $F(x) \geq 1/3$

Basic Formula for Integration

1. $\int x^n = x^{n+1}/n+1 + C$
2. $\int \cos x = \sin x + C$
3. $\int \sin x = -\cos x + C$
4. $\int \sec^2 x = \tan x + C$
5. $\int \mathrm{cosec}^2 x = -\cot x + C$
6. $\int \sec x \tan x = \sec x + C$
7. $\int \mathrm{cosec}\, x \cot x = -\mathrm{cosec}\, x + C$
8. $\int dx/\sqrt{1-x^2} = \sin^{-1} x + C$
9. $\int dx/\sqrt{1-x^2} = -\cos^{-1} x + C$
10. $\int dx/\sqrt{1+x^2} = \tan^{-1} x + C$

10. $\int dx/\sqrt{1+x^2} = \tan^{-1} x + C$

11. $\int dx/\sqrt{1+x^2} = -\cot^{-1} x + C$
12. $\int e^x = e^x + C$
13. $\int a^x = a^x / \log a + C$
14. $\int dx/x\sqrt{x^2 - 1} = \sec^{-1} x + C$
15. $\int dx/x\sqrt{x^2 - 1} = \mathrm{cosec}^{-1} x + C$
16. $\int 1/x = \log |x| + c$
17. $\int \tan x = \log |\sec x| + c$
18. $\int \cot x = \log |\sin x| + c$
19. $\int \sec x = \log |\sec x + \tan x| + c$
20. $\int \mathrm{cosec}\, x = \log |\mathrm{cosec}\, x - \cot x| + c$

Integration by parts

1. $\int f(x)\, g(x)\, dx = f(x) \int g(x)\, dx - \int (f'(x) \int g(x)\, dx)\, dx$

Example - Continuous random variable

1. If x is c.r.v with pdf $f(x) = k(x-x^2)$, $0 \le x \le 1$, then find k.

   By definition

   $$\int_0^1 f(x)dx = 1 \quad \Rightarrow \quad \int_0^1 k(x-x^2)dx = 1$$

   $$k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1 \qquad \left\{ \int x^n = x^{n+1}/n+1 \right\}$$

   $$k\left[\frac{1}{2} - \frac{1}{3} - 0 + 0\right] = 1 \quad \Rightarrow \quad k \cdot \frac{1}{6} = 1 \quad \Rightarrow k=6$$

Example - Continuous random variable

2. A petrol pump is supplied with petrol once a day. If it's daily volume x of sale in thousands is distributed by the function $f(x) = 5(1-x)^4$ such that $0 \le x \le 1$. What must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0.01

Solution : Given a c.r.v denoting amount of petrol sales in a day with pdf

$$f(x) = 5(1-x)^4 \text{ s.t } 0 \le x \le 1$$

Let k be the capacity of tank in thousands, given that supply will be exhausted in a day shall be 0.01 i.e $P(x \ge k)$

$$\int_k^1 f(x)dx = 0.01 \quad \Rightarrow \quad \int_k^1 5(1-x)^4 dx = 0.01$$

$$\left[-5\,\frac{(1-x)^5}{5}\right]_k^1 = 0.01 \quad \Rightarrow \quad 1-k = 0.01^5 \quad \Rightarrow k = 0.6019$$

Therefore the capacity of the tank is 0.6019 x 1000 = 601.9 Litres

Problems

1. A c.r.v has the following pdf

$$f(x) = \begin{cases} k(2-x) & 0 \le x < 2 \\ kx(x-2) & 2 \le x < 3 \\ 0 & \text{otherwise} \end{cases}$$

   Find k and the median of the distribution.

2. CDF of the distribution of a c.r.v is given by

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

   Find pdf.

3. Diameter of an electric cable is assumed to be c.r.v say x defined by function $f(x) = 6x(1-x) \quad 0 \le x \le 1$, check if f(x) is pdf , determine a number b such that $P(x<b)= P(x>b)$.

Solution

1. A c.r.v has the following pdf

$$f(x) = \begin{cases} k(2-x) & 0 \le x < 2 \\ kx(x-2) & 2 \le x < 3 \\ 0 & \text{otherwise} \end{cases}$$

Solution:

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

$$\int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)dx + \int_{2}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$$

$$0 + \int_{0}^{2} k(2-x)dx + \int_{2}^{3} kx(x-2)dx + 0 = 1$$

$$k\left[2x - x^2/2\right]_0^2 + k\left[x^3/3 - 2x^2/2\right]_2^3 = 1$$

$$k\left[2x2 - 2^2/2\right] + k\left[3^3/3 - 2x3^2/2 - 2^3/3 + 2x2^2/2\right] = 1$$

$$2k + 4k/3 = 1 \Rightarrow k = 3/10$$

Median

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

$$k \times \left[2x - x^2/2\right]_0^m = 1/2$$

$$3/10 \times [2m - m^2/2] = ½$$

$$2m - m^2/2 = 5/3$$

$$3m^2 - 12m + 10 = 0$$

$$m = 1.18$$

Relation between pdf and CDF

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

Fundamental Theorem of Calculus

$$P(a < X < b) = F(b) - F(a)$$

Problem 1

Let X be a random variable with PDF given by

$$f_X(x) = \begin{cases} cx^2 & |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the constant c. Find $P(X \ge \frac{1}{2})$.

To find c, we can use $\int_{-\infty}^{\infty} f_X(u)du = 1$:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(u)du \\ &= \int_{-1}^{1} cu^2 du \\ &= \frac{2}{3}c. \end{aligned}$$

Thus, we must have $c = \frac{3}{2}$.

To find $P(X \ge \frac{1}{2})$, we can write

$$P(X \ge \frac{1}{2}) = \frac{3}{2}\int_{\frac{1}{2}}^{1} x^2 dx = \frac{7}{16}.$$

Problem 2 :- *The mileage C in thousands of miles which car owners get with a certain kind of tyre is a random variable having probability density function*

$$f(x) = \frac{1}{20} e^{-x/20}, \text{ for } x > 0$$
$$= 0, \text{ for } x \le 0$$

Find the probabilities that one of these tyres will last

(i) at most 10,000 miles,

(ii) anywhere from 16,000 to 24,000 miles.

(iii) at least 30,000 miles.

Solution. Let *r.v.* X denote the mileage (in '000 miles) with a certain kind of tyre. Then required probability is given by:

(i) $$P(X \le 10) = \int_0^{10} f(x)\,dx = \frac{1}{20}\int_0^{10} e^{-x/20}\,dx$$

$$= \frac{1}{20}\left|\frac{e^{-x/20}}{-1/20}\right|_0^{10} = 1 - e^{-1/2}$$

$$= 1 - 0{\cdot}6065 = 0{\cdot}3935$$

(*ii*) $P(16 \leq X \leq 24) = \dfrac{1}{20} \int_{16}^{24} \exp\left(-\dfrac{x}{20}\right) dx = \left| -e^{-x/20} \right|_{16}^{24}$

$$= e^{-16/20} - e^{-24/20} = e^{-4/5} - e^{-6/5}$$

$$= 0{\cdot}4493 - 0{\cdot}3012 = 0{\cdot}1481$$

(*iii*) $P(X \geq 30) = \int_{30}^{\infty} f(x)\, dx = \dfrac{1}{20} \left| \dfrac{e^{-x/20}}{-1/20} \right|_{30}^{\infty}$

$$= e^{-1\cdot 5} = 0{\cdot}2231$$

Two-dimensional Random Variable

Let S be a sample space associated with an experiment E and X,Y be functions each assigning a real number X(S) and Y(S) respectively to every outcome s ϵ S. Then the pair (X,Y) is called a 2-dimensional random variable.

If X,Y takes finite values then it is called discrete random variable and if it takes all the values in a specified region R in X Y plane it is called a 2 dimensional continuous random variable.

Two-dimensional Discrete Random Variable

Probability Mass Function:

The probability function or joint probability mass function of a discrete two dimensional random variable (x, y) is the function P_{ij} which gives the value $P(X = x_i, Y = y_j)$ for each pair (x_i, y_j) provided the following conditions are satisfied.

(i) $P_{ij} \geq 0$ for all i, j

(ii) $\sum_{j} \sum_{i} P_{ij} = 1$

Properties:

(1) $F(-\infty, \infty) = 1$ and $F(-\infty, y) = 0 = F(x, -\infty)$

(2) For real numbers a_1, a_2, b_1, b_2

$P(a_1 < x \leq b_1, a_2 < y \leq b_2)$

$$= Fxy\,(b_1, b_2) + Fxy\,(a_1, a_2) - Fxy(a_1, b_2) - Fxy(b_1, a_2)$$

(3) If F(x, y) is the density function continuous at (x, y) then $\dfrac{\partial^2 f}{\partial x\, \partial y} = f(x, y)$.

Marginal Probability Distribution:

Let (X, Y) be two dimensional discrete RV.

If X takes values $x_1, x_2, \ldots\ldots, x_n$ and Y take $y_1, y_2 \ldots\ldots, y_m$ then probability distribution of (X, Y) is given by.

X \ Y	y_1	y_2	y_3	……	y_m	Total
x_1	P_{11}	P_{12}	P_{13}	……	P_{1m}	P_1
x_2	P_{21}	P_{22}	P_{23}	……	P_{2m}	P_2
x_3	P_{31}	P_{32}	P_{33}	……	P_{3m}	P_3
x_n	P_{n1}	P_{n2}	P_{n3}	……	P_{nm}	P_n
Total	P_1'	P_2'	P_3'	……	P_m'	1

Marginal probability distribution of X is given by

X	x_1	x_2	x_3	……	x_n	Total
P(X)	P_1	P_2	P_3	……	P_n	1

and

Marginal probability distribution of Y is given by

Y	y_1	y_2	y_3	……	y_m	Total
P(Y)	P_1'	P_2'	P_3'	……	P_m'	1

Examples

Two random variable X and Y have joint probability distribution function given by the following table:

X/Y	1	2	3
1	1/16	1/16	3/16
2	1/8	1/8	1/8
3	3/16	1/16	1/16

Find

i) $P(X<1.5)$ ii) X is even iii) XY is even
iv) marginal distribution of X v) marginal distribution of Y

Example

i) $P(X<1.5) = P(X=1) = 1/16 + 1/16 + 3/16 = 5/16$

ii) $P(X \text{ is even}) = P(X=2) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$

ii) $P(XY \text{ is even}) = P(X=1,Y=2) + P(X=2,Y=1) + P(X=2,Y=2) + P(X=2,Y=3) + P(X=3,Y=2)$

$$= 1/16 + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + 1/16 = 8/16 = \frac{1}{2}$$

iv) marginal distribution of X

X	1	2	3	TOTAL
P(X)	5/16	6/16	5/16	1

v) marginal distribution of Y

Y	1	2	3	TOTAL
P(Y)	6/16	4/16	6/16	1

Example

Joint probability mass function of discrete random variables(X,Y) is given by $P(x,y)= k(2x+3y)$ where x= 0,1,2 and y= 1,2 ,3. Find k.

Joint probability mass function of discrete random variables(X,Y) can be obtained by substituting all the possible values of x, y in joint pmf.

We put the values in tabular format

X/Y	1	2	3
0	3k	6k	9k
1	5k	8k	11k
2	7k	10k	13k

By property of joint pmf

$\sum\sum P(x,y) =1$

$3k+6k+9k+5k+8k+11k+7k+10k+13k =1$
$72k=1$
$k=1/72$

Example

An urn contains 2 white, 4 black and 3 red balls. Three balls are drawn at random without replacement. If X denotes the number of white balls drawn and Y denotes the number of red balls drawn. Find the joint probability distribution.

Solution:

Since there are only 2 white balls and 3 red balls X can take values 0, 1 or 2 and Y can take values 0, 1, 2, or 3.

There are total 9 balls in an urn and we have to draw 3 of them, this can be done is ${}^9C_3$ ways

$\therefore \quad n(S) = {}^9C_3$

then

$$P(X=0, Y=0) = P\text{ (No white or red ball is drawn)}$$

$$= \frac{{}^4C_3}{{}^9C_3} = \frac{1}{21}$$

$$P(X=1, Y=0) = P\text{ (One white and No red ball is shown)}$$

$$= \frac{{}^2C_1 \times {}^4C_2}{{}^9C_3} = \frac{1}{7}$$

$P(X = 2, Y = 0) = P$ (Two white and No red balls is drawn)

$$= \frac{{}^{2}C_{2}\, {}^{4}C_{1}}{{}^{9}C_{3}} = \frac{1}{21}$$

$P(X = 0, Y = 1) = P$ (No white and one red ball is drawn)

$$= \frac{{}^{4}C_{2}\, {}^{3}C_{1}}{{}^{9}C_{3}} = \frac{3}{14}$$

$P(X = 1, Y = 1) = P$ (One white and one red ball is drawn)

$$= \frac{{}^{2}C_{1} \times {}^{3}C_{1} \times {}^{4}C_{1}}{{}^{9}C_{3}} = \frac{2}{7}$$

$P(X = 2, Y = 1) = P$ (Two white and one red ball is drawn)

$$= \frac{{}^{2}C_{2} \times {}^{3}C_{1}}{{}^{9}C_{3}} = \frac{1}{28}$$

$P(X = 0, Y = 2) = P$ (No white and two red balls are drawn)

$$= \frac{{}^{4}C_{1} \times {}^{3}C_{2}}{{}^{9}C_{3}} = \frac{1}{7}$$

$P(X = 1, Y = 2) = P$ (one white and two red balls are drawn)

$$= \frac{{}^{2}C_{1} \times {}^{3}C_{2}}{{}^{9}C_{3}} = \frac{1}{14}$$

$P(X = 2, Y = 2) = P$ (Two white and two red balls are drawn)

$= 0$ (Since it is an impossible event, as we have to draw only 3 balls)

$P(X = 0, Y = 3) + P$(No white and 3 red balls are drawn)

$$= \frac{{}^3C_3}{{}^9C_3} = \frac{1}{84}$$

$P(X = 1, Y = 3) = P$ (One white and 3 red balls are drawn)

$= 0$ (Since it is an impossible event)

$P(X = 2, Y = 3) = P$ (Two white and 3 red balls are drawn)

$= 0$ (since it is an impossible event)

The joint probability distribution of (X, Y) can be represented in a tabular format as given below.

X \ Y	0	1	2	3
0	$\frac{1}{21}$	$\frac{3}{14}$	$\frac{1}{7}$	$\frac{1}{84}$
1	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{14}$	0
2	$\frac{1}{21}$	$\frac{1}{28}$	0	0

Problems

Question 1- *For the following bivariate probability distribution of X and Y, find*

(i) $P(X \leq 1, Y = 2)$, (ii) $P(X \leq 1)$, (iii) $P(Y = 3)$, (iv) $P(Y \leq 3)$ and (v) $P(X < 3, Y \leq 4)$

X \ Y	1	2	3	4	5	6
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

Question **2. The joint probability distribution of a pair of random variables is given by the following table :-**

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.3	0.1

Find :

(i) **The marginal distributions.**

(ii) $P\{(X + Y) < 4\}$.

Question 3 **The following table represents the joint probability distribution of the discrete random variable (X, Y)**

Y \ X	1	2	3
1	1/12	1/6	0
2	0	1/9	1/5
3	1/18	1/4	2/15

Evaluate marginal distribution of X and Y

Two-dimensional Continuous Random Variable

Joint Density Function or Probability Density Function:

The probability density function of a continuous two dimensional random variable (X, y) is the function $F_{xy}(x, y)$ such that

(i) $f_{xy}(x,y) \geq 0$ for all x, y

(ii) $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{xy}(x,y)\, dx\, dy = 1$$

The probability that X takes values in the interval $[a_1, a_2]$ and simultaneously Y takes values in the interval $[b_1, b_2]$ is calculated as.

$$P(a_1 \leq x \leq a_2, b_1 \leq y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{xy}(x,y)\,dx\,dy$$

Joint probability distribution function

$$F_{xy}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{xy}(x,y)\, dx\, dy$$

for two dimensional continuous RV.

The marginal probability density function of X is defined as

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x, y)\, dy$$

and the marginal probability density function of y is defined as

$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x, y)\, dx$$

Examples

A two dimensional random variable (X, Y) has the joint probability density function.

$$f_{xy}(x,y) = \begin{cases} 6e^{-2x-3y} & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

then find

(i) $P(1 < X < 2, 2 < Y < 3)$

(ii) $P(0 < X < 2, Y > 2)$

(iii) Marginal probability density functions of X and Y.

Solution:

The given joint probability density function is

$$f_{xy}(x,y) = \begin{cases} 6\, e^{-2x} \cdot e^{-3y} & x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) $P(1 < X < 2, 2 < Y < 3) = \int_{2}^{3} \int_{1}^{2} f_{xy}(x,y)\,dx\,dy$

$$= \int_{2}^{3} \int_{1}^{2} 6 \cdot e^{-2x} \cdot e^{-3y}\,dx\,dy$$

$$= \int_{2}^{3} 6\,e^{-3y} \cdot \left[\frac{e^{-2x}}{-2}\right]_{1}^{2} dy$$

$$= \int_{2}^{3} 6\,e^{-3y} \left[\frac{e^{-4}}{-2} - \frac{e^{-2}}{-2}\right] dy$$

$$= \int_{2}^{3} -3\,e^{-3y}\,(e^{-4} - e^{-2})\,dy$$

$$= -3\,(e^{-4} - e^{-2}) \cdot \left[\frac{e^{-3y}}{-3}\right]_{2}^{3}$$

$$= (e^{-4} - e^{-2})\,(e^{-9} - e^{-6})$$

(ii) $P(0 < X < 2, Y < 2) = \int_{2}^{\infty} \int_{0}^{2} f_{xy}(x,y)\, dx\, dy$

$$= \int_{2}^{\infty} \int_{0}^{2} 6 \cdot e^{-2x} \cdot e^{-3y}\, dx\, dy$$

$$= \int_{2}^{\infty} 6 \cdot e^{-3y} \left[\frac{e^{-2x}}{-2}\right]_0^2 dy$$

$$= \int_{2}^{\infty} -3\, e^{-3y}\, [e^{-4} - 1]\, dy$$

$$= -3\,(e^{-4} - 1) \left[\frac{e^{-3y}}{-3}\right]_2^{\infty}$$

$$= (e^{-4} - 1)\, [e^{-\infty} - e^{-6}]$$

$$= (e^{-4} - 1)\, [0 - e^{-6}]$$

$$= -(e^{-4} - 1)\, e^{-6}$$

(iii) By definition Marginal probability density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$$

$$= \int_{-\infty}^{0} f_{XY}(x, y)\, dy + \int_{0}^{\infty} f_{XY}(x, y)\, dy$$

$$= 0 + \int_{0}^{\infty} 6 \cdot e^{-2x} \cdot e^{-3y}\, dy$$

$$= 6 \cdot e^{-2x} \cdot \left[\frac{e^{-3y}}{-3}\right]_0^{\infty}$$

$$= -2\, e^{-2x}\, [e^{-\infty} - 1]$$

$$= -2\, e^{-2x}\, (0 - 1)$$

$$= 2\, e^{-2x} \qquad \text{when } x > 0$$

$$= 0 \qquad \text{otherwise}$$

Similarly, Marginal probability density function of Y is given by

$$f_Y(x, y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx$$

$$= \int_{-\infty}^{0} f_{XY}(x, y)\, dx + \int_{0}^{\infty} f_{XY}(x, y)\, dx$$

$$= 0 + \int_{0}^{\infty} 6 \cdot e^{-2x} \cdot e^{-3y}\, dx$$

$$= 6 \cdot e^{-3y} \left[\frac{e^{-2x}}{-2}\right]_0^{\infty}$$

$$= -3\, e^{-3y}\, [e^{-\infty} - e^{0}]$$

$$= -3\, e^{-3y}\, (0 - 1)$$

$$= 3e^{-3y} \qquad \text{when } y > 0$$

$$= 0 \qquad \text{otherwise.}$$

Conditional Probability Distribution

For 2-dimensional discrete random variable

If (X, Y) is two dimensional RV then the probability of event $(X = x_i, Y = y_j)$ such that $Y = y_j$ is given is called conditional probability of X and is given by

$$P(X = x_i \,/\, Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

$$= \frac{P_{ij}}{\sum\limits_{i} P_{ij}}$$

Similarly,

The probability of event $(X = x_i, Y= y_j)$ such that $X = x_i$ is given is called conditional probability of Y and is given by

$$P(X = x_i / Y = y_j) = \frac{P(X = x_i \; Y = y_j)}{P(X = x_i)}$$

$$= \frac{P_{ij}}{\sum\limits_{j} P_{ij}}$$

Conditional Probability Distribution

For 2-dimensional continuous random variable

The conditional density of X, for given Y is denoted by F(X/Y) and is defined as

$$f(x/y) = \frac{f(x, y)}{f_y(y)}$$

where F(x, y) is joint density of (X, Y) and $F_Y(y)$ is marginal density of Y.

Similarly,

The conditional density of Y, for given X is denoted by F(y/x) and is defined as F(y/x) =

$$f(y/x) = \frac{f(x, y)}{f_X(x)}$$

Examples

(X, Y) is a bivariate random variable having following probability mass distribution.

X / Y	0	1	2
0	0	$\frac{1}{3}$	$\frac{2}{3}$
1	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$
2	$\frac{4}{15}$	$\frac{5}{15}$	$\frac{6}{15}$

Find the conditional probability of Y when

(i) $X = 0$ (ii) $X = 1$ (iii) $X = 3$

Solution:

(i) Conditional probability of Y when $X = 0$

By definition

$$P(Y = y_j / X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

Here $x_i = 0$

$$\therefore \quad P(X = x_i) = P(X = 0) = 0 + \frac{1}{3} + \frac{2}{3}$$

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

$$P(Y = 0 / X = 0) = \frac{P(X = 0, Y = 0)}{P(X = 0)} = \frac{0}{1} = 0$$

$$P(Y = 1 / X = 0) = \frac{P(X = 0, Y = 1)}{P(X = 0)} = \frac{\frac{1}{3}}{1} = \frac{1}{3}$$

$$\therefore \quad P(Y = 2 / X = 0) = \frac{P(X = 0, Y = 2)}{P(X = 0)} = \frac{\frac{2}{3}}{1} = \frac{2}{3}$$

$\therefore$ Conditional probability of Y, given $X = 0$

Y	0	1	2	Total
$P(Y = y_j / X = 0)$	0	$\frac{1}{3}$	$\frac{2}{3}$	1

Similarly one can show that

(ii) Conditional probability of Y givn $X = 1$ is

Y	0	1	2	Total
$P(Y = y_j / X = 1)$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$	1

and

(iii) conditional probability of Y when $X = 2$ is

Y	0	1	2	Total
$P(Y = y_j / X = 2)$	$\frac{4}{15}$	$\frac{5}{15}$	$\frac{6}{15}$	1

Examples

The joint pdf of (X, Y) is $f(x, y) = 6e^{-2x-3y}$, $x \geq 0$, $y \geq 0$. Find the marginal density of X and conditional density of Y given X.

Solution:

Given joint pdf of (X, Y) is $f(x, y) = 6e^{-2x-3y}$, $x \geq 0, y \geq 0$.

$\therefore$ By definition, marginal density of X is given by

$$F_x(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_{0}^{\infty} f_{XY}(x, y)\, dy$$

$$= \int_{0}^{\infty} 6\, e^{-2x-3y}\, dy$$

$$= \left[\frac{6e^{-2x-3y}}{-3}\right]_0^{\infty} = -2\,[0 - e^{-2x}]\ \{\because e^{-\infty} = 0, e^0 = 1\}$$

$$= 2e^{-2x}, x \geq 0.$$

$\therefore$ By definition

conditional density of Y, given X is

$$f(y/x) = \frac{f(x, y)}{f_X(x)}$$

$$= \frac{6e^{-2x-3y}}{2e^{-2x}}$$

$$= 3e^{-3y}, y \geq 0$$

Examples

If the joint pdf of (X, Y) is given by $f(x, y) = k, 0 \le x \le y \le 2$. Find k and also the marginal and conditional density functions.

Solution:

Since f(x, y) is joint pdf

$$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1 \qquad \text{where } 0 \le x \le y \le 2$$

$$\text{i.e., } 0 \le x \le y$$

$$\& \; 0 \le y \le 2$$

$$\therefore \int_0^2\int_0^y k\,dy\,dx = 1$$

$$\therefore \int_0^2 k\,(y - 0)\,dy = 1$$

$$\therefore k\left[\frac{y^2}{2}\right]_0^2 = 1$$

$$\therefore k\,[2 - 0] = 1$$

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

Now, by definition Marginal densities of X and Y are respectively given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_{x}^{2} k\, dy = \int_{x}^{2} \frac{1}{2}\, dy = \left[\frac{y}{2}\right]_{x}^{2}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx = \int_{0}^{y} k\, dx = [k\,x]_0^y = \left[\frac{1}{2}x\right]_0^y$$

$$= \frac{y}{2}$$

$\therefore$ Conditional density of X given Y is

$$f(x/y) = \frac{f(x, y)}{f_Y(y)} = \frac{k}{\frac{y}{2}} = \frac{\frac{1}{2}}{\frac{y}{2}} = \frac{1}{y}\quad 0 < x < y.$$

and conditional density of Y given X is

$$f(y/x) = \frac{f(x, y)}{f_X(x)} = \frac{k}{\frac{(2-x)}{2}} = \frac{\frac{1}{2}}{\frac{(2-x)}{2}} = \frac{1}{2-x}; x < y < 2$$

Problems

1. If (X, Y) is two dimensional discrete RV and it's joint probability distribution is as given below then find the conditional distribution of X given Y = 2.

X / Y	0	1	2
–1	$\frac{1}{15}$	$\frac{3}{15}$	$\frac{2}{15}$
0	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
1	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$

2. The joint probability density function of the two dimensional random variable (X, Y) is given by

$$f(x, y) = \begin{cases} \frac{8}{9}xy, & 1 \leq x \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find the marginal densities of X and Y.

(ii) Find the conditional density function of Y given X = x, and the conditional density function of X given Y = y.

Stochastic Independence

An **independent random variable** is a random variable that doesn't have an effect on the other random variables in your experiment. In other words, it doesn't affect the probability of another event happening. For example, let's say you wanted to know the average weight of a bag of sugar so you randomly sample 50 bags from various grocery stores. You wouldn't expect the weight of one bag to affect another, so the variables are independent. The opposite is a **dependent random variable**, which *does* affect probabilities of other random variables.

Knowing the value of X, an independent random variable, doesn't help us to predict a value for Y and vice versa.

Two random variables are independent if **either of the following statements are true**:

1. $P(x|y) = P(x)$, for all values of X and Y.
2. $P(x \cap y) = P(x) * P(y)$, for all values of X and Y.

The two are equivalent.

The first statement, $P(x|y) = P(x)$, for all values of X and Y, is stating "the probability of x, given y, is x." In other words, knowing y should make no difference on the probability, x — it's still going to be just x no matter what the value of y. **Second statement is fundamental counting principle** , which states that if you have two independent events, multiply their probabilities together.

Stochastic Independence. Let us consider two random variables X and Y (of discrete or continuous type) with joint p.d.f. $f_{XY}(x, y)$ and marginal p.d.f.'s $f_X(x)$ and $g_Y(y)$ respectively.

Independent Random variables. *Two r.v.'s X and Y with joint p.d.f. $f_{XY}(x, y)$ and marginal p.d.f.'s $f_X(x)$ and $g_Y(y)$ respectively are said to be stochastically independent if and only if*

$$f_{X,Y}(x, y) = f_X(x)\ g_Y(y)$$

Examples

The joint probability density function of a two-dimensional random variable (X,Y) is given by

$$f(x,y) = 2 \; ; \; 0 < x < 1, \; 0 < y < x$$
$$= 0, \; \text{elsewhere}$$

(i) Find the marginal density functions of X and Y,

(ii) find the conditional density function of Y given $X = x$ and conditional density function of X given $Y = y$, and

(iii) check for independence of X and Y.

Solution. Evidently $f(x,y) \geq 0$ and

$$\int_0^1 \int_0^x 2\, dx\, dy = 2 \int_0^1 x\, dx = 1$$

(i) The marginal p.d.f.'s of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \int_0^x 2\, dy = 2x, \quad 0 < x < 1$$

$$= 0, \text{ elsewhere}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx = \int_y^1 2dx = 2(1-y), \quad 0 < y < 1$$

$$= 0, \text{ elsewhere}$$

(ii) The conditional density function of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}, \quad 0 < x < 1$$

The conditional density function of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{(1-y)}, \quad 0 < y < 1$$

(iii) Since $f_X(x)\, f_Y(y) = 2(2x)(1-y) \neq f_{XY}(x,y)$, X and Y are not independent.

Examples

The joint probability density function of a two-dimensional random variable (X,Y) is given by

$$f(x,y) = 2 \; ; \; 0 < x < 1, \; 0 < y < x$$
$$= 0, \; elsewhere$$

(i) Find the marginal density functions of X and Y,

(ii) find the conditional density function of Y given X = x and conditional density function of X given Y = y, and

(iii) check for independence of X and Y.

Solution. $f(x,y) \geq 0$ and $\int_0^1 \int_0^x 2\,dx\,dy = 2\int_0^1 x\,dx = 1$

(i) The marginal p.d.f.'s of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^x 2\,dy = 2x, \quad 0 < x < 1$$
$$= 0, \; \text{elsewhere}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_y^1 2dx = 2(1-y), \quad 0 < y < 1$$
$$= 0, \; \text{elsewhere}$$

(ii) The conditional density function of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{2}{2x} = \frac{1}{x}, \quad 0 < x < 1$$

The conditional density function of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{2}{2(1-y)} = \frac{1}{(1-y)}, \quad 0 < y < 1$$

(iii) Since $f_X(x)\, f_Y(y) = 2(2x)(1-y) \neq f_{XY}(x,y)$, X and Y are not independent.

Examples

Joint distribution of X and Y is given by

$$f(x, y) = 4xy\, e^{-(x^2+y^2)}; \quad x \geq 0,\ y \geq 0.$$

Test whether X and Y are independent.

For the above joint distribution, find the conditional density of X given Y = y.

Solution. Joint p.d.f. of X and Y is

$$f(x, y) = 4xy \cdot e^{-(x^2+y^2)}; \quad x \geq 0,\ y \geq 0.$$

Marginal density of X is given by

$$f_1(x) = \int_0^\infty f(x, y)\, dy = \int_0^\infty 4xy\, e^{-(x^2+y^2)}\, dy$$

$$= 4x\, e^{-x^2} \int_0^\infty y\, e^{-y^2}\, dy = 4x\, e^{-x^2} \cdot \int_0^\infty e^{-t} \cdot \frac{dt}{2} \quad \left(Put\ y^2 = t\right)$$

$$= 2x \cdot e^{-x^2} \left| -e^{-t} \right|_0^\infty$$

$$\Rightarrow \quad f_1(x) = 2x\, e^{-x^2}; \quad x \geq 0$$

Similarly, the marginal p.d.f. of Y is given by

$$f_2(y) = \int_0^\infty f(x, y)\, dx = 2y\, e^{-y^2}\ ; \quad y \geq 0$$

Since $f(x, y) = f_1(x) \cdot f_2(y)$, X and Y are independently distributed.
The conditional distribution of X for given Y is given by :

$$f(X = x \mid Y = y) = \frac{f(x, y)}{f_2(y)}$$

$$= 2x\, e^{-x^2}; \quad x \geq 0.$$

Problems

1 The random variables X and Y have the joint distribution given by the probability density function :

$$f(x,y)=\begin{cases} 6(1-x-y), & \text{for } x>0,\ y>0,\ x+y<1 \\ 0, & \textit{elsewhere} \end{cases}$$

Find the marginal distributions of X and Y. Hence examine if X and Y are independent.

2 *(a)* Two-dimensional random variable (X, Y) have the joint density

$$f(x,y) = 8xy,\quad 0<x<y<1$$
$$= 0,\quad \text{otherwise}$$

(i) Find the marginal and conditional distributions.

(ii) Are X and Y independent ? Give reasons for your answer.

Expectation

Mathematical expectation, also known as the expected value, is the summation or integration of a possible values from a random variable. It is also known as the product of the probability of an event occurring, denoted P(x), and the value corresponding with the actual observed occurrence of the event. The expected value is a useful property of any random variable.

Usually notated as E(X), the expect value can be computed by the summation overall the distinct values that the random variable can take.

The mathematical expectation will be given by the mathematical formula as,

$E(X)= \Sigma (x_1p_1, x_2p_2, \ldots, x_np_n)$, where x is a random variable with the probability function, f(x), p is the probability of the occurrence, and n is the number of all possible values

For d.r.v $\quad E(X) = \Sigma\, xP(x)$

For c.r.v $\quad E(X) = \int xf(x)\, dx$

Properties of Expectation and Variance

Theorem 1. *If X and Y are random variables then*

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

provided all the expectations exist.

Proof. Let X and Y be continuous r.v.'s with joint p.d.f. $f(x, y)$ and marginal p.d.f 's $f(x)$ and $f(y)$ respectively. Then by definition :

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

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

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

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

$$= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y)\, dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y)\, dx \right] dy$$

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

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

Multiplication Theorem of Expectation

Theorem 2. *If X and Y are independent random variables, then*

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

Proof. Proceeding as in Theorem 6·1, we have :

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

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

[Since X and Y are independent]

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

$$= E(X) E(Y),$$

provided X and Y are independent.

Variance

The variance of random variable X is denoted by V(X) and is defined as second moment about mean denoted by

$$Var\,(X) = E(X - \bar{X})^2$$
$$= E\,(X^2) - 2XE(X) + (E(X))^2$$
$$\mathbf{Var(X) = E(X^2) - [E(X)]^2}$$

Properties of Expectation and Variance

- $\mathbf{E(g(x)) = \Sigma\, g(x)\, P(x) \Rightarrow \int g(x)f(x)dx}$
- $\mathbf{E(g(x) \pm h(x)) = E(g(x)) \pm E(h(x))}$
- $\mathbf{E(ax+b) = aE(x) + b}$
- $\mathbf{V(x \pm y) = V(x) \pm V(y)}$
- $\mathbf{V(ax+b) = a^2\, V(x)}$

Theroem 3 *If X is a random variable and a and b are constants, then*

$$E(aX+b) = a\,E(X)+b$$

provided all the expectations exist.

Proof. By definition, we have

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

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

$$= a\,E(X) + b$$

Theorem 4: if X is a random variable and a and b are two constants then

$$\mathbf{V(ax+b) = a^2\ V(x)}$$

Proof: L.H.S $\quad V(ax+b)$

$= E((ax+b)^2) - [E(ax+b)]^2$

$= E(a^2x^2 + b^2 + 2abx] - [\ aE(x) + b\ ]^2$

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

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

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

$= a^2V(x)$ By definition of V(X)

$$\mathbf{V(ax+b) = a^2\ V(x)}$$

Examples

A random variable X has the following probability mass function

X	−2	−1	0	1	2
P(x)	0.2	0.1	0.2	0.2	0.3

Find (i) $E(X)$ (ii) $E(2x + 1)$ (iii) $E(X^2)$.

Solution:

(i) By definition, Expected value of discrete random variable is

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

$$= (-2)(0.2) + (-1)(0.1) + 0(0.2) + 1(0.2) + 2(0.3)$$

$$= -0.4 - 0.1 + 0 + 0.2 + 0.6$$

$$= 0.3$$

(ii) Let $F(X) = 2X + 1$

Then probability distribution for $F(X) = 2X + 1$ is given as

X	−2	−1	0	1	2
F(X)	−3	−1	1	3	5
P(X)	0.2	0.1	0.2	0.2	0.3

∴ By definition, Expected value of function of discrete random variable is

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

$$= (-3)(0.2) + (-1)(0.1) + (1)(0.2) + 3(0.2) + 5(0.3)$$

$$= -0.6 - 0.1 + 0.2 + 0.6 + 1.5$$

$$= 1.6$$

(iii) Let $g(X) = x^2$

then by given probability distribution we have

X	–2	–1	0	1	2
P(X)	0.2	0.1	0.2	0.2	0.3
g(X)	4	1	0	1	4

∴ Probability distribution for g(X) is

g(X)	0	1	4
P(X)	0.2	0.3	0.5

∴ By definition, expected value of g(X) is

$$
\begin{aligned}
E(g(x)) &= \sum_{i} g(x_i)\, P(x_i) \\
&= 0\,(0.2) + 1(0.3) + 4(0.5) \\
&= 0 + 0.3 + 2 \\
&= 2.3
\end{aligned}
$$

2. Consider a random variable with pdf

$$f(x) = e^{-x}, \qquad x > 0$$

Find E(X)

Solution:

By definition

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

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

$$= \left[x \frac{e^{-x}}{-1} - \int \frac{e^{-x}}{-1}(1)\, dx \right]_0^{\infty}$$

$$= [-x\, e^{-x} - e^{-x}]_0^{\infty}$$

$$= 1$$

Example

If two unbiased dice are thrown, find the expected value of the sum of the number of points on them.

Solution:

Let X be RV representing number of point on a die. then X can take values 1, 2, 3, 4, 5 or 6 with following distribution.

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\text{Expectation } E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \ldots\ldots + 6 \times \frac{1}{6}$$

$$= (1 + 2 + 3 + 4 + 5 + 6) \times \frac{1}{6}$$

$$= \frac{7}{2}$$

Example

The joint probability distribution of a bivariate random variable (X, Y) is given below

Y \ X	1	2	3	Total
1	0.1	0.1	0.2	0.4
2	0.2	0.3	0.1	0.6
Total	0.3	0.4	0.3	1

then find (1) $E(X + Y)$ (2) $E(X) + E(Y)$

(3) $E(XY)$.

Solution:

(1) By definition $E(X + Y) = \sum_j \sum_i P(x = x_i, y = y_j)\,(x_i + y_j)$

$$= (0.1) \times (1 + 1) + (0.1) \times (2 + 1) + (0.2) \times (3 + 1)$$
$$+ (0.2) \times (1 + 2) + (0.3) \times (2 + 2) + (0.1) \times (3 + 2)$$
$$= 0.2 + 0.3 + 0.8 + 0.6 + 1.2 + 0.5$$
$$= 3.6$$

(2) By definition $E(X) = \sum_{i} x_i P(x_i)$

$$= 1 \times (0.3) + 2 \times (0.4) + 3 \times (0.3)$$

$$= 0.3 + 0.8 + 0.9$$

$$= 2$$

similarly $E(Y) = \sum_{j} y_j P(y_j)$

$$= 1 \times (0.4) + 2 \times (0.6)$$

$$= 0.4 + 1.2$$

$$= 1.6$$

(3) $E(XY) = \sum_{j} \sum_{i} P(X = x_i, Y = y_j)\,(x_i\, y_j)$

$$= (0.1) \times (1 \times 1) + (0.1) \times (2 \times 1) + (0.2)\,(3 \times 1)$$

$$+ (0.2)\,(1 \times 2) + (0.3)\,92 \times 2) + (0.1) \times (3 \times 2)$$

$$= 0.1 + 0.2 + 0.6 + 0.4 + 1.2 + 0.6$$

$$= 3.1$$

Problems

(1) A two dimensional random variable (X, Y) has the following joint probability density function

$$f(X, Y) = 2 - \frac{(x+y)}{2} \qquad , 0 \le x \le 2, 0 \le y \le 2$$

$$= 0 \qquad \text{otherwise}$$

then find (i) $E(X+Y)$ (ii) $E(XY)$

(2) **Example 4.11:**

Let X be a random variable for which E(X) = 10 and V(X) = 25. Find the value of 'a' and 'b' such that Y = aX – b has expectation zero and variance 1.

Example The number of hardware failure system in a week of operation has the following probability mass function.

No. of failures	0	1	2	3	4	5	6
Probability	0.18	0.28	0.25	0.18	0.06	0.04	0.01

Find the expectation and variance of the number of failures.

Solution:

(i) By definition

$$\begin{aligned} \text{Expectation } E(x) &= \Sigma x P(x) \\ &= (0 \times 0.18) + (1 \times 0.28) + (2 \times 0.25) + (3 \times 0.18) + (4 \times 0.06) + (5 \times 0.04) \\ &\quad + (6 \times 0.01) \\ &= 1.82 \text{ failure per week} \end{aligned}$$

$$\begin{aligned} \text{Also } E(x^2) \Sigma x^2 P(x) &= (0 \times 0.18) + (1 \times 0.28) + (4 \times 0.25) + (9 \times 0.18) \\ &\quad + (16 \times 0.06) + (25 \times 0.04) + (36 \times 0.01) \\ &= 5.22 \end{aligned}$$

By definition

$$\begin{aligned} \text{Var}(x) &= E(x^2) - (E(x))^2 \\ &= 5.22 - (1.82)^2 \\ &= 1.91 \end{aligned}$$

Example A continuous random variable X has p. d. f.

$f(x) = kx^2 e^{-x}, x \geq 0$. Find k, mean and variance.

Solution:

Given that p.d.f. of x, $f(x) = kx^2 e^{-x}, x \geq 0$

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

$$\therefore \quad \int_{0}^{\infty} k\, x^2 e^{-x}\, dx = 1$$

$$\therefore \quad k\left[-x^2 e^{-x} + \int e^{-x}\, 2x\, dx\right] = 1$$

$$\therefore \quad k\left[-x^2 e^{-x} + \int e^{-x}\, 2x\, dx\right] = 1$$

$$\therefore \quad k\left[-x^2 e^{-x} + 2x\,(-e^{-x}) + \int e^{-x}\, 2\, dx\right] = 1$$

$$\therefore \quad k\left[-x^2 e^{-x} - 2x\, e^{-x} - 2e^{-x}\right]_0^{\infty} = 1$$

$$\therefore \quad k\,[0 - 0 - 0 + 0 + 0 + 2] = 1$$

$$\therefore \quad 2k = 1$$

$$\therefore \quad k = \frac{1}{2}$$

$\therefore$ P. d. f $f(x) = \frac{1}{2}\, x^2 e^{-x}$

$\therefore$ By definition Mean = E(X)

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

$$= \int_{0}^{\infty} \frac{x}{2} \cdot x^2 e^{-x}\, dx$$

$$= \frac{1}{2} \int_{0}^{\infty} x^3 e^{-x}\, dx$$

$$= \frac{1}{2} \left[-x^3 e^{-x} - 3x^2 e^{-x} - 6\, x e^{-x} - 6\, e^{-x} \right]_0^{\infty}$$

$$= \frac{1}{2}\, [6]$$

$$= 3$$

By definition Var (X) = $E(X^2) - (E(X))^2$

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

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

$$= \frac{1}{2}\left[-x^4 e^{-x} - 4x^3 e^{-x} - 12x^2 e^{-x} - 24\, xe^{-x} - 24e^{-x}\right]_0^{\infty}$$

$$= \frac{1}{2}\,[0 + 24]$$

$$= 12$$

$$\therefore \text{Var}(X) = 12 - (3)^2$$

$$= 12 - 9$$

$$= 3$$

Covariance

In probability, covariance is the measure of the joint probability for two random variables. It describes how the two variables change together.

It is denoted as the function cov(X, Y), where X and Y are the two random variables being considered.

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

Two random variable X and Y are said to be uncorrelated if cov(x,y)=0

Properties

1. $cov(ax,by) = ab\ cov(x,y)$
2. $cov(x+a, y+b) = cov(x,y)$
3. $cov(x+y, z) = cov(x,z) + cov(y,z)$

$$\text{Coefficient of correlation } r=cov(x,y)/ \sqrt{V(x)V(y)} = cov(x,y)/\boldsymbol{\sigma}_x\boldsymbol{\sigma}_y$$

Examples

1. Probability density function for two random variables x and y is given by

   $f(x,y) = 2-x-y \qquad 0<x<1,\ 0<y<1$

   i) find variance of x and variance of y

   ii) covariance of x and y

Solution- find marginal density function of x i.e $f(x) = (3/2) - x$, $f(y) =(3/2) -y$

$E(x) = \int xf(x)\,dx = \int x((3/2) - x)dx = [\,3x^2/4 - x^3/3]_0^1 = 5/12$

$E(y) = \int yf(y)\,dy = \int y((3/2) - y)dy = [\,3y^2/4 - y^3/3]_0^1 = 5/12$

$$E(x^2) = \int x^2 f(x)\,dx = \int x^2((3/2) - x)dx = \left[\, 3x^3/6 - x^4/4 \right]_0^1 = 1/4$$

$$E(y^2) = \int y^2 f(y)\,dy = \int y^2((3/2) - y)dy = \left[\, 3y^3/6 - y^4/4 \right]_0^1 = 1/4$$

$$\mathbf{V(X) = E(X^2) - [E(X)]^2 = 1/4 - (5/12)^2 = 11/144}$$

$$\mathbf{V(Y) = E(Y^2) - [E(Y)]^2 = 1/4 - (5/12)^2 = 11/144}$$

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

$$E(XY) = \int_0^1\int_0^1 xy(2-x-y)dx\,dy = \int_0^1 \left[2x/3 - x^2/2\right]dx = \left[2x^2/6 - x^3/6\right]_0^1$$

$$= 2/6 - 1/6 = 1/6$$

$$\mathrm{cov}(X,Y) = 1/6 - 5/12 \times 5/12 = -0.006944$$

Problem

For the following problem check whether X and Y are uncorrelated.

X/Y	-1	0	1
-2	1/16	1/16	1/16
-2	1/8	1/16	1/8
1	1/8	1/16	1/8
2	1/16	1/16	1/16

Queuing Models

Module 5

Syllabus

Queuing Models: Essential features of queuing systems, operating characteristics of queuing system, probability distribution in queuing systems, M/M/1 : N/FCFS.

Queuing Theory

Queuing theory is the mathematical study of the congestion and delays of waiting in line.

Queuing theory examines every component of waiting in line to be served, including the arrival process, service process, number of servers, number of system places, and the number of customers—which might be people, data packets, cars, etc.

As a branch of operations research, queuing theory can help users make informed business decisions on how to build efficient and cost-effective workflow systems. Real-life applications of queuing theory cover a wide range of applications, such as how to provide faster customer service, improve traffic flow, efficiently ship orders from a warehouse, and design of telecommunications systems, from data networks to call centers.

Introduction

Waiting line problems arise either because

(*i*) there is too much demand on the facilities so that we say that there is an excess of waiting time or inadequate number of service facilities.

(*ii*) there is too less demand, in which case there is too much idle facility time or too many facilities.

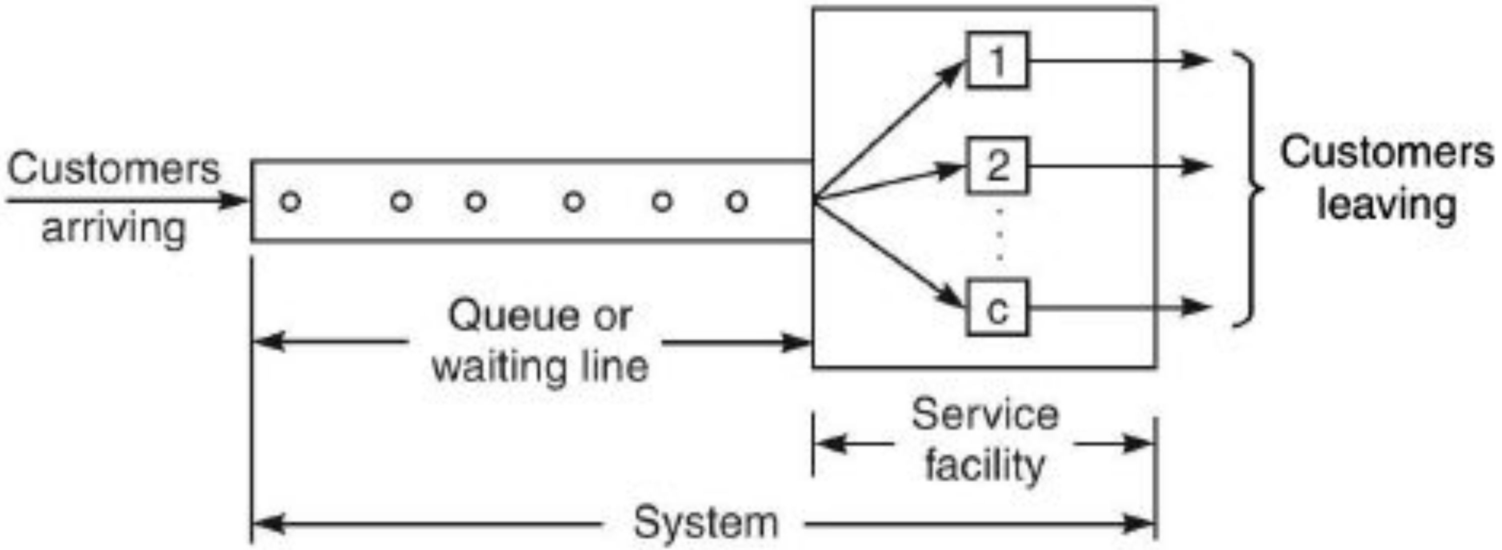
In either case, the problem is to either *schedule arrivals* or *provide proper number of facilities or both* so as to obtain an optimum balance between the costs associated with waiting time and idle time.

Contd..

Units arrive, at regular or irregular intervals of time, at a given point called the service centre. All these units are called *entries* or *arrivals of customers*.

One or more *service channels* or *service stations* or *service facilities* (ticket windows, are assembled at the service centre.

Queuing System



Contd..

1. **Customer:** The arriving unit that requires some service to be performed. As already described, the customers may be persons, machines, vehicles, parts, etc.
2. **Queue (Waiting line):** The number of customers waiting to be serviced. The queue does not include the customer(s) currently being serviced.
3. **Service Channel:** The process or facility which is performing the services to the customer. This may be single or multi-channel. The number of service channels is denoted by the symbol c.

Elements of queuing system

A queuing system is specified completely by seven main elements:

1. Input or arrival (inter-arrival) distribution
2. Output or departure (service) distribution
3. Service channels
4. Service discipline
5. Maximum number of customers allowed in the system
6. Calling source or population
7. Customer's behaviour.

Customer’s Behaviour

- **Balking** - customers deciding not to join the **queue** if it is too long)
- **Reneging** - customers leave the **queue** if they have waited too long for service
- **Jockeying** -customers switch between queues if they think they will get served faster by so doing

Operating characteristics

Analysis of a queuing system involves a study of its different operating characteristics. Some of them are

1. *Queue length* (L_q) – the average number of customers in the queue waiting to get service. This excludes the customer(s) being served.
2. *System length* (L_s) – the average number of customers in the system including those waiting as well as those being served.
3. *Waiting time in the queue* (W_q) – the average time for which a customer has to wait in the queue to get service.
4. *Total time in the system* (W_s) – the average total time spent by a customer in the system from the moment he arrives till he leaves the system. It is taken to be the waiting time plus the service time.
5. *Utilization factor* (ρ) – it is the proportion of time a server actually spends with the customers. It is also called *traffic intensity*.

Transient and Steady State of the system

- **Transient State:**
  - the operating characteristics vary with time
  - Early stages of operation-transient state

- **Steady State**
  - Behaviour becomes independent of its initial conditions and of the elapsed time
  - Long-run behaviour

- **Avg arrival rate < Avg service rate** and constant then eventually settles to **steady state**
- If the rates are **not constant**, system will not reach a steady state and could remain **unstable**
- **Avg arrival rate > Avg service rate**, system cannot attain a steady state (queue length increases steadily and may reach to infinity……**explosive state of system**

Arrival time

The Poisson distribution involves the probability
of occurrence of an arrival. Poisson assumption is quite restrictive in some cases. It assumes that arrivals are random and independent of all other operating conditions. The mean arrival rate (*i.e.*, the number of arrivals per unit of time) λ is assumed to be constant over time and is independent of the number of units already serviced, queue length or any other random property of the queue.

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for x} = 0,1,2 \ldots . \infty$$

Since the mean arrival rate is constant over time, it follows that the probability of an arrival between time t and $t + dt$ is $\lambda \, . \, dt$.

Thus probability of an arrival in time $dt = \lambda \, dt$. ...(10.1)

The following characteristics of Poisson distribution are written here without proof :

$$\text{Probability of } n \text{ arrivals in time } t = \frac{(\lambda t)^n \, . \, e^{-\lambda t^*}}{n!}, \; n = 0, 1, 2, \ldots\ldots, \qquad \ldots(10.2)$$

Probability density function of inter-arrival time (time interval between two consecutive arrivals)

$$= \lambda \, . \, e^{-\lambda t}. \qquad \ldots(10.3)$$

Finally, Poisson distribution assumes that the time period dt is very small so that $(dt)^2$, $(dt)^3$, etc. $\rightarrow 0$ and can be ignored.

Service time

Service time is the time required for completion of a service *i.e.*, it is the time interval between beginning of a service and its completion. The mean service rate is the number of customers served per unit of time (assuming the service to be continuous throughout the entire time unit), while the average service time $1/\mu$ is the time required to serve one customer. The most common type of distribution used for service times is exponential distribution. It involves the probability of completion of a service. It should be noted that Poisson distribution cannot be applied to servicing because of the possibility of the service facility remaining idle for some time. Poisson distribution assumes fixed time interval of continuous servicing, which can never be assured in all services.

Exponential Distribution Formula

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

$f(x;\lambda)$ = probability density function

λ = rate parameter

x = random variable

Mean service rate μ is also assumed to be constant over time and independent of number of units already serviced, queue length or any other random property of the system. Thus probability that a service is completed between t and $t + dt$, provided that the service is continuous

$$= \mu dt.$$

Under the condition of continuous service, the following characteristics of exponential distribution are written, without proof :

$$\text{Probability of } n \text{ complete services in time } t = \frac{(\mu t)^n \cdot e^{-\mu t}}{n!}. \qquad \text{...(10.4)}$$

Probability density function (*p.d.f*) of interservice time, *i.e.*, time between two consecutive services $= \mu \cdot e^{-\mu t}$. ...(10.5)

Probability that a customer shall be serviced in more than time $t = e^{-\mu t}$. ...(10.6)

Contd…

The symbols e and f represent a finite (N) or infinite (∞) number of customers in the system and calling source respectively. For instance, $(M/E_k/1) : (FCFS/N/\infty)$ represents Poisson arrival (exponential interarrival), Erlangian departure, single server, 'first come, first served' discipline, maximum allowable customers N in the system and infinite population model.

Notations:

n = number of customers in the system (waiting line + service facility) at time t.

λ = mean arrival rate (number of arrivals per unit of time).

μ = mean service rate per busy server (number of customers served per unit of time).

L_q = expected (average) number of customers in the queue.

L_s = expected number of customers in the system (waiting + being served).

W_q = expected waiting time per customer in the queue (expected time a customer keeps waiting in line).

W_s = expected time a customer spends in the system. (in waiting + being served)

L_n = expected number of customers waiting in line excluding those *times when the line is empty i.e.*, expected number in *non-empty queue* (expected number of customers in a *queue that is formed from time to time)*.

W_n = expected time a customer waits in line if *he has to wait at all i.e.*, expected time in the queue for *non-empty queue*.

p_n = steady state probability of exactly n customers in the system.

Model I. Single-Channel Poisson Arrivals with Exponential Service, Infinite Population Model [(M/M/I) : (FCFS/∞/∞)]

1. *Expected number of units in the system (waiting + being served)*, L_s is obtained by using the definition of an expected value:

$$E(x) = \sum_{i=0}^{i=\infty} x_i \, p_i$$

$$\therefore \qquad L_s = \sum_{n=0}^{n=\infty} np_n$$

$$\therefore \qquad L_s = \left(1 - \frac{\lambda}{\mu}\right)\left[\frac{\lambda/\mu}{(1-\lambda/\mu)^2}\right] = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}. \qquad ...(10.14)$$

2. *Expected number of units in the queue,* L_q = Expected number of units in the system – Expected number in service (single server).

$$\therefore \qquad L_q = L_s - \frac{\lambda}{\mu} = \frac{\lambda}{\mu-\lambda} - \frac{\lambda}{\mu} = \lambda\left[\frac{\mu-\mu+\lambda}{\mu(\mu-\lambda)}\right] = \frac{\lambda}{\mu} \cdot \frac{\lambda}{\mu-\lambda}. \qquad ...(10.15)$$

Note that the expected number in service is 1 times the probability that the service channel is busy *i.e.*, $1 \cdot \frac{\lambda}{\mu}$.

3. *Expected time per unit in the system (expected time a unit spends in the system)*,

$$W_s = \frac{\text{Expected number of units in the system}}{\text{Arrival rate}} = \frac{L_s}{\lambda} = \frac{\lambda}{(\mu - \lambda) \cdot \lambda} = \frac{1}{\mu - \lambda}. \quad \text{...(}$$

4. *Expected waiting time per unit in the queue, W_q* = Expected time in system – time in service.

$$\therefore \qquad W_q = W_s - \frac{1}{\mu} = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu} \cdot \frac{1}{\mu - \lambda}. \qquad \text{...(10.17)}$$

∴ Variance of queue length

$$= \frac{\lambda/\mu\,(1+\lambda/\mu)}{(1-\lambda/\mu)^2} - \frac{\frac{\lambda^2}{\mu^2}}{\left(1-\frac{\lambda}{\mu}\right)^2} = \frac{\frac{\lambda}{\mu}}{\left(1-\frac{\lambda}{\mu}\right)^2}. \qquad \text{...(10.18)}$$

6. *Average length of non-empty queue (length of queue that is formed from time to time),* L_n :
For a non-empty queue, the number of units in the system should be at least 2 (one in service and the other in the queue). Probability of a non-empty queue

$$= \sum_{n=0}^{\infty} p_n - (p_0 + p_1) = 1 - \left(p_0 + \frac{\lambda}{\mu}\,p_0\right) = 1 - p_0\left(1+\frac{\lambda}{\mu}\right)$$

$$= 1 - \left(1-\frac{\lambda}{\mu}\right)\left(1+\frac{\lambda}{\mu}\right) = \left(\frac{\lambda}{\mu}\right)^2.$$

average length of non-empty queue,

$$L_n = \frac{\text{Average length of queue}}{\text{Probability of non-empty queue}} = \frac{\frac{\lambda}{\mu} \cdot \frac{\lambda}{\mu - \lambda}}{\left(\frac{\lambda}{\mu}\right)^2} = \frac{\mu}{\mu - \lambda}.$$

7. *Average waiting time in non-empty queue (average waiting time of an arrival who waits), or expected waiting time per busy period,*

$$W_n = \frac{1}{\mu - \lambda}. \qquad \text{...(10.20)}$$

8. *Probability of queue being greater than or equal to k,* $= \left(\frac{\lambda}{\mu}\right)^k$.

9. *Probability of queue being greater than k,*

$$p\,(>k) = \left(\frac{\lambda}{\mu}\right)^{k+1}.$$

10. *Probability that the queue is non-empty,*

$$p(n>1) = 1 - p_0 - p_1 = 1 - \left(1 - \frac{\lambda}{\mu}\right) - \frac{\lambda}{\mu}\left(1 - \frac{\lambda}{\mu}\right) = \left(\frac{\lambda}{\mu}\right)^2.$$

11. *Probability density function of waiting time (excluding service) distribution*

$$= \begin{cases} \dfrac{\lambda}{\mu}(\mu - \lambda)\,.\,e^{-(\mu - \lambda)t}, & t > 0 \\[2ex] \dfrac{\lambda}{\mu}(\mu - \lambda), & t = 0. \end{cases}$$

12. *Probability density function of (waiting + service) time distribution*

$$= (\mu - \lambda)\,.\,e^{-(\mu - \lambda)t}.$$

Ex:1

A self-service store employs one cashier at its counter. Nine customers arrive on an average every 5 minutes while the cashier can serve 10 customers in 5 minutes. Assuming Poisson distribution for arrival rate and exponential distribution for service time, find

1. *Average number of customers in the system.*
2. *Average number of customers in the queue or average queue length.*
3. *Average time a customer spends in the system.*
4. *Average time a customer waits before being served.*

Arrival rate $\lambda = 9/5 = 1.8$ customers/minute,

service rate $\mu = 10/5 = 2$ customers/minute.

1. Average number of customers in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{1.8}{2 - 1.8} = 9.$$

2. Average number of customers in the queue,

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu} \cdot \frac{\lambda}{(\mu - \lambda)} = \frac{1.8}{2} \times \frac{1.8}{2 - 1.8} = 8.1.$$

3. Average time a customer spends in the system,

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{2 - 1.8} = 5 \text{ minutes.}$$

4. Average time a costumer waits in the queue,

$$W_q = \frac{\lambda}{\mu}\left(\frac{1}{\mu - \lambda}\right) = \frac{1.8}{2}\left(\frac{1}{2 - 1.8}\right) = 4.5 \text{ minutes.}$$

Ex 2:

A person repairing radios finds that the time spent on the radio sets has exponential distribution with mean 20 minutes. If the radios are repaired in the order in which they come in and their arrival is approximately Poisson with an average rate of 15 for 8-hour day, what is the repairman's expected idle time each day ? How many jobs are ahead of the average set just brought in ?

Solution

$$\text{Arrival rate } \lambda = \frac{15}{8 \times 60} = \frac{1}{32} \text{ units/minute,}$$

$$\text{service rate } \mu = \frac{1}{20} \text{ units/minute.}$$

Number of jobs ahead of the set brought in = Average number of jobs in the system,

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{1/32}{1/20 - 1/32} = \frac{5}{3}.$$

Number of hours for which the repairman remains busy in an 8-hour day

$$= 8\ \frac{\lambda}{\mu} = 8 \times \frac{1/32}{1/20} = 8 \times \frac{20}{32} = 5 \text{ hours.}$$

$\therefore$ Time for which repairman remains idle in an 8-hour day

$$= 8 - 5 = 3 \text{ hours.}$$

Model (M/M/1:FCFS/N/∞) Finite length Model

This model differs from model I in that the maximum number of customers in the system is limited to N and hence the difference equations of model I are valid only so long as $n < N$.

Thus in this model, $\lambda_n = \lambda,\ \mu_n = \mu$ for $n < N$,

$\lambda_n = 0,\ \mu_n = \mu$ for $n \geq N$,

because when the number of customers in the system becomes N, no new arrivals can be accommodated.

Characteristics:

1. Average number of customers in the system,

$$L_s = \sum_{n=0}^{N} np_n = \sum_{n=0}^{N} n \cdot \left(\frac{1-\rho}{1-\rho^{N+1}}\right) \rho^n$$

$$= \frac{1-\rho}{1-\rho^{N+1}} \cdot \sum_{n=0}^{N} n\rho^n$$

$$= \frac{1-\rho}{1-\rho^{N+1}} \cdot [0 + \rho + 2\rho^2 + 3\rho^3 + \ldots + N\rho^N].$$

$$L_s = \frac{\rho[1-(1+N)\rho^N + N\rho^{N+1}]}{(1-\rho)(1-\rho^{N+1})}.$$

2. Average number of customers in the queue,

$$
\begin{aligned}
L_q &= \sum_{n=1}^{N} (n-1)\, p_n = \sum_{n=1}^{N} np_n - \sum_{n=1}^{N} p_n \\
&= \sum_{n=1}^{N} (0 + np_n) - \left[\sum_{n=0}^{N} p_n - p_0 \right] \\
&= \sum_{n=0}^{N} np_n - \sum_{n=0}^{N} p_n + p_0 = L_s - 1 + p_0 \\
&= \frac{\rho\,[1 - (1+N)\rho^N + N\rho^{N+1}]}{(1-\rho)\,(1-\rho^{N+1})} - 1 + \frac{1-\rho}{1-\rho^{N+1}} \\
&= \frac{\{\rho - (1+N)\rho^{N+1} + N\rho^{N+2}\} - \{1 - \rho - \rho^{N+1} + \rho^{N+2}\} + \{1 - 2\rho + \rho^2\}}{(1-\rho)\,(1-\rho^{N+1})} \\
&= \frac{\rho^2 - N\,\rho^{N+1} + (N-1)\rho^{N+2}}{(1-\rho)\,(1-\rho^{N+1})} \\
&= \frac{1 - N\,\rho^{N-1} + (N-1)\rho^N}{(1-\rho)\,(1-\rho^{N+1})} \,.\, \rho^2.
\end{aligned}
$$

3. Average time a customer spends in the system,

$$W_s = \frac{L_s}{\lambda'}, \qquad \text{where } \lambda' = \lambda\,(1 - p_N).$$

4. Average waiting time in the queue,

$$W_q = \frac{L_q}{\lambda'}.$$

Ex 1:

Trains arrive at the yard every 15 minutes and the service time is 33 minutes. If the line capacity of the yard is limited to 4 trains, find

(i) the probability that the yard is empty,

(ii) the average number of trains in the system.

Solution

$$\lambda = 1/15 \text{ per minute,}$$
$$\mu = 1/33 \text{ per minute.}$$
$$\therefore \qquad \rho = \lambda/\mu = 33/15 = 2.2,$$
$$N = 4.$$

(*i*)
$$p_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{\rho - 1}{\rho^{N+1} - 1} = \frac{2.2-1}{2.2^5 - 1} = \frac{1.2}{51.5 - 1} = 0.0237.$$

(*ii*) Average number of trains in the system,

$$L_s = \sum_{n=0}^{N} np_n = 0 + p_1 + 2p_2 + 3p_3 + 4p_4$$
$$= p_0(\rho + 2\rho^2 + 3\rho^3 + 4\rho^4) \qquad (\because p_n = \rho^n \cdot p_0.)$$
$$= 0.0237[2.2 + 2 \times 2.2^2 + 3 \times 2.2^3 + 4 \times 2.2^4]$$
$$= 0.0237[2.2 + 9.68 + 31.94 + 93.70] = 3.26.$$

Ex 2:

If for a period of 2 hours in a day (8 A.M. to 10 A.M.) trains arrive at the yard every 20 minutes but the service time is 36 minutes, calculate for this period

(a) the probability that the yard is empty,

(b) the average number of trains at the yard.

Line capacity of the yard is limited to 4 trains only.

Solution

Here, $\lambda = \dfrac{60}{20} = 3$ trains/hour, $\mu = \dfrac{60}{36} = \dfrac{5}{3}$ trains/hour.

$\therefore$ $$\rho = \frac{\lambda}{\mu} = \frac{3 \times 3}{5} = 1.8.$$

(*a*) $$p_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{\rho - 1}{\rho^{N+1} - 1} = \frac{1.8 - 1}{1.8^5 - 1} = \frac{0.8}{18.9 - 1} = 0.045.$$

$\therefore$ $$L_s = p_0 \sum_{n=0}^{4} n \cdot p_n = p_0\, [0 + \rho + 2\rho^2 + 3\rho^3 + 4\rho^4] \; (\because p_n = \rho^n . p_0 .)$$

$$= 0.045 \times 1.8\, [1 + 2 \times 1.8 + 3 \times 1.8^2 + 4 \times 1.8^3]$$
$$= 0.081\, [1 + 3.6 + 9.72 + 23.328] = 3.05 \text{ trains.}$$

Ex 3:

At a railway station only one train is handled at a time. The railway yard is sufficient only for two trains to wait while the other is given signal to leave the station. Trains arrive at the station at an average rate of 6 per hour and the railway station can handle them on an average of 12 per hour. Assuming Poisson arrivals and exponential service distribution, find the steady state probabilities for the various number of trains in the system. Also find the average number of trains at the railway station and the average waiting time of a new train coming into the yard.

Solution

Here, $\lambda = 6$ trains/hour, $\mu = 12$ trains/hour, $\rho = \dfrac{\lambda}{\mu} = \dfrac{6}{12} = 0.5$.

Since the maximum queue length is 2, the maximum number of trains in the system is 3 (= N).

Now, probability that there is no train in the system, $p_0 = \dfrac{1-\rho}{1-\rho^{N+1}}$.

$$\therefore \qquad p_0 = \frac{1-0.5}{1-(0.5)^{3+1}} = \frac{0.5}{1-(0.5)^4} = 0.53.$$

Since $p_n = p_0 \,.\, \rho^n$, $p_1 = 0.53 \times (.5)^1 = 0.265$, $p_2 = 0.53\ (.5)^2 = 0.13$, $p_3 = 0.53\ (.5)^3 = 0.066$.

$\therefore$ Average number of trains in the system,

$$\mathrm{L}_s = 1p_1 + 2p_2 + 3p_3$$

or $$\mathrm{L}_s = 1 \times 0.265 + 2 \times 0.13 + 3 \times 0.066 = 0.723.$$

To find W_q

$$\mathrm{L}_q = \frac{1 - \mathrm{N}\,\rho^{\mathrm{N}-1} + (\mathrm{N}-1)\,\rho^{\mathrm{N}}}{(1-\rho)\,(1-\rho^{\mathrm{N}+1})} \,.\, \rho^2 = \frac{1 - 3 \times 0.5^2 + 2 \times 0.5^3}{(1-0.5)\,[1-(0.5)^4]} \times (0.5)^2$$

$$= \frac{1 - 0.75 + 0.25}{0.5 \times 0.9375} \times 0.25 = 0.267.$$

$$\lambda' = \lambda\,(1 - p_{\mathrm{N}}) = 6\,(1 - p_3) = 6\,(1 - 0.066) = 5.604.$$

$\therefore$ $$\mathrm{W}_q = \frac{\mathrm{L}_q}{\lambda'} = \frac{0.267}{5.604} = 0.0476 \text{ hours} = 2.85 \text{ minutes.}$$

Simulation

Module6

Syllabus

Introduction to simulation, steps in simulation, advantages of simulation, limitations of simulation, applications of simulation, Monte-Carlo method: simple examples, single server queue model.

Introduction to Simulation

- A *simulation* is the imitation of the operation of real-world process or system over time.
  - Generation of artificial history and observation of that observation history
- A *model* construct a conceptual framework that describes a system
- The behavior of a system that evolves over time is studied by developing a simulation *model*.
- The model takes a set of expressed assumptions:
  - Mathematical, logical
  - Symbolic relationship between the *entities*

Goal of modeling and simulation

- A model can be used to investigate a wide verity of "what if" questions about real-world system.
  - Potential changes to the system can be simulated and predicate their impact on the system.
  - Find adequate parameters before implementation
- So simulation can be used as
  - Analysis tool for predicating the effect of changes
  - Design tool to predicate the performance of new system
- It is better to do simulation before Implementation.

How a model can be developed?

- Mathematical Methods
  - Probability theory, algebraic method ,…
  - Their results are accurate
  - They have a few Number of parameters
  - It is impossible for complex systems
- Numerical computer-based simulation
  - It is simple
  - It is useful for complex system

When Simulation Is the Appropriate Tool

- Simulation enable the study of internal interaction of a subsystem with complex system
- Informational, organizational and environmental changes can be simulated and find their effects
- A simulation model help us to gain knowledge about improvement of system
- Finding important input parameters with changing simulation inputs
- Simulation can be used with new design and policies before implementation
- Simulating different capabilities for a machine can help determine the requirement
- Simulation models designed for training make learning possible without the cost disruption
- A plan can be visualized with animated simulation
- The modern system (factory, wafer fabrication plant, service organization) is too complex that its internal interaction can be treated only by simulation

When Simulation Is Not Appropriate

- When the problem can be solved by common sense.
- When the problem can be solved analytically.
- If it is easier to perform direct experiments.
- If cost exceed savings.
- If resource or time are not available.
- If system behavior is too complex.
  - Like human behavior

Advantages and disadvantages of simulation

- In contrast to optimization models, simulation models are “run” rather than solved.
  - Given as a set of inputs and model characteristics the model is run and the simulated behavior is observed

Advantages of simulation

- New policies, operating procedures, information flows and son on can be explored without disrupting ongoing operation of the real system.
- New hardware designs, physical layouts, transportation systems and … can be tested without committing resources for their acquisition.
- Time can be compressed or expanded to allow for a speed-up or slow-down of the phenomenon(clock is self-control).
- Insight can be obtained about interaction of variables and important variables to the performance.
- Bottleneck analysis can be performed to discover where work in process, the system is delayed.
- A simulation study can help in understanding how the system operates.
- “What if” questions can be answered.

Disadvantages of simulation

- Model building requires special training.
  - Vendors of simulation software have been actively developing packages that contain models that only need input (templates).
- Simulation results can be difficult to interpret.
- Simulation modeling and analysis can be time consuming and expensive.
  - Many simulation software have output-analysis.

Areas of application

- Manufacturing Applications
- Semiconductor Manufacturing
- Construction Engineering and project management
- Military application
- Logistics, Supply chain and distribution application
- Transportation modes and Traffic
- Business Process Simulation
- Health Care
- Automated Material Handling System (AMHS)
  - Test beds for functional testing of control-system software
- Risk analysis
  - Insurance, portfolio,...
- Computer Simulation
  - CPU, Memory,…
- Network simulation
  - Internet backbone, LAN (Switch/Router), Wireless, PSTN (call center),...

Systems and System Environment

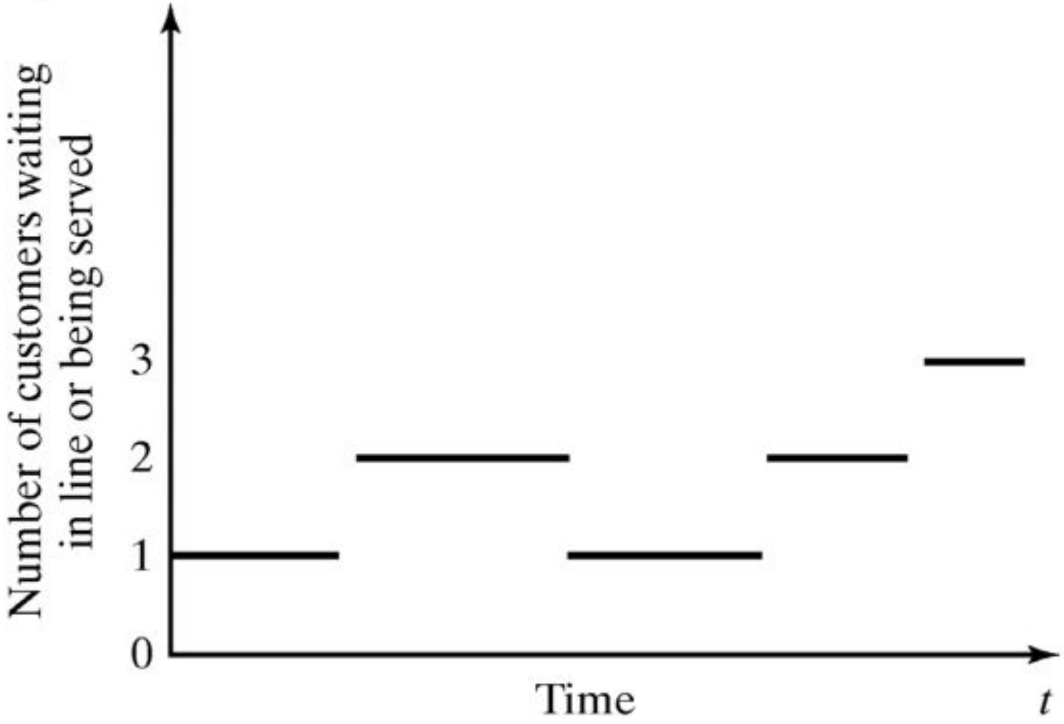
- A *system* is defined as a groups of objects that are joined together in some regular interaction toward the accomplishment of some purpose.
  - An automobile factory: Machines, components parts and workers operate jointly along assembly line
- A system is often affected by changes occurring outside the system: system environment.
  - Factory : Arrival orders
    - Effect of supply on demand : relationship between factory output and arrival (activity of system)
  - Banks : arrival of customers

Components of system

- Entity
  - An object of interest in the system : Machines in factory
- Attribute
  - The property of an entity : speed, capacity
- Activity
  - A time period of specified length :welding, stamping
- State
  - A collection of variables that describe the system in any time : status of machine (busy, idle, down,…)
- Event
  - A instantaneous occurrence that might change the state of the system: breakdown
- Endogenous
  - Activities and events occurring with the system
- Exogenous
  - Activities and events occurring with the environment

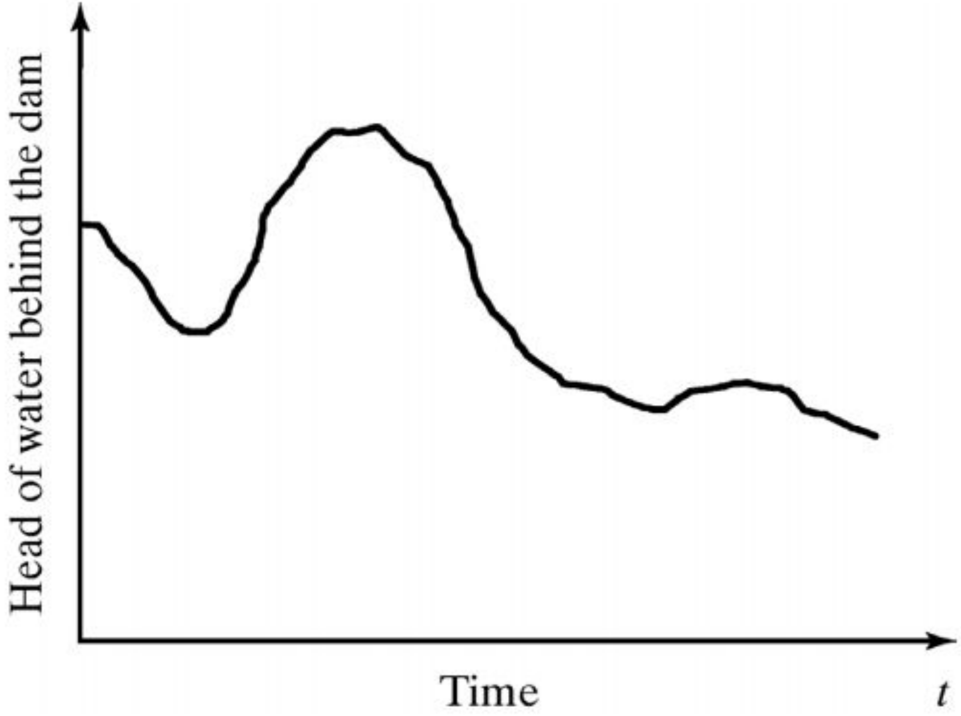
Discrete and Continues Systems

- A discrete system is one in which the state variables change only at a discrete set of points in time : Bank example



Discrete and Continues Systems (cont.)

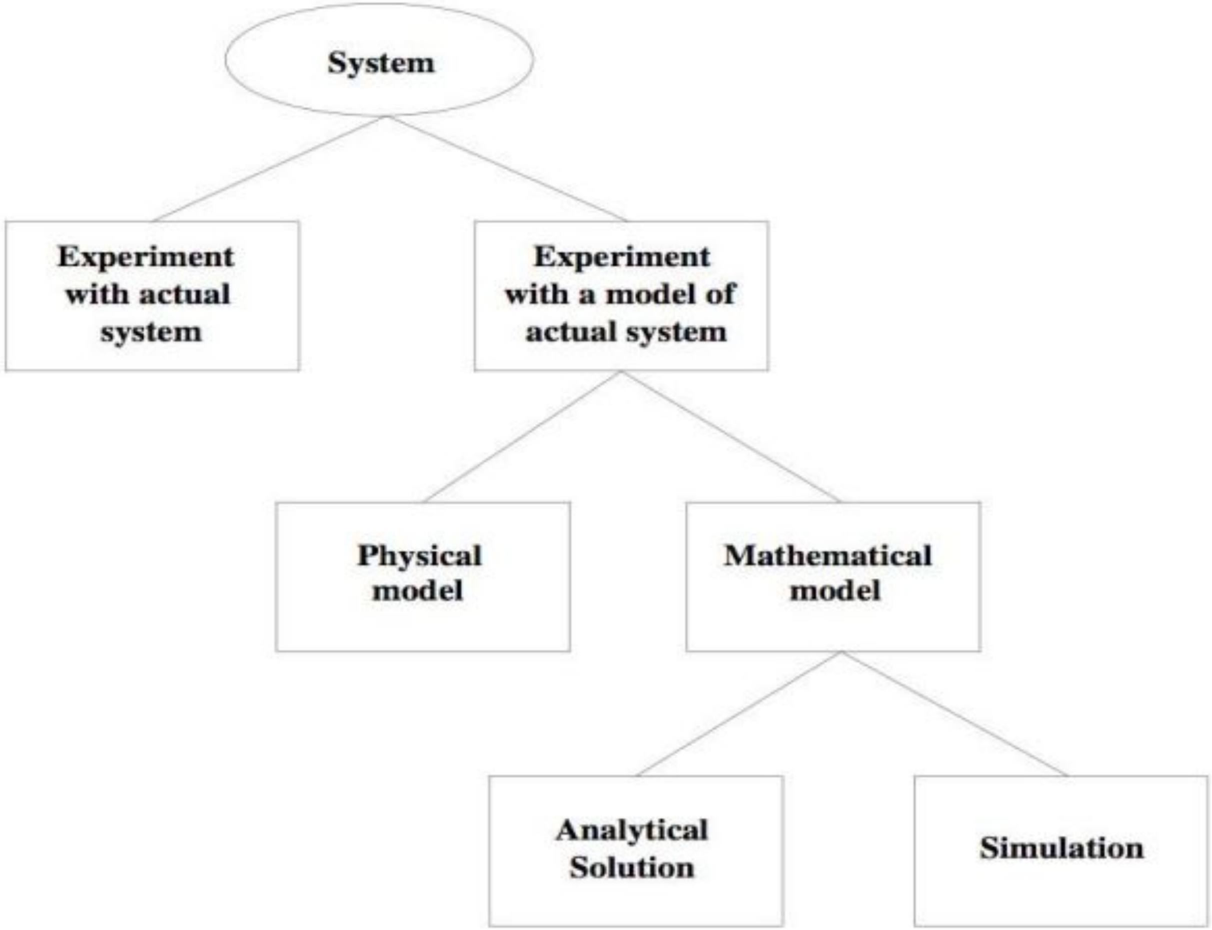
- A continues system is one in which the state variables change continuously over time: Head of water behind the dam



Model of a System

- To study the system
  - it is sometimes possible to experiments with system
    - This is not always possible (bank, factory,…)
    - A new system may not yet exist
- Model: construct a conceptual framework that describes a system
  - It is necessary to consider those accepts of systems that affect the problem under investigation (unnecessary details must remove)

Types of Models



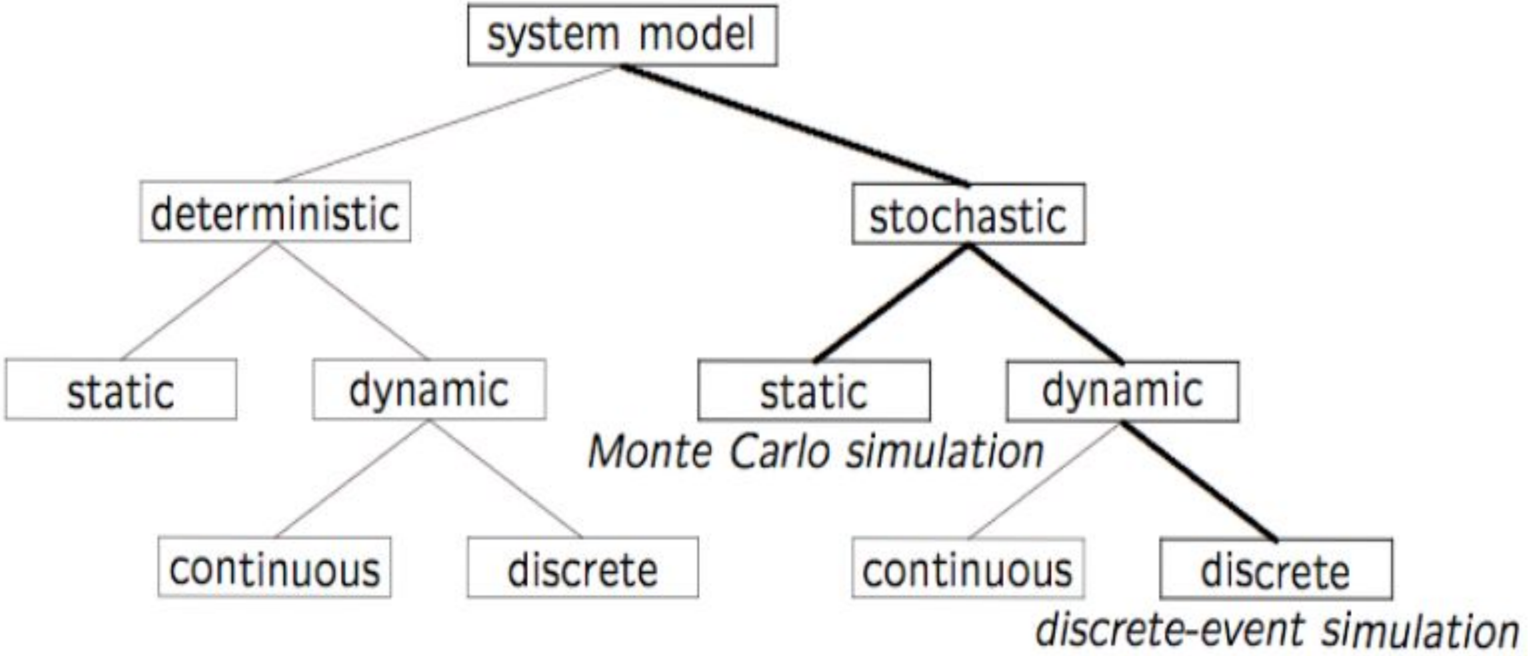
Characterizing a Simulation Model

- Deterministic or Stochastic
  - Does the model contain stochastic components?
  - Randomness is easy to add to a DES
- Static or Dynamic
  - Is time a significant variable?
- Continuous or Discrete
  - Does the system state evolve continuously or only at discrete points in time?
  - Continuous: classical mechanics
  - Discrete: queuing, inventory, machine shop models

Discrete-Event Simulation Model

- Stochastic: some state variables are random
- *Dynamic*: time evolution is important
- *Discrete-Event*: significant changes occur at discrete time instances

Model Taxonomy



DES Model Development

How to develop a model:

1) Determine the goals and objectives
2) Build a ***conceptual*** model
3) Convert into a ***specification*** model
4) Convert into a ***computational*** model
5) Verify
6) Validate

Typically an iterative process

Three Model Levels

- Conceptual
  - Very high level
  - How comprehensive should the model be?
  - What are the *state variables,* which are dynamic, and which are important?
- Specification
  - On paper
  - May involve equations, pseudocode, etc.
  - How will the model receive input?
- Computational
  - A computer program
  - General-purpose PL or simulation language?

Verification vs. Validation

- *Verification*
  - Computational model should be consistent with specification model
  - Did we build the model right?
- *Validation*
  - Computational model should be consistent with the system being analyzed
  - Did we build the right model?
  - Can an expert distinguish simulation output from system output?
- Interactive graphics can prove valuable

Steps in simulation study

1. - Clearly understand problem
   - Reformulation of the problem

2. - Which questions should be answered?
   - Is simulation appropriate?
   - Alternate sys designs,Costs?

3. -complexity
   - involve Model user

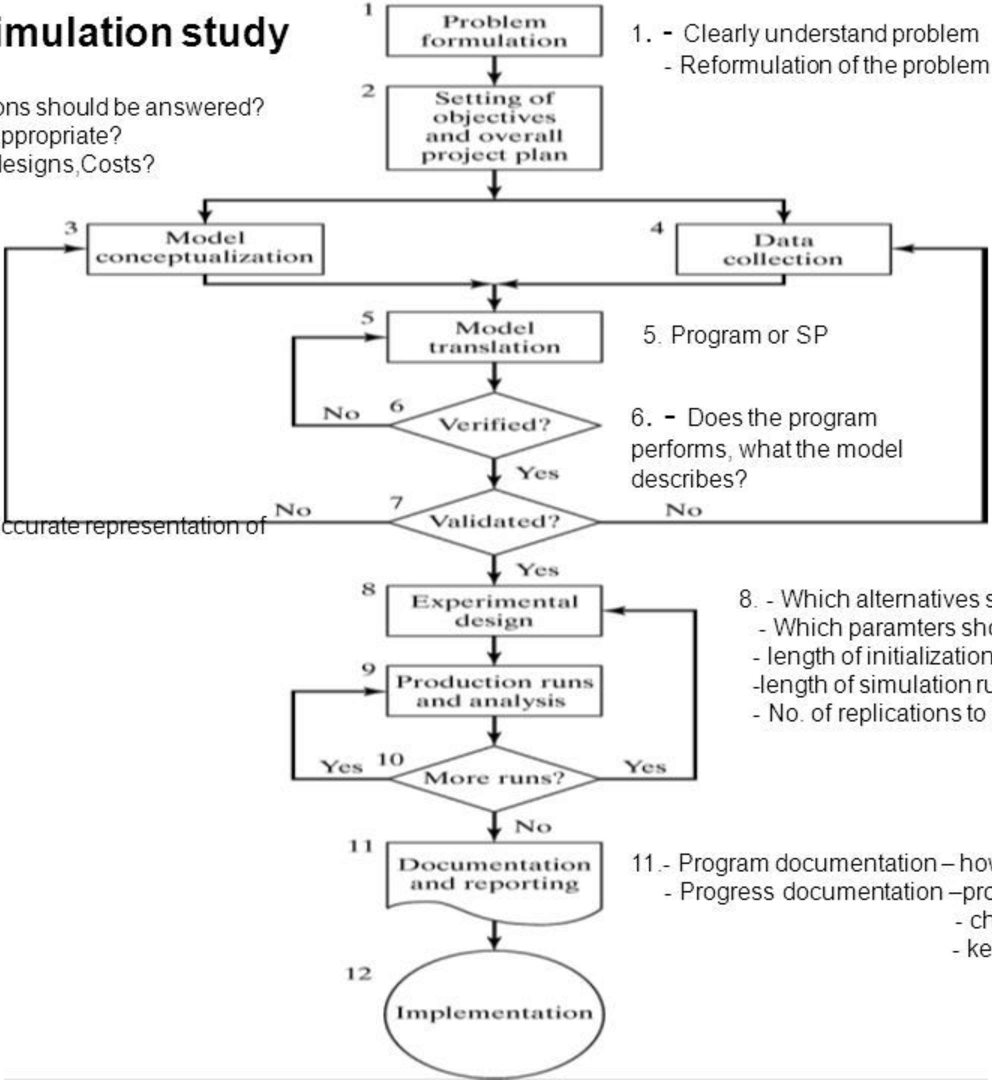
5. Program or SP

6. - Does the program performs, what the model describes?

7. –Is the model accurate representation of real system

8. - Which alternatives should be run?
   - Which paramters should be varied?
   - length of initialization period
   -length of simulation run
   - No. of replications to be made of each run.

11.- Program documentation – how does the prog work
   - Progress documentation –project history
   - chronology of the work
   - keep project on course



Monte Carlo Simulation

https://www.youtube.com/watch?v=7ESK5SaP-bc

Simulation Examples

Simulation steps using Simulation Table

1. **Determine the characteristics of each of the inputs to the simulation (probability distributions).**

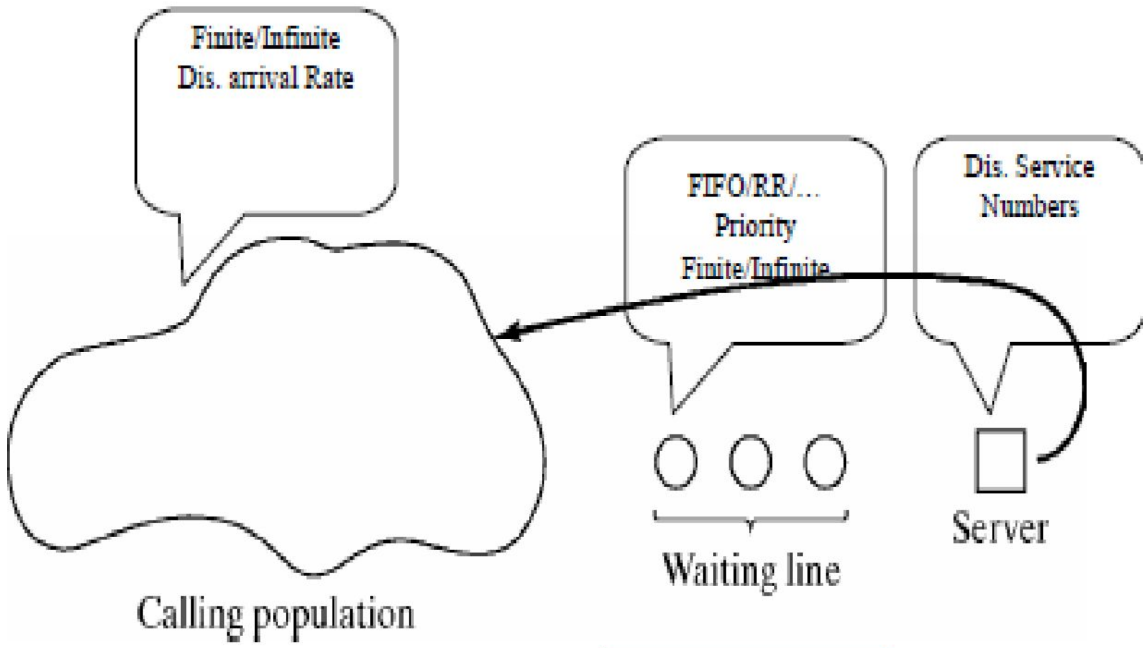
2. **Construct a simulation table (repetition 1).**

3. **For each repetition *i*, generate a value for the inputs, and evaluate function, calculating a value of response yi.**

Simulation Table

	Inputs				Response
Repetition	X1	x2	….	xp	yi
1					
2					
:					
n					

Simulation of Queuing System (Details in pre. unit)

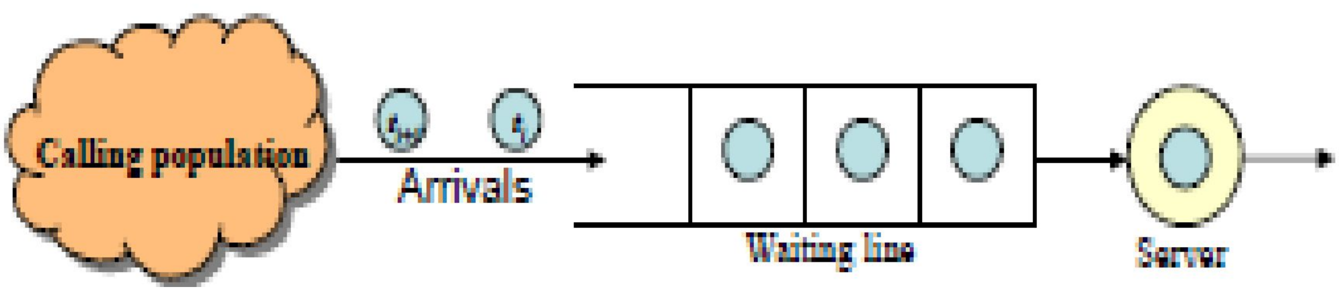


Attributes:

1. **Calling Population**
2. **Nature of arrivals**
3. **Service Mechanism**
4. **System capacity**
5. **Queuing Discipline**

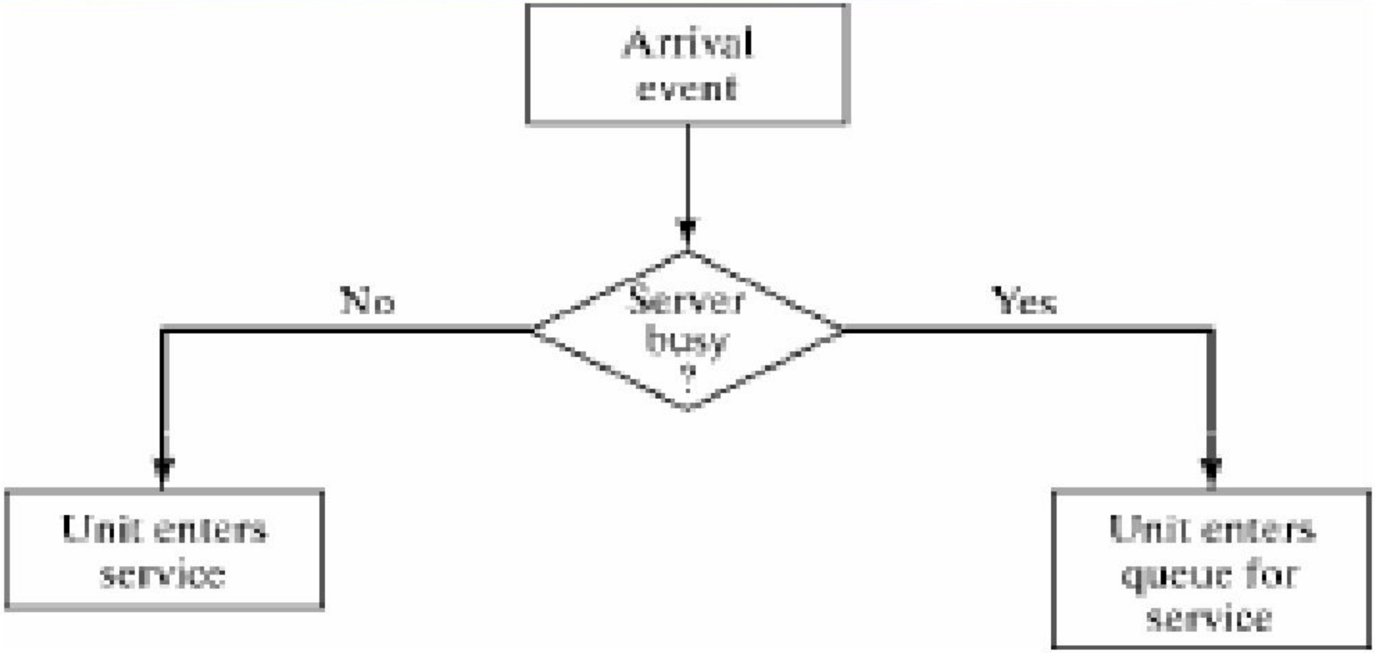
- **Single server queue:**
  - **Calling population is infinite-Arrival rate does not change**
  - **Units are served according FIFO**
  - **Arrivals are defined by the distribution of the time between arrivals - inter-arrival time**
  - **Service times are according to distribution**
  - **Arrival rate must be less than service rate- stable system**

Otherwise waiting line will grow unbounded.



- **Queuing system :**
  - *System*
    - **Server**
    - **Units (in queue or being served)**
    - **Clock**
  - *State of the system*
    - **Number of units in the system**
    - **Status of server (idle, busy)**
  - *Events*
    - **Arrival of a unit**
    - **Departure of a unit**

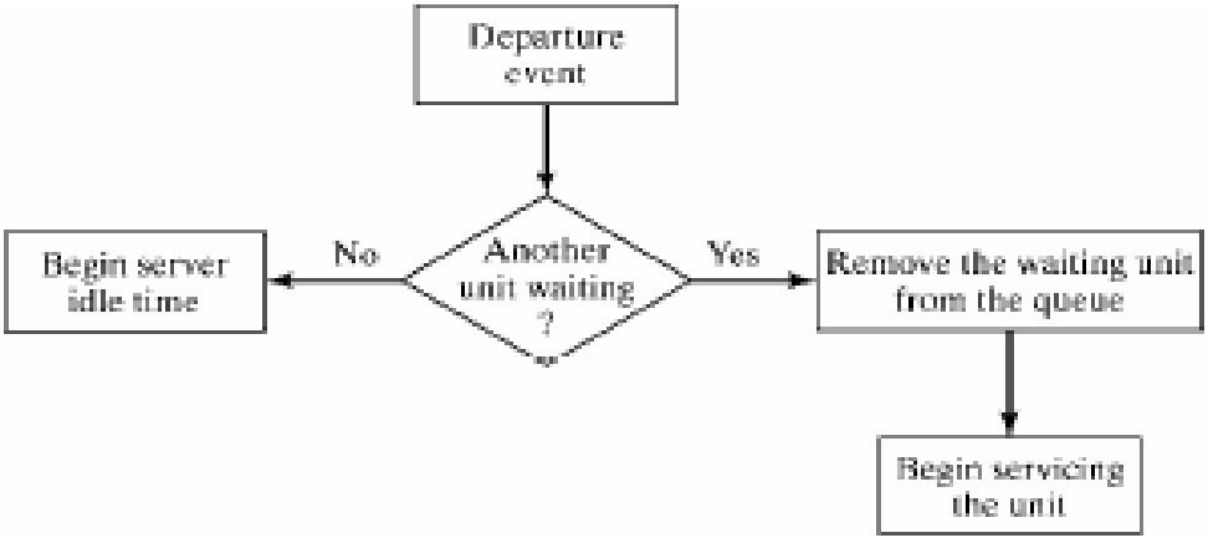
- **Arrival Event**
  - **If server is idle customer gets service, otherwise customer enters queue.**



		Queue status	
		Not empty	Empty
Server status	Busy	Enter queue	Enter queue
	Idle	Impossible	Enter service

Departure Event

If queue is not empty then server begin servicing next unit, otherwise server will be idle.



		Queue status	
		Not empty	Empty
Server status	Busy		Impossible
	Idle	Impossible	

Grocery Store Example(Ex 1)

• Producing Random Numbers from Random Digits

• Select randomly a number, e.g. 99219
- One digit: 0.9
- Two digits: 0.19
- Three digits: 0.219

• Proceed in a systematic direction, e.g.
- first down then right
- first up then left

The interarrival and service times are taken from distributions!

Customer	Interarrival Time	Arrival Time on Clock	Service Time
1	-	0	2
2	2	2	1
3	4	6	3
4	1	7	2
5	2	9	1
6	6	15	4

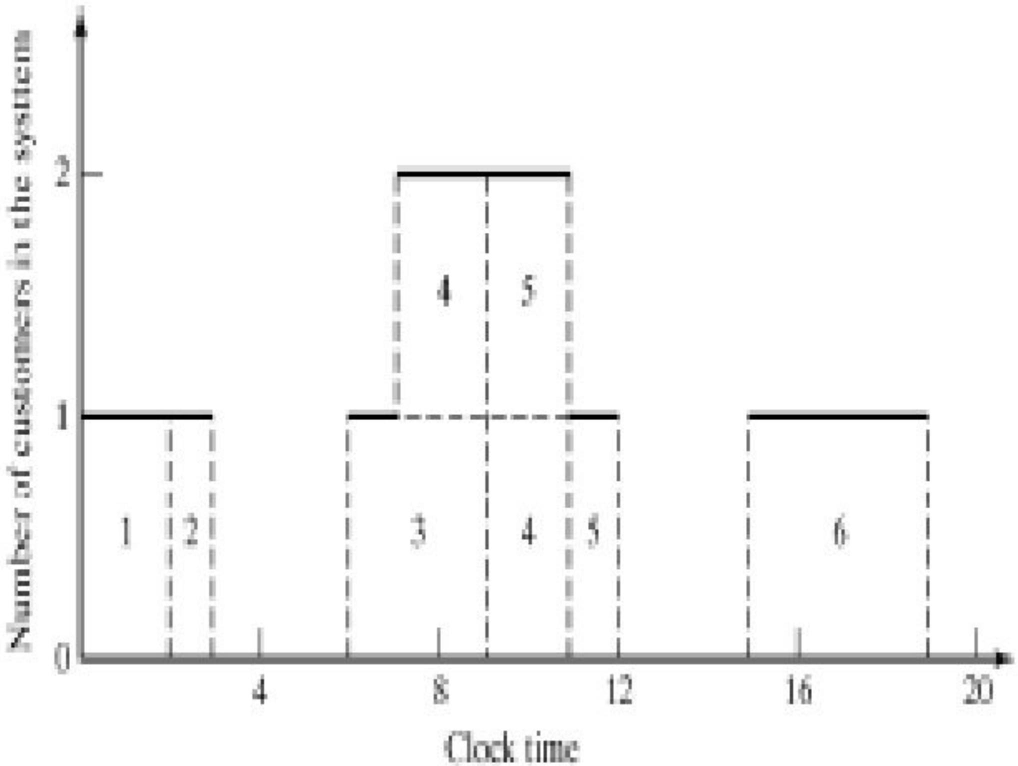
The simulation run is build by meshing clock, arrival and service times!

Customer Number	Arrival Time [Clock]	Time Service Begins [Clock]	Service Time [Duration]	Time Service Ends [Clock]
1	0	0	2	2
2	2	2	1	3
3	6	6	3	9
4	7	9	2	11
5	9	11	1	12
6	15	15	4	19

Chronological ordering of events

Clock Time	Customer Number	Event Type	Number of customers
0	1	Arrival	1
2	1	Departure	0
2	2	Arrival	1
3	2	Departure	0
6	3	Arrival	1
7	4	Arrival	2
9	3	Departure	1
9	5	Arrival	2
11	4	Departure	1
12	5	Departure	0
16	6	Arrival	1
19	6	Departure	0

Number of customers in the system



• Example1: A Grocery Store

- **Analysis of a small grocery store**
  - One checkout counter
  - Customers arrive at random times from {1,2,…,8} minutes
  - Service times vary from {1,2,…,6} minutes
  - Consider the system for 100 customers
- **Problems/Simplifications**
  - Sample size is too small to be able to draw reliable conclusions
  - Initial condition is not considered

Interarrival Time	Probability	Cumulative Probability
1	0.125	0.125
2	0.125	0.250
3	0.125	0.375
4	0.125	0.500
5	0.125	0.625
6	0.125	0.750
7	0.125	0.875
8	0.125	1.000

Service Time	Probability	Cumulative Probability
1	0.10	0.10
2	0.20	0.30
3	0.30	0.60
4	0.25	0.85
5	0.10	0.95
6	0.05	1.00

Customer	Interarrival Time [Minutes]	Arrival Time [Clock]	Service Time [Minutes]	Time Service Begins [Clock]	Time Service Ends [Clock]	Waiting Time in Queue [Minutes]	Time Customer in System [Minutes]	Idle Time of Server [Minutes]
1	-	0	4	0	4	0	4	0
2	1	1	2	4	6	3	5	0
3	1	2	5	6	11	4	9	0
4	6	8	4	11	15	3	7	0
5	3	11	1	15	16	4	5	0
6	7	18	5	18	23	0	5	2
...								
100	5	415	2	416	418	1	3	0
Total	415		317			174	491	101

- **Average waiting time**

$$\bar{w} = \frac{\sum \text{Waiting time in queue}}{\text{Number of customers}} = \frac{174}{100} = 1.74\,\text{min}$$

- **Probability that a customer has to w**

$$p(\text{wait}) = \frac{\text{Number of customer who wait}}{\text{Number of customers}} = \frac{46}{100} = 0.46$$

- **Proportion of server idle time**

$$p(\text{idle server}) = \frac{\sum \text{Idle time of server}}{\text{Simulation run time}} = \frac{101}{418} = 0.24$$

- **Average service time**

$$\bar{s} = \frac{\sum \text{Service time}}{\text{Number of customers}} = \frac{317}{100} = 3.17\,\text{min}$$

$$E(s) = \sum_{s=0}^{\infty} s \cdot p(s) = 0.1 \cdot 10 + 0.2 \cdot 20 + \cdots + 0.05 \cdot 6 = 3.2\,\text{min}$$

- **Average time between arrivals**

$$\bar{\lambda} = \frac{\sum \text{Times between arrivals}}{\text{Number of arrivals - 1}} = \frac{415}{99} = 4.19 \text{ min}$$

$$E(\lambda) = \frac{a+b}{2} = \frac{1+8}{2} = 4.5 \text{ min}$$

- **Average waiting time of those who wait**

$$\bar{w}_{waited} = \frac{\sum \text{Waiting time in queue}}{\text{Number of customers that wait}} = \frac{174}{54} = 3.22 \text{ min}$$

- **Average time a customer spends in system**

$$\bar{t} = \frac{\sum \text{Time customers spend in system}}{\text{Number of customers}} = \frac{491}{100} = 4.91 \text{ min}$$

$$\bar{t} = \bar{w} + \bar{s} = 1.74 + 3.17 = 4.91 \text{min}$$

Example 2

Simulation Problem

Problem : A dentist schedule her patients for 30 min appointments. Some take more or less time than 30 min depending the type of dental work. Following table shows Probabilities and the actual time taken to complete the job.

Category	Time required	No. of patients	Probability
Filling	45 min	40	0.40
Crown	60 min	15	0.15
Cleaning	15 min	15	0.15
Extracting	45 min	10	0.10
Check up	15 min	20	0.20

Simulate the doctor’s clinic for 4 hrs.Find the average waiting time for patients and idle time for doctor. Assuming all patients show up at their scheduled time, arrival time starting at 8 am. Use following random numbers to solve the problem

40, 82, 11, 34, 25, 66, 17, 79

Simulating for 8 patients.
Cumulative distribution table.

Table 1

Category	Probability	Cumulative Probability	Random number
Filling	0.40	0.40	0- 39
Crown	0.15	0.55	40-54
Cleaning	0.15	0.70	55-69
Extracting	0.10	0.80	70-79
Check up	0.20	1.00	80-99

Setting random number intervals

Table 2

Patient	Arrival time	Random number	category	Service time needed
1	8:00	40	Crown	60
2	8:30	82	Checkup	15
3	9:00	11	Filling	45
4	9:30	34	Filling	45
5	10:00	25	Filling	45
6	10:30	66	Cleaning	15
7	11:00	17	Filling	45
8	11:30	79	Extracting	45

Simulation table

Table 3

Patient	Arrival time	Service start	service duration	Service ends	Waiting time	Idle time
1	8:00	8:00	60	9:00	0	0
2	8:30	9:00	15	9:15	30	0
3	9:00	9:15	45	10:00	15	0
4	9:30	10:00	45	10:45	30	0
5	10:00	10:45	45	11:30	45	0
6	10:30	11:30	15	11:45	60	0
7	11:00	11:45	45	12:30	45	0
8	11:30	12:30	45	1:15	60	0
Total					285	

Average waiting time = 285 /8 = 35.62 minutes

Idle time for doctor =0

Monte Carlo Simulation

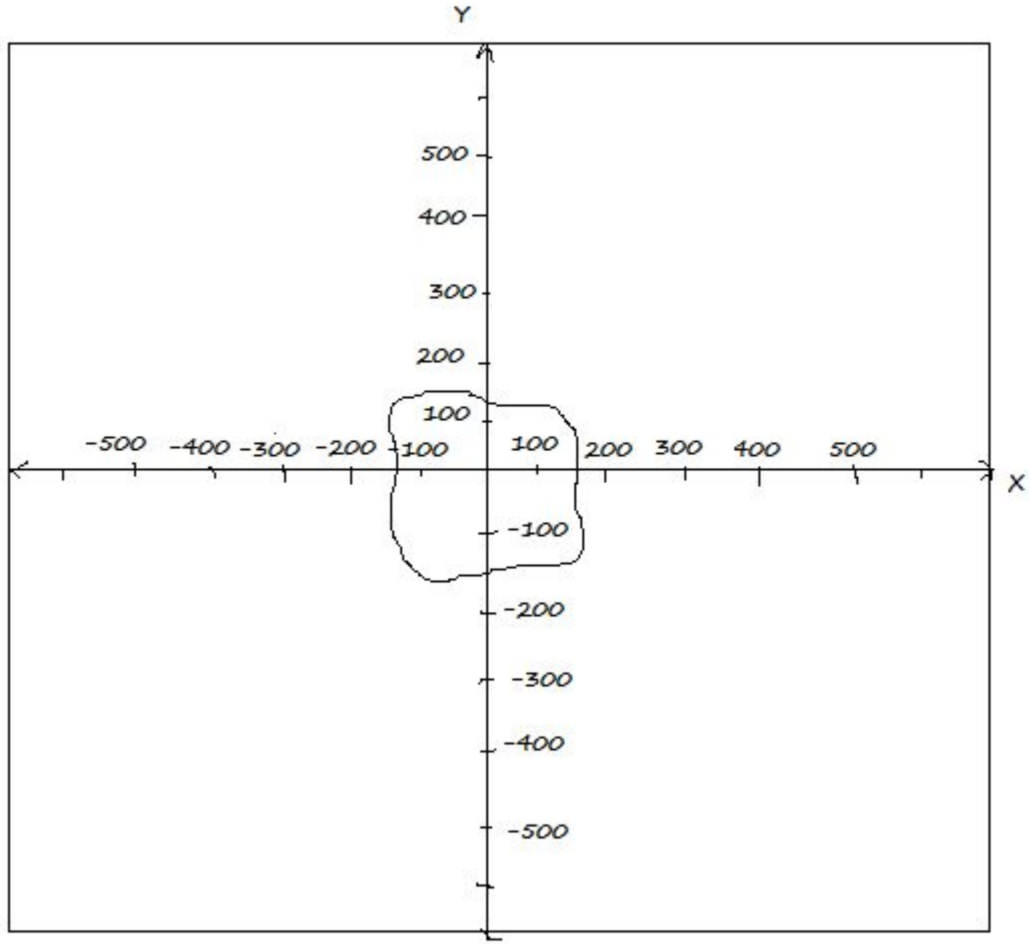
Monte Carlo simulation is a computerized mathematical technique that allows people to account for risk in quantitative analysis and decision making. The technique is used by professionals in such widely disparate fields as finance, project management, energy, manufacturing, engineering, research and development, insurance, oil & gas, transportation, and the environment.

Monte Carlo simulation furnishes the decision-maker with a range of possible outcomes and the probabilities they will occur for any choice of action. It shows the extreme possibilities—the outcomes of going for broke and for the most conservative decision—along with all possible consequences for middle-of-the-road decisions.

The technique was first used by scientists working on the atom bomb; it was named for Monte Carlo, the Monaco resort town renowned for its casinos. Since its introduction in World War II, Monte Carlo simulation has been used to model a variety of physical and conceptual systems.

Example :

A squad of bombers are attempting to destroy an ammunition depot as shown in figure. If the bomb lands anywhere on the depot it's hit otherwise bomb is a miss. The aiming point is the dot located at the heart of the ammunition depot. The point of impact is assumed to be normally distributed around the aiming point with a standard deviation of 600 mtrs in horizontal direction and 300 in vertical ditrection. The problem is to simulate the operation and make a statement about the number of bombs on the target.



Ammunition Depot

Random numbers for x- coordinate -0.84,1.03,0.92,-1.82,-0.16,-1.78,2.04 , 1.08,-1,50,-0.42
Random numbers for y- coordinate 0.66,-0.13,0.06,-1.40,0.23 , 1.33,0.69,-1.10,-0.72,-0.60

20/11/

In this eg the aiming point is considered as (0,0) hence mean in horizontal direction & vertical direction is assumed to be zero.

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

$$z\sigma = x - \mu$$

$$z\sigma + \mu = x$$

$$\therefore \underline{\mu = 0}$$

$$\therefore y = z\sigma$$

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

$$x = z\sigma + \mu$$

$$\therefore x = z\sigma$$

σ = 600 m (horizontal) 300 m (vertical)

$$\therefore \cancel{x = z\,600}$$

$$x = 600z$$

$$\cancel{x = z\,300}$$

$$y = 300z$$

Bomber	Random digits (x)	X co-ordinate (600 Z)	Random digits (y)	Y co-ordinate (300 Z)	Result
1	-0.84	$-0.84 \times 600 =$ -504	0.66	$0.66 \times 300 =$ 198	miss.
2	1.03	$1.03 \times 600 =$ 618	-0.13	-39	miss
3	0.92	552	0.06	18	miss
4	-1.82	-1092	-1.40	-420	miss
5	-0.16	-96	0.23	69	Hit
6	-1.78	-1068	1.33	399	miss
7	2.04	1224	0.69	207	miss
8	1.08	648	-1.10	-330	miss
9	-1.50	-900	-0.72	-216	miss
10	-0.42	-252	-0.60	-180	hit.

~~So~~ The result is that 10 bombers had 2 hits & 8 misses. This is an eg of Monty Carlo or Static simulation because time is not a factor in the solution.

https://www.youtube.com/watch?v=7ESK5SaP-bc