PROBABILITY FOR COMPUTING PRACTICALS

RAMANUJAN COLLEGE

UNIVERSITY OF DELHI

Discipline Specific Course: 06

Probability for Computing

Semester – 02

SESSION (2024 -25)

Submitted By

Mansi Gupta

College Roll No.: 24570036

University Roll No.: 24020570035

B.Sc. (H) Computer Science

Submitted To

Dr. Aakash (Assistant Professor) (Operational Research)

Department of Computer Science

Ramanujan College

University of Delhi

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to Dr. Aakash, Assistant Professor (Operational Research) Ramanujan College, University of Delhi for providing me the opportunity to work on this practical file of Probability for Computing, his guidance and encouragement have been invaluable throughout the preparation of this file.

I am also thankful to my peers and the faculty members of the Computer Science Department for their continuous support and invaluable insights, which have greatly enriched my understanding of the subject.

Finally, I extend my sincere appreciation to my family and friends for their constant support and encouragement which has been a source of strength in completing this file successfully.

Mansi Gupta

College Roll No. – 24570036

University Roll No. – 24020570035

INDEX

1. Plotting and fitting of Binomial distribution and graphical representation of probabilities

<u>Binomial Distribution:</u>

The binomial distribution is a discrete probability distribution. It describes the outcome of binary scenarios, e.g. toss of a coin.

<u>Binomial Distribution Formula:</u>

$$\mathrm{P(x;n,p)} = C_x^n \, p^x (1 - p)^{n-x}$$

Where:

P(x;n,p) is the probability of x successes in n trials in an experiment which can result in exactly two outcomes (success or failure).

p is the probability of success on an individual trial.

n is the number of trials.

x is the total number of success.

Mean ($\mu = \mathrm{np}$)

Variance ($\sigma^2 = npq$)

Implementation in Excel:

=BINOM.DIST(number_s, trials, probability_s, cumulative)

Where:
number_s: number of successes

trials: total number of trials

probability_s: probability of success on each trial

cumulative: TRUE returns the cumulative probability; FALSE returns the exact probability

<u>Skewness in Case Binomial Distribution can be defined as follow:</u>

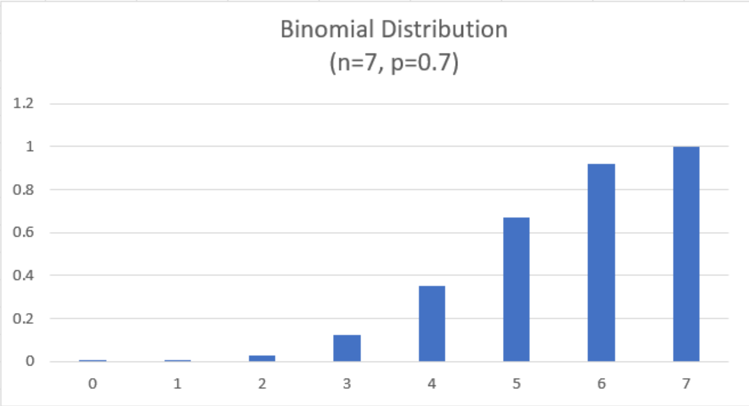
If p = 0.5, the binomial distribution will be symmetrical, regardless of the value of n.

If p ≠ 0.5, the distribution will be skewed.

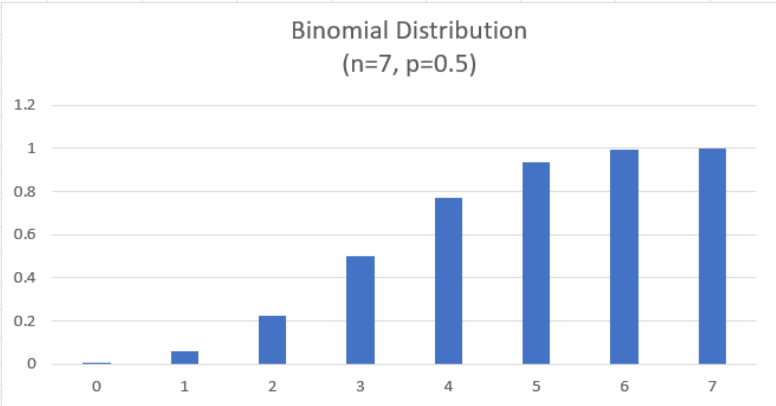
If p < 0.5, the distribution will be positively skewed or right – skewed. This means the bulk of the probability falls in the smaller numbers and the distribution tails off to the right.

If p > 0.5, the distribution will be negatively skewed or left – skewed. This means the bulk of the probability falls in the larger numbers and the distribution tails off to the left.

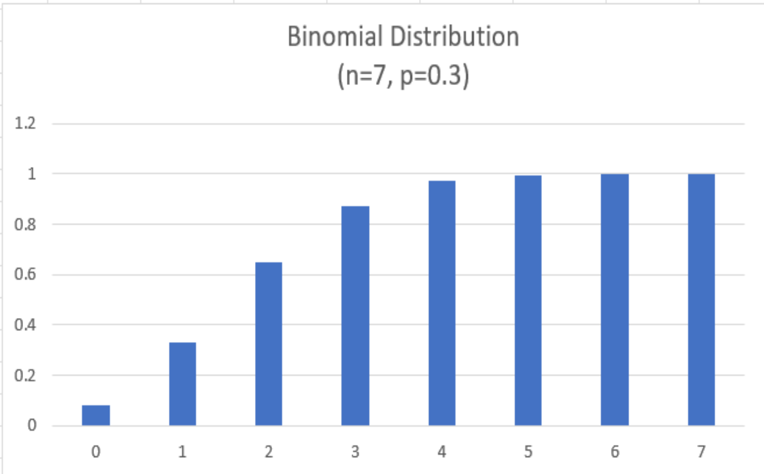
n(Number of Trials)	7
p(Probability of Success on given Trails)	0.7
k(Number of Success)	Binomial Distribution
0	0.0002187
1	0.0037908
2	0.0287955
3	0.126036
4	0.3529305
5	0.6705828
6	0.9176457
7	1



n(Number of Trials)	7
p(Probability of Success on given Trails)	0.5
k(Number of Success)	Binomial Distribution
0	0.0078125
1	0.0625
2	0.2265625
3	0.5
4	0.7734375
5	0.9375
6	0.9921875
7	1



n(Number of Trials)	7
p(Probability of Success on given Trails)	0.3
k(Number of Success)	Binomial Distribution
0	0.0823543
1	0.3294172
2	0.6470695
3	0.873964
4	0.9712045
5	0.9962092
6	0.9997813
7	1



2. Plotting and fitting of Multinomial Distribution and Graphical Representation of Probabilities

The multinomial distribution is a multivariate generalization of the binomial distribution. Consider a trial that results in exactly one of some fixed finite number k **of possible outcomes, with probabilities** $p_1, p_2, \ldots., p_k$ **(so that** $p_i \geq 0$ for $\mathrm{i} = 1, \ldots, \mathrm{k}$ **and** $\sum i = 1kpi = 1$**), and there are** n **independent trials. Then let the random variables** X_i **indicate the number of times outcome number** i **was observed over the** n **trials. Then** $\mathrm{X} = (X_{1,}\ X_{2,} \ldots, X_k)$ **follows a multinomial distribution with parameters** n **and** p**, where** $\mathrm{p} = (p_{1,}\ p_2, \ldots, p_k)$**.**

Multinomial Distribution Formula

$$p(x_1, x_2 \ldots \ldots x_k) = \left[\frac{n!}{x_1!\,.\,x_2!\ldots\ldots x_k!}\right] \cdot p_1^{x_1} \cdot p_2^{x_2} \ldots \ldots p_k^{x_k}$$

$$\mathrm{Cov}(X_i, X_j) = -np_ip_j \quad (i \neq j)$$

When $\mathrm{X} = (x_1, x_2, \ldots, x_k)$ **follows a multinomial distribution with the PMF given above,** X_i **follows a binomial distribution with** n **trials and success probability** p_i**.**

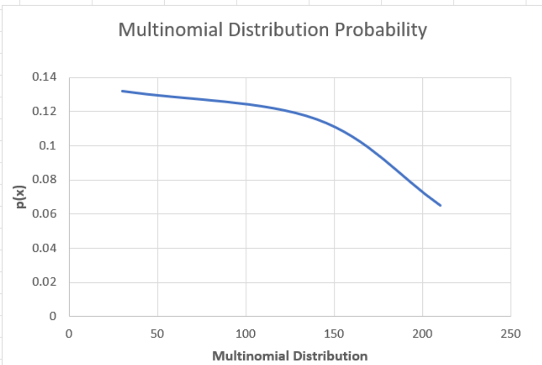
Implementation in Excel

Multinomial = MULTINOMIAL(X1, X2, X3)

Probability = MULTINOMIAL*PRODUCT($p1^{X1}$, $p2^{X2}$, $p3^{X3}$**)**

		1st Case	2nd Case	3rd Case
Apple	X1	4	3	2
Berry	X2	0	1	1
Orange	X3	6	3	2
P_Apple	P1	0.375	0.375	0.375
P_Berry	P2	0.125	0.125	0.125
P_Orange	P3	0.5	0.5	0.5
	Multinomial	210	140	30
	P1^X1	0.019775391	0.052734375	0.140625
	P2^X2	1	0.125	0.125
	P3^X3	0.015625	0.125	0.25
	Probability	0.064888	0.115356445	0.131835938

Multinomial	Multinomial Distribution Probability
210	0.064888
140	0.115356445
30	0.131835938



3. Plotting and fitting of Poisson distribution and graphical representation of probabilities

The Poisson distribution is a type of discrete probability distribution that determines the likelihood of an event occurring a specific number of times (k) within a designated time or space interval. This distribution is characterized by a single parameter, λ (lambda), representing the average number of occurrences of the event.

Poisson Distribution Formula

$$\mathrm{P}\ (\mathrm{K}) = e^{-\lambda}\ \frac{\lambda^{k}}{k\,!}$$

Mean	$\mu = E(X) = \lambda$
Variance	$\sigma^2 = V(X) = \lambda$
Standard Deviation	$\sigma = \sqrt{\sigma^2} = \sqrt{\lambda}$

Where:

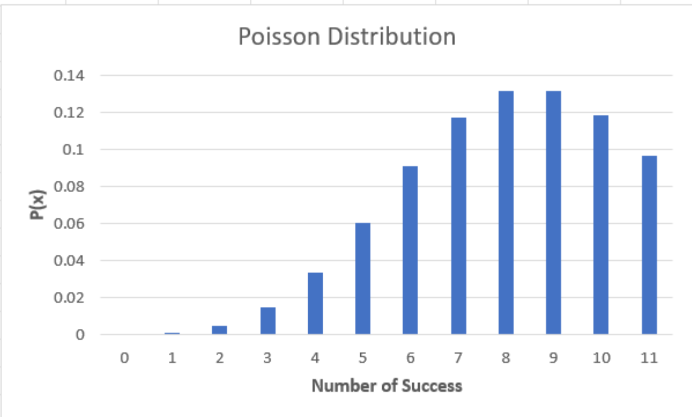
- P(X = k) is the probability of observing k events
- e is the base of the natural logarithm (approx. 2.71828)
- λ means number of successes that occur during a specific time interval, $\lambda = np$
- k is the number of successes

Excel Implementation

POISSON.DIST(number_s, average, cumulative)

POISSON.DIST(k, λ, FALSE)

λ(Mean Number of Success)	9
k(Number of Success)	Poisson Distribution
0	0.00012341
1	0.001110688
2	0.004998097
3	0.014994291
4	0.033737155
5	0.060726879
6	0.091090319
7	0.117116124
8	0.13175564
9	0.13175564
10	0.118580076
11	0.097020062



4. Plotting and fitting of Geometric Distribution and Graphical Representation of probabilities

In a Bernoulli Trial, the likelihood of the number of successive failures before success is obtained is represented by a geometric distribution, which is a sort of discrete probability distribution. A Bernoulli trial is a test that can only have one of two outcomes: success or failure. In other words, a Bernoulli trial is repeated until success is obtained and then stopped in geometric distribution.

A geometric distribution is a discrete probability distribution that indicates the likelihood of achieving one's first success after a series of failures. The number of attempts in a geometric distribution can go on indefinitely until the first success is achieved. Geometric distributions are probability distributions that are based on three key assumptions.

- **The trials that are being undertaken are self – contained.**
- **Each trial may only have one of two outcomes: success or failure.**
- **For each trial, the success probability, represented by p is the same.**

Geometric Distribution Formula

P(X = k) = $(1-p)^k p$

Mean	$\mu = E(X) = \frac{1}{p}$
Variance	$\sigma = V(X) = \frac{(1-p)}{p^2}$

Where:

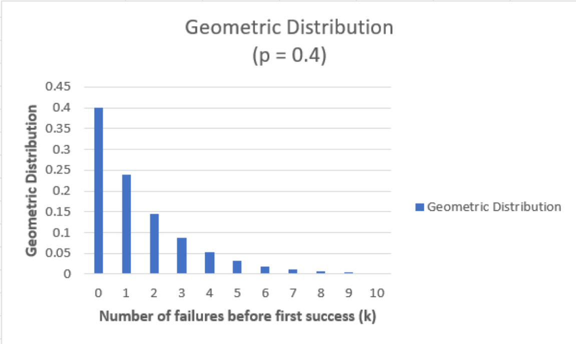
- **k: number of failures before first success**
- **p: probability of success on each trial**

The chance of a trial's success is denoted by p, whereas the likelihood of failure is denoted by q, q = 1 – p in this case. X ~ G (p) represents a discrete random variable, X with a geometric probability distribution.

Excel implementation

Probability = $(1-p)^k p$

Probability of success on given trial (p)	0.4
Number of failures before first success (k)	Geometric Distribution
0	0.4
1	0.24
2	0.144
3	0.0864
4	0.05184
5	0.031104
6	0.0186624
7	0.01119744
8	0.006718464
9	0.004031078
10	0.002418647



5. Plotting and fitting of Uniform distribution and graphical representation of probabilities

A uniform distribution is a distribution that has constant probability due to equally likely occurring events. It is also known as rectangular distribution (contribution uniform distribution). It has two parameters a and b: a = minimum and b = maximum. The distribution is written as U (a, b).

A uniform distribution is a type of probability distribution where every possible outcome has an equal probability of occurring. This means that all values within a given range are equally likely to be observed.

Uniform Distribution Formula

The probability density function (PDF) of a continuous uniform distribution defines the probability of a random variable falling within a particular interval. For a continuous uniform distribution over the interval [a, b].

$$f(x) = \frac{1}{b-a} for\ a \ \leq x \ \leq b$$

$$Mean\ \mu = \frac{a+b}{2}$$

$$Variance\ \sigma^2 = \frac{(b-a)^2}{12}$$

Excel Implementation

$$P = \frac{(x_{2-}\,x_1)}{(b-a)}$$

For calculating probability, we need:

- a: minimum value in the distribution
- b: maximum value in the distribution

- x_1: **minimum value you're interested in**
- x_2: **maximum value you're interested in**

a	14			
b	28			
		case 1	case 2	case 3
x1		16	19	23
x2		18	22	24
Uniform Distribution		3.428571429	7.321428571	9.178571429

6. Plotting and fitting of Exponential Distribution and graphical representation of probabilities

The support (set of values the Random Variable can take) of an Exponential Random Variable is the set of all positive real numbers. Suppose we are posed with the question – How much time do we need to wait before a given event occurs? The answer to this question can be given in probabilistic terms if we model the given problem using the Exponential Distribution. Since the time we need to wait is unknown, we can think of it as a Random Variable, If the probability of the event happening in a given interval is proportional to the length of the interval, then the random variable has an exponential distribution.

This distribution can be used to solve following type of real life problems:

- How long does a shop owner need to wait until a customer enter a shop.
- How long will a battery continue to work before it dies.
- How long will a computer continue to work before it breaks down.

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

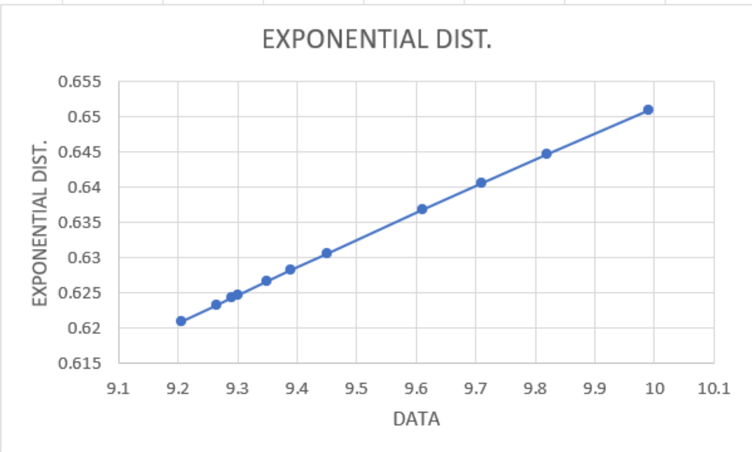
Here λ is the rate parameter and it effects on the density function.

e is a constant roughly equal to 2.718.

<u>Excel Implementation</u>

EXPON.DIST(X, lambda, FALSE)

DATA	EXPONENTIAL DIST.
9.206	0.620979545
9.264	0.623289155
9.29	0.624319924
9.3	0.624715622
9.35	0.62668787
9.39	0.628258203
9.45	0.630601326
9.61	0.636777702
9.71	0.640585385
9.82	0.644727749
9.99	0.651035876
MEAN	9.489090909
λ	0.105384



7. Plotting and fitting of Normal Distribution and graphical representation of probabilities.

Normal Distribution

A normal distribution is a continuous probability distribution defined by the probability density function (PDF).

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

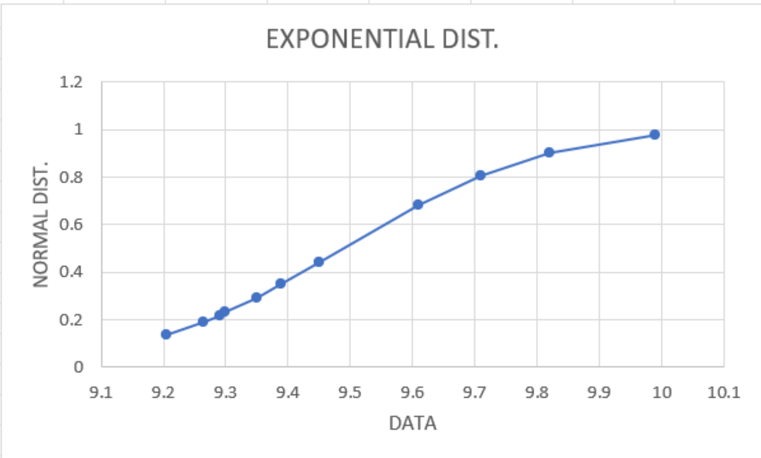
Where:

- μ **is the mean of the distribution.**
- σ **is the standard deviation.**
- e **is the base of the natural logarithm (approx. 2.718)**
- π **is the constant.**
- x **is the variable.**

Excel Implementation

=NORM.DIST(x, mean, standard deviation, TRUE)

DATA	EXPONENTIAL DIST.
9.206	0.135376919
9.264	0.190600166
9.29	0.219306596
9.3	0.230976889
9.35	0.294214155
9.39	0.349933414
9.45	0.439562678
9.61	0.680958591
9.71	0.804945019
9.82	0.901015336
9.99	0.974335647
MEAN	9.489090909
ST.DEV	0.257045



8. Calculation of cumulative functions for Exponential and Normal Distribution.

CDF of Exponential Distribution

For an exponential distribution with rate parameter λ, the CDF is:

$$F\,(x) = P(X \le x) = 1 - e^{-\lambda x}, for\ x \ge 0$$

CDF of Normal Distribution

For a normal distribution with mean μ and standard deviation σ, the CDF is given by:

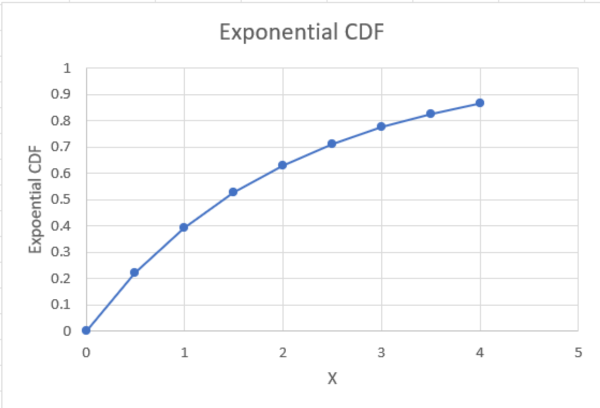
$$F(x) = P(X \le x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

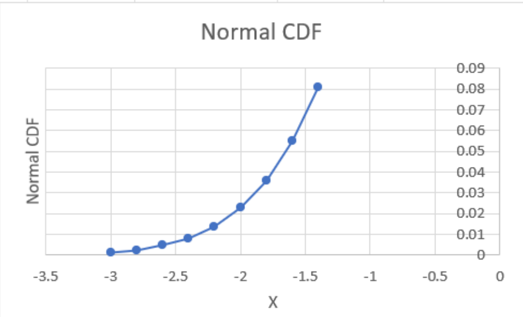
Excel Implementation

Exponential Distribution: =EXPON.DIST(C5, 0.5, TRUE)

Normal Distribution: =NORM.DIST(F5, 0, 1, TRUE)

X	Exponential CDF	X	Normal CDF
0	0	-3	0.001349898
0.5	0.221199217	-2.8	0.00255513
1	0.39346934	-2.6	0.004661188
1.5	0.527633447	-2.4	0.008197536
2	0.632120559	-2.2	0.013903448
2.5	0.713495203	-2	0.022750132
3	0.77686984	-1.8	0.035930319
3.5	0.826226057	-1.6	0.054799292
4	0.864664717	-1.4	0.080756659





9. Given data from two distributions, find the distance between the distributions.

Euclidean Distance:

The Euclidean Distance between two points $P = (p_1, p_2, \dots, p_n)$ **and** $Q = (q_1, q_2, \dots, q_n)$ **is given by the formula:**

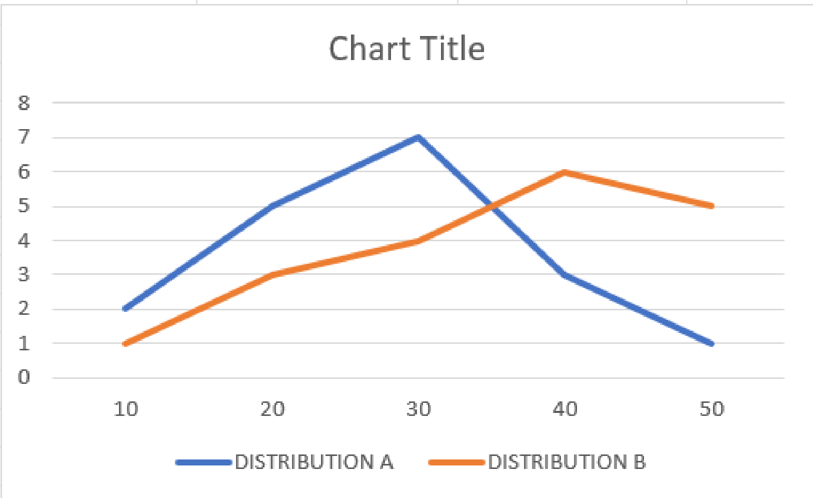
$$d(P, Q) = \sqrt{\sum_{i=1}^{n} (p_1 - q_i)^2}$$

For distributions, we consider the probability density function (PDF's) of each distribution over a range of values, say $x_1, x_2, \dots, x_n$, **and computer the distance between the value of their PDF's.**

Excel Implementation

=SQRT(SUM(E4:E8))

SCORE	DISTRIBUTION A	DISTRIBUTION B	(C - E) ^ 2
10	2	1	1
20	5	3	4
30	7	4	9
40	3	6	9
50	1	5	16
EUCLIDEAN DISTANCE	6.244997998		



10. Application problems based on Binomial Distribution.

A factory produces bulbs, and 5% of them are defective. If a quality control inspector selects 20 bulbs at random, what is the probability that exactly 2 bulbs are defective?

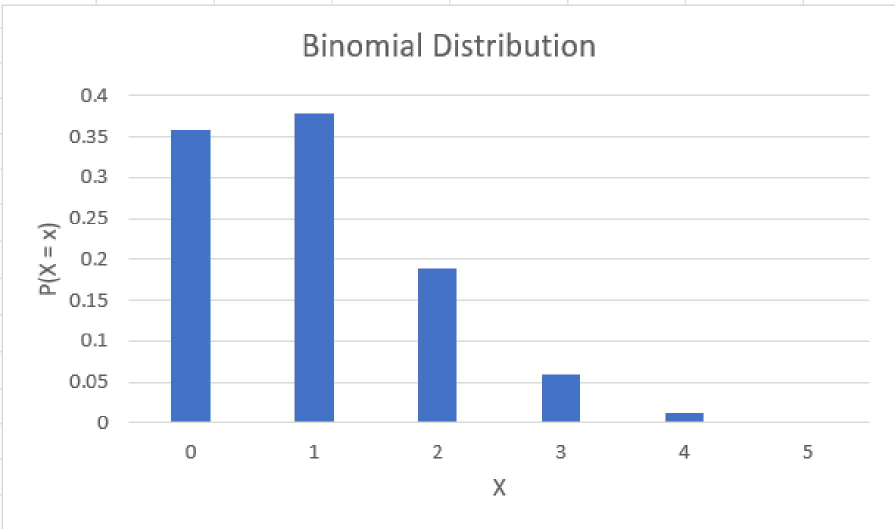
- $n = 20$ **(number of trials)**
- $p = 0.05$ **(probability of success – defective bulbs)**
- $q = 0.95$ **(probability of non – defective bulbs)**
- $x = 0 \; to \; 5$ **(we'll calculate 0 to 5 bulbs)**

<u>Excel Implementation</u>

=BINOM.DIST(C4, D13, D12, FALSE)

X	P(X = x)
0	0.358485922
1	0.377353603
2	0.188676801
3	0.059582148
4	0.013327586
5	0.002244646

p	0.05
q	0.05
n	20



11. Application problems based on the Poisson Distribution.

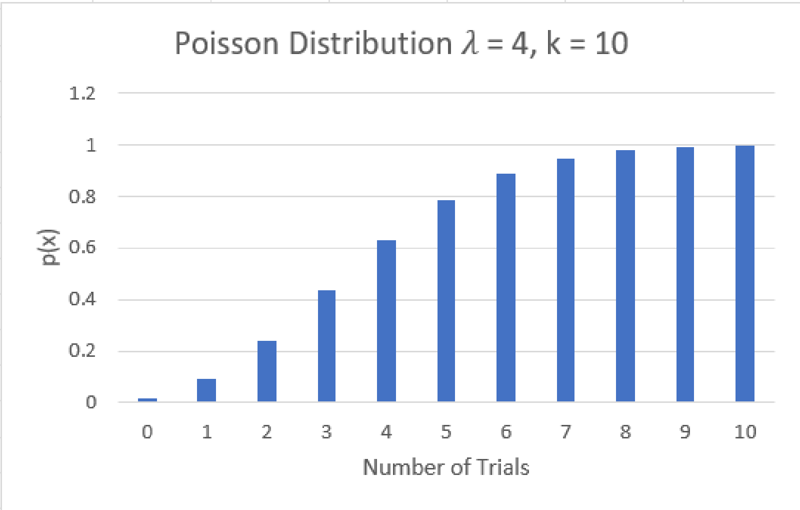
A call centre receives an average of 4 calls per minute. What is the probability of receiving 0 to 10 calls in a minute?

- $\Lambda = 4$ (average rate of calls)

<u>Excel Implementation</u>

=POISSON.DIST(C4, 4, TRUE)

X	P(X = x)
0	0.018315639
1	0.091578194
2	0.238103306
3	0.43347012
4	0.628836935
5	0.785130387
6	0.889326022
7	0.948866384
8	0.978636566
9	0.991867757
10	0.997160234



12. Application problems based on the Normal Distribution.

The average daily demand for a product n a store is 500 units, with standard deviation of 50 units. The demand follows a normal distribution. Create a graph showing the normal distribution curve for daily demand from 350 – 400 units.

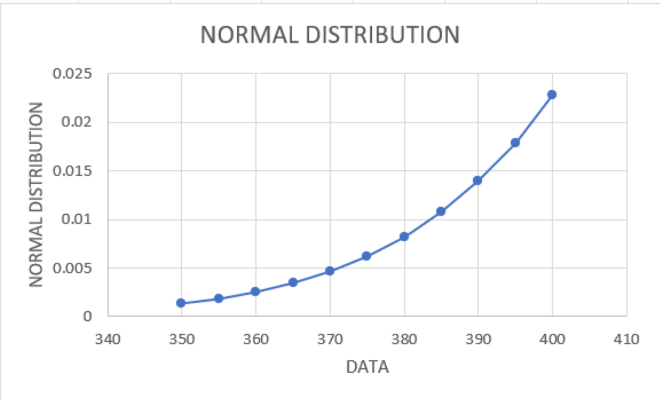
- Mean (μ) = 500
- Standard Deviation (σ) = 50

<u>Excel Implementation</u>

=NORM.DIST(E4, E17, E16, TRUE)

DATA	NORMAL DISTRIBUTION
350	0.001349898
355	0.001865813
360	0.00255513
365	0.003466974
370	0.004661188
375	0.006209665
380	0.008197536
385	0.01072411
390	0.013903448
395	0.017864421
400	0.022750132

stdv (σ)	50
mean (μ)	500



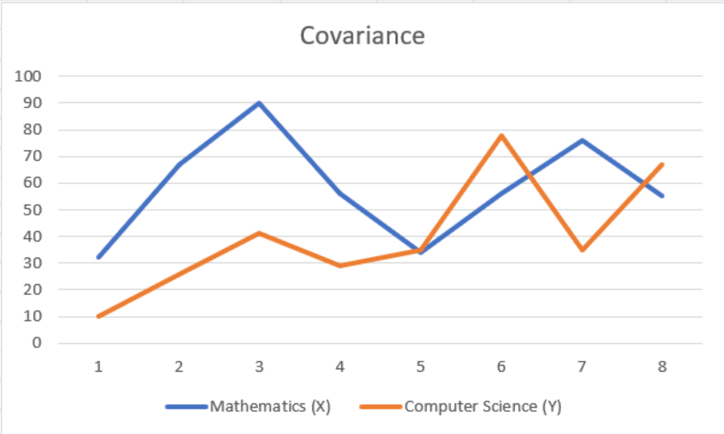
13. Presentation of Bivariate Data through scatter – plot diagrams and calculation of covariance.

The following data shows the marks obtained by 8 students in two different subjects: Mathematics (X) and Computer Science (Y). Calculate the covariance between the marks of Mathematics and Computer Science.

<u>Excel Implementation</u>

=COVARIANCE.S(D6:13, E6:E13)

Mathematics (X)	Computer Science (Y)
32	10
67	26
90	41
56	29
34	35
56	78
76	35
55	67
Covariance	82.96428571



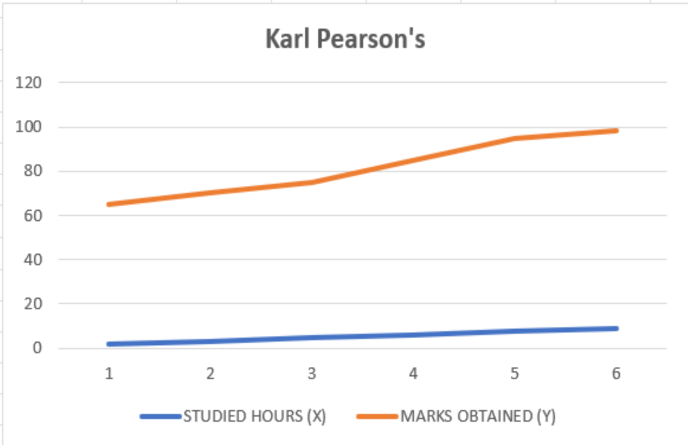
14. Calculation of Karl Pearson's correlation coefficients.

The following table shows the number of hours studied and marks obtained by six students in a test. Now calculate the Karl Pearson's Correlation Coefficient between hours studied and marks obtained.

Excel Implementation

=PEARSON(C4:C9, D4:D9)

STUDENTS	STUDIED HOURS (X)	MARKS OBTAINED (Y)
A	2	65
B	3	70
C	5	75
D	6	85
E	8	95
F	9	98
Karl Pearson's Coefficient	0.988828105	



15. To find the correlation coefficient for a bivariate frequency distribution.

A teacher recorded the number of students scoring combinations of marks in two subjects: Math (X) and Science (Y). Calculate the correlation coefficient for a bivariate frequency distribution.

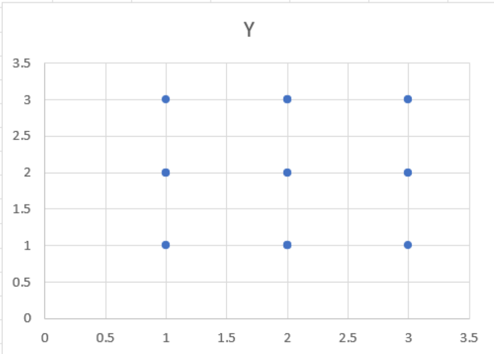
<u>Excel Implementation</u>

=CORREL(G4:G23, H4:H23)

X/Y	1	2	3
1	2	3	1
2	4	2	3
3	1	2	2

Correlation Coefficient	0.163754283

X	Y
1	1
1	1
1	2
1	2
1	2
1	3
2	1
2	1
2	1
2	1
2	2
2	2
2	3
2	3
2	3
3	1
3	2
3	2
3	3
3	3



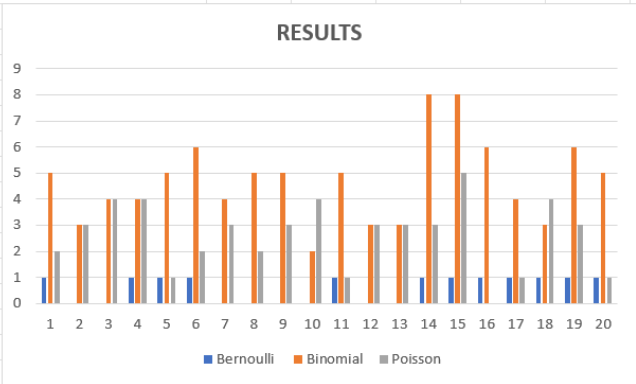
16. Generating random numbers from discrete (Bernoulli, Binomial, Poisson) distributions.

Formula to generate random numbers from:

- **Bernoulli Distribution: =IF(RAND() <= 0.6, 1, 0)**
- **Binomial Distribution: =BINOM.INV(10, 0.5, RAND())**
- **Poisson Distribution: =POISSON.INV(RAND(), 3)**

Trial	Bernoulli	Binomial	Poisson
1	1	5	2
2	0	3	3
3	0	4	4
4	1	4	4
5	1	5	1
6	1	6	2
7	0	4	3
8	0	5	2
9	0	5	3
10	0	2	4
11	1	5	1
12	0	3	3
13	0	3	3
14	1	8	3
15	1	8	5
16	1	6	0
17	1	4	1
18	1	3	4
19	1	6	3
20	1	5	1

number of trial (n)	20
probability of success (p)	0.5
mean (λ)	3

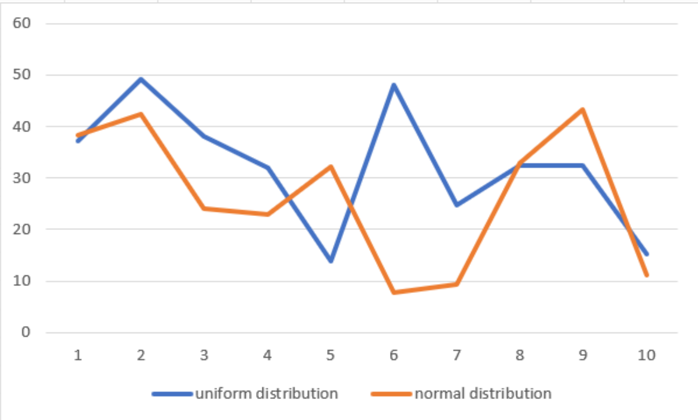


17. Generating random numbers from continuous (Uniform, Normal distributions).

Formula to find the random numbers from continuous:

- **Uniform Distribution: =RAND()*(upper limit – lower limit) + lower limit**
- **Normal Distribution: =NORMINV(RAND(), mean, standard deviation)**

uniform distribution	normal distribution
37.16024649	38.36753464
49.23606602	42.35072339
38.05476137	23.95025664
32.02511762	23.02666085
13.81139165	32.23947051
48.15912599	7.789431721
24.68380279	9.222263716
32.46960207	32.94017155
32.51325128	43.42330944
15.24330048	11.18098877



18. Find the entropy from the given data set.

Given the weather conditions and the corresponding decision to play or not, calculate the entropy of the "Play" decision.

<u>Excel Implementation</u>

=(p*log2(p))*2

Weather	Play
sunny	0
sunny	0
overcast	1
rainy	1
rainy	0
rainy	1
overcast	1
sunny	0
sunny	1
rainy	1

yes = 1	
no = 0	

Class	Count	(p)	p*log2(p)	Entropy
yes	5	0.5	-0.5	
no	5	0.5	-0.5	1
total	10