PROBABILITY FOR COMPUTING

PRACTICAL FILE

RAMANUJAN COLLEGE

UNIVERSITY OF DELHI

SUBMITTED BY :

NAME : OMPRAKASH
ROLL NO. : 24570043
EXAMINATION ROLL NO : 24020570032
CLASS : BSC(H) COMPUTER SCIENCE SEM 1

SUBMITTED TO:

DR. AAKASH
ASSISTANT PROFESSOR
(OPERATIONAL RESEARCH),
DEPARTMENT OF COMPUTER SCIENCE,
RAMANUJAN COLLEGE,
UNIVERSITY OF DELHI,
CR PARK MAIN ROAD,BLOCK H , KALKAJI
NEW DELHI 110019

Acknowledgement

I would like to take this opportunity to acknowledge everyone who has helped us in every stage of this project.

I am deeply indebted to my mathematics Professor, Dr Aakash, assistant professor, Department of computer science, ramanujan college, Delhi University for his guidance and suggestions in completing this project. The completion of this project was possible under his guidance and support.

I am also very thankful to my parents and my friends who have boosted me up morally with their continuous support.

At last but not least, I am very thankful to God almighty for showering his blessings upon me.

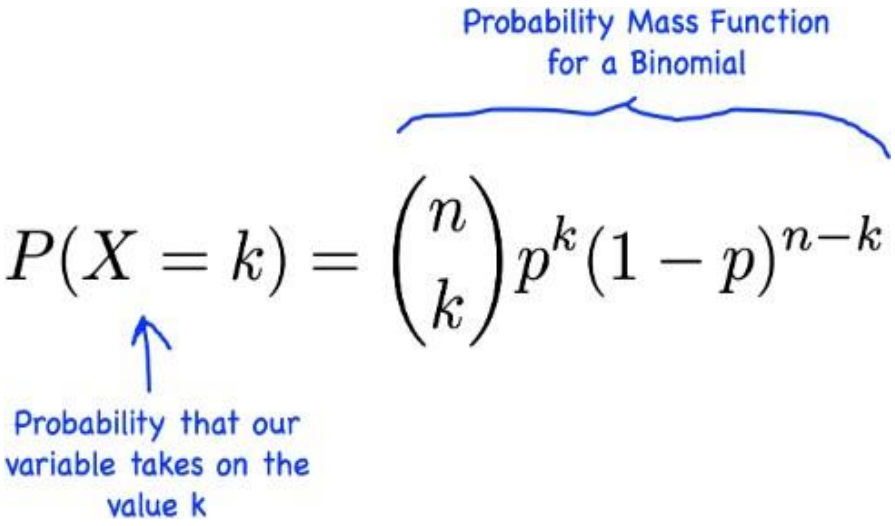
Binomial distribution

Binomial Distribution in Probability gives information about only two types of possible outcomes i.e. Success or Failure. Binomial Probability Distribution is a discrete probability distribution used for the events that give results in ***'Yes or No' or 'Success or Failure'.***

the probability of success (usually denoted as "p") and the probability of failure (usually denoted as "q") is constant for each trial.

Binomial Distribution Formula

The Binomial Distribution Formula which is used to calculate the probability, for a random variable X = 0, 1, 2, 3,….,n is given as



Mean $\mu = n \cdot p$

Variance $\sigma^2 = n \cdot p \cdot q$

Std. Dev. $\sigma = \sqrt{n \cdot p \cdot q}$

, k = 0, 1, 2, 3….

Where ,

- *p is probability of success*
- *q is probability of failure and q = 1 – p*
- *p, q > 0 such that p + q = 1*
- *n is the number of independent trials*
- *k is the number of success*
- *nCk is the number of ways to obtain k success in n trials.*

Understanding Binomial Distribution Symmetry

- *A binomial distribution is symmetric if its success probability, p, is equal to 0.5 (i.e., the probability of failure, q, is also 0.5). When p = 0.5, the distribution is said to be balanced, which means that the chances of success and failure are equal. In this case, the given binomial distribution has p = 0.5, which means that it is* ***symmetric****. The shape of the probability distribution will be* ***symmetric and not skewed***.

Determining the Skewness of the Distribution

- *Since we know that the distribution is symmetric when p = 0.5, it follows that the **skewness is zero**. Skewness is a measure of the asymmetry of a distribution. A distribution is **positively skewed** (skewed right) when there are more smaller values, and **negatively skewed** (skewed left) when there are more larger values. A symmetric distribution has a skewness of zero because there is an equal balance of larger and smaller values in the distribution. This means that the shape of the probability distribution is not skewed to the left or the right. Therefore, the shape of the probability distribution in this case is symmetric due to p = 0.5.*

how to implement in excel

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

= BINOM.DIST(k,n,p,FALSE)

The BINOM.DIST utilizes the accompanying contentions:

1. ***Number_s** (required argument) – This is the number of successes in trials.*
2. ***Trials** (required argument) – This is the number of independent trials. It must be greater than or equal to 0.*
3. ***Probability_s** (required argument) – This is the probability of success in each trial.*
4. ***Cumulative** (required argument) – This is a logical value that determines the form of the function. It can either be:*
    1. *TRUE – Uses the cumulative distribution function.*
    2. *FALSE – Uses the probability mass function*

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

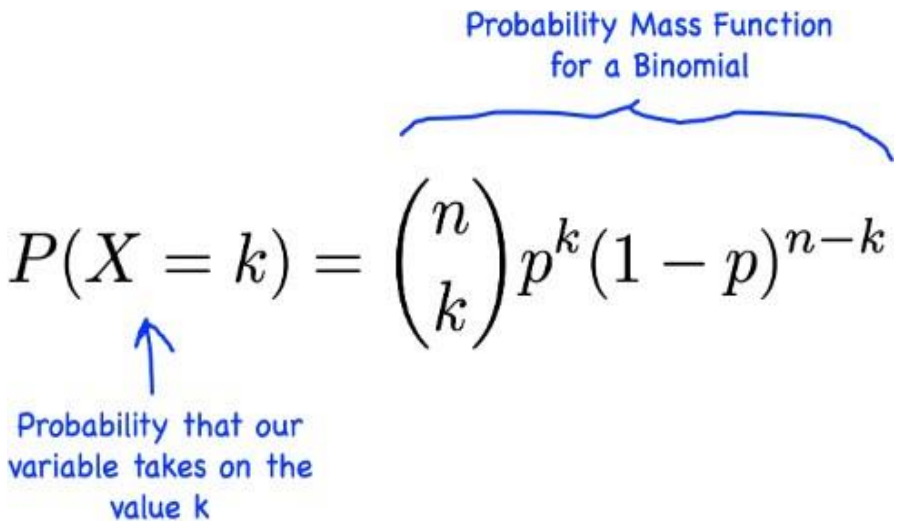
Binomial distribution

Binomial Distribution in Probability gives information about only two types of possible outcomes i.e. Success or Failure. Binomial Probability Distribution is a discrete probability distribution used for the events that give results in ***'Yes or No' or 'Success or Failure'.***

the probability of success (usually denoted as "p") and the probability of failure (usually denoted as "q") is constant for each trial.

Binomial Distribution Formula

The Binomial Distribution Formula which is used to calculate the probability, for a random variable X = 0, 1, 2, 3,….,n is given as



Mean $\mu = n \cdot p$

Variance $\sigma^2 = n \cdot p \cdot q$

Std. Dev. $\sigma = \sqrt{n \cdot p \cdot q}$

, k = 0, 1, 2, 3….

Where ,

- *p is probability of success*

- *q is probability of failure and q = 1 – p*
- *p, q > 0 such that p + q = 1*
- *n is the number of independent trials*
- *k is the number of success*
- *nCk is the number of ways to obtain k success in n trials.*

Understanding Binomial Distribution Symmetry

- *A binomial distribution is symmetric if its success probability, p, is equal to 0.5 (i.e., the probability of failure, q, is also 0.5). When p = 0.5, the distribution is said to be balanced, which means that the chances of success and failure are equal. In this case, the given binomial distribution has p = 0.5, which means that it is **symmetric**. The shape of the probability distribution will be **symmetric and not skewed**.*

Determining the Skewness of the Distribution

- *Since we know that the distribution is symmetric when p = 0.5, it follows that the **skewness is zero**. Skewness is a measure of the asymmetry of a distribution. A distribution is **positively skewed** (skewed right) when there are more smaller values, and **negatively skewed** (skewed left) when there are more larger values. A symmetric distribution has a skewness of zero because there is an equal balance of larger and smaller values in the distribution. This means that the shape of the probability distribution is not skewed to the left or the right. Therefore, the shape of the probability distribution in this case is symmetric due to p = 0.5.*

how to implement in excel

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

= BINOM.DIST(k,n,p,FALSE)

The BINOM.DIST utilizes the accompanying contentions:

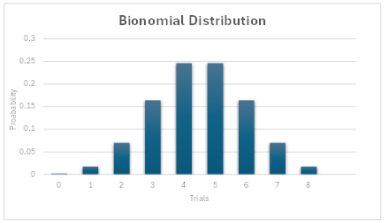
5. ***Number_s** (required argument) – This is the number of successes in trials.*
6. ***Trials** (required argument) – This is the number of independent trials. It must be greater than or equal to 0.*
7. ***Probability_s** (required argument) – This is the probability of success in each trial.*
8. ***Cumulative** (required argument) – This is a logical value that determines the form of the function. It can either be:*
    1. *TRUE – Uses the cumulative distribution function.*
    2. *FALSE – Uses the probability mass function*

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

Suppose that a x, y, and z electronic company produce both wired and wireless mouse . The product mix is 50% of the mouse are wired and 50% are wireless .If we choose 10 mouse at random and choosing wireless mouse is defined as a success. The probability distribution of the number of success during these 8 trials with probability p=0.5 then plot the graph of this probability in excel ?

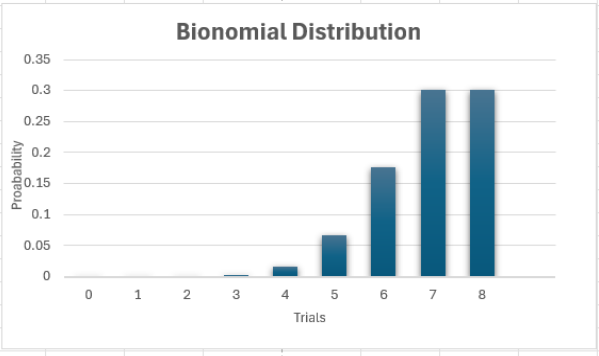
Number of Success(X=K)	9
Proabability of success P(X=K)	0.5

Trials	Binomial Distribution of trials
0	0.001953125
1	0.017578125
2	0.0703125
3	0.1640625
4	0.24609375
5	0.24609375
6	0.1640625
7	0.0703125
8	0.017578125



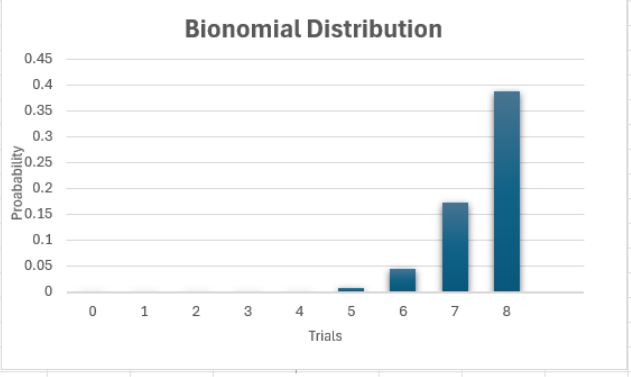
Number of Success(X=K)	9
Proabability of success P(X=K)	0.8

Trials	Binomial Distribution of trials
0	0.000000512
1	0.000018432
2	0.000294912
3	0.002752512
4	0.016515072
5	0.066060288
6	0.176160768
7	0.301989888
8	0.301989888



Number of Success(X=K)	9
Proabability of success P(X=K)	0.9

Trials	Binomial Distribution of trials
0	1E-09
1	0.000000081
2	0.000002916
3	0.000061236
4	0.000826686
5	0.007440174
6	0.044641044
7	0.172186884
8	0.387420489



Multinomial distribution

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* $i = 1, \ldots, k$ *and* $\sum_{i=1}^{k} p_i = 1$*), 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* $X = (X_1, X_2, \ldots, X_k)$ *follows a multinomial distribution with parameters* **n** *and* ***p****, where* $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$*.*

Multinomial Distribution Formula

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

Mean	$E\{X_i\} = np_i$
Variance	$\mathrm{Var}(X_i) = np_i(1 - p_i)$ $\mathrm{Cov}(X_i, X_j) = -np_ip_j \quad (i \neq j)$

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

how to implement in excel

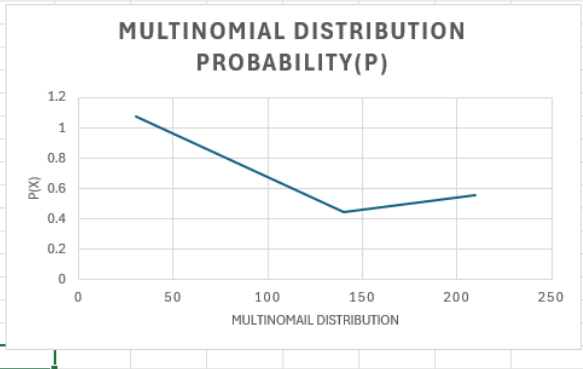
Multinomial = MULTINOMIAL(X1,X2,X3)

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

2. Plotting and fitting of Multinomial distribution and graphical representation of probabilities.

Suppose that a bag contain 8 balls 3 red, 1 green and 4 blue to reach in a bag to pull ball at random and then pull the ball back and pull out another ball .experiment is repeated at total of 10 time . What is the probability outcome will result in 4 red and 6 blues?

		1st case	2nd case	3rd case
Red	X1	5	2	2
Green	X2	0	1	1
Blue	X3	6	5	3
P_Red	P1	0.3	0.445	0.324
P_green	P2	0.15	0.189	0.42
P_blue	P3	0.89	0.56	0.74
	Multinomia	42	21	10
	P1^X1	0.00243	0.198025	0.104976
	P2^X2	7.59375E-05	0.189	0.42
	P3^X3	0.558405945	0.05507318	0.5476
	Probability	0.560911882	0.44209818	1.072576
	Multinmial	multinomial distribution proabability(P)		
	210	0.560911882		
	140	0.442098178		
	30	1.072576		



Poisson distribution

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

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

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 (approximately 2.71828)*
- ***λ*** *mean number of success that occur during a specific interval,* ***λ = np***
- ***k*** *is the number of success*

how to implement in excel

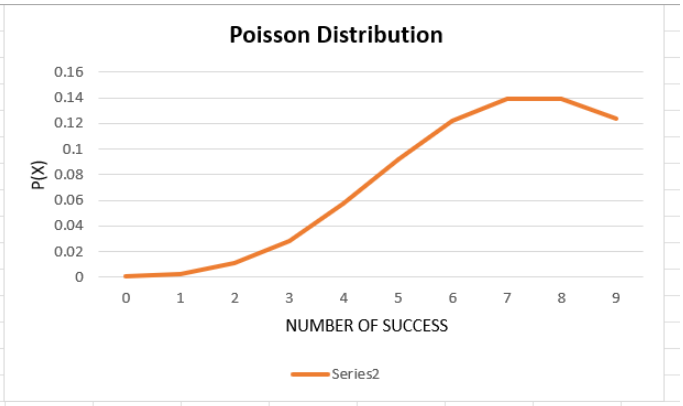
POISSON.DIST(number_s,average,cumulative)

POISSON.DIST(k, λ ,FALSE)

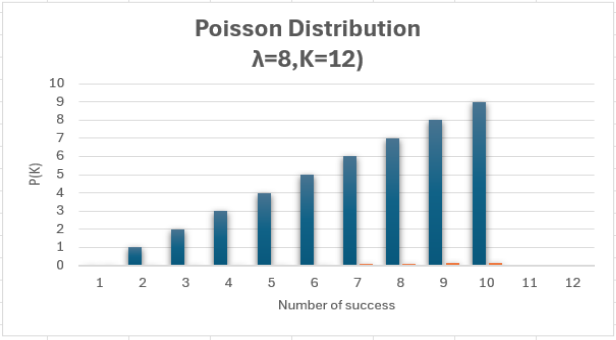
3. Plotting and fitting of Poisson distribution and graphical representation of probabilities.

An electronic store sells on average 8 Desktop in a week. Assuming that purchases are as described above then what is the probability that the store will have turn away potential buyers before the end if the stock 9 computers ? how many computers should the store stock in order to make sure that it has 99% probability of being able to supply a week's demand ?

λ(Mean numbet of success)	8
K(Number of Success)	Poisson Distribution
0	0.000335463
1	0.002683701
2	0.010734804
3	0.028626144
4	0.057252288
5	0.091603662
6	0.122138215
7	0.139586532
8	0.139586532
9	0.124076917



λ(Mean numbet 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



Geometric distribution

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:	$\sigma^2 = V(X) = \frac{(1-p)}{p^2}$

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.

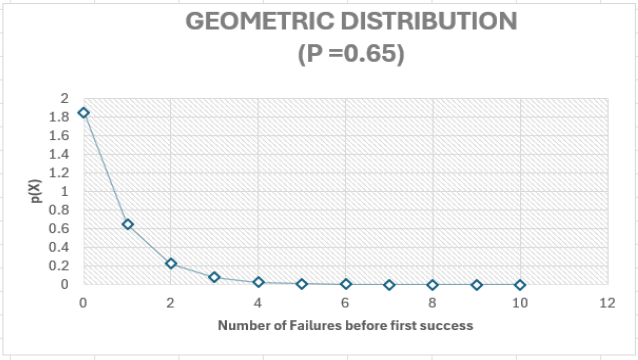
how to implement in excel

Probability = (1-p)^k*p

4. Plotting and fitting of Geometric distribution and graphical representation of probabilities.

Suppose a programmer is waiting outside a PM office to take his views on artificial intelligence like they support AI or not .The probability that a PM supports the AI is p = 0.7 . what is the probability that the fourth PM , the programmer talk to is the first PM to support AI ?

Prob. Of Success on given trial(P)	0.65
Number of failuires before first success	Gemotric Distribution
0	1.857142857
1	0.65
2	0.2275
3	0.079625
4	0.02786875
5	0.009754063
6	0.003413922
7	0.001194873
8	0.000418205
9	0.000146372
10	5.12302E-05



Uniform Distribution Function

A uniform distribution is a distribution that has constant probability due to equally likely occurring events. It is also known as rectangular distribution (continuous 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].

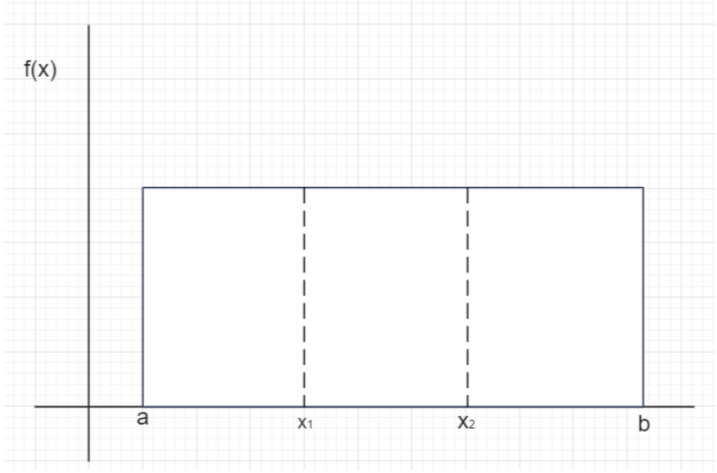
$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

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

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

how to implement in excel

$$P = (x_2 - x_1) / (b - a)$$



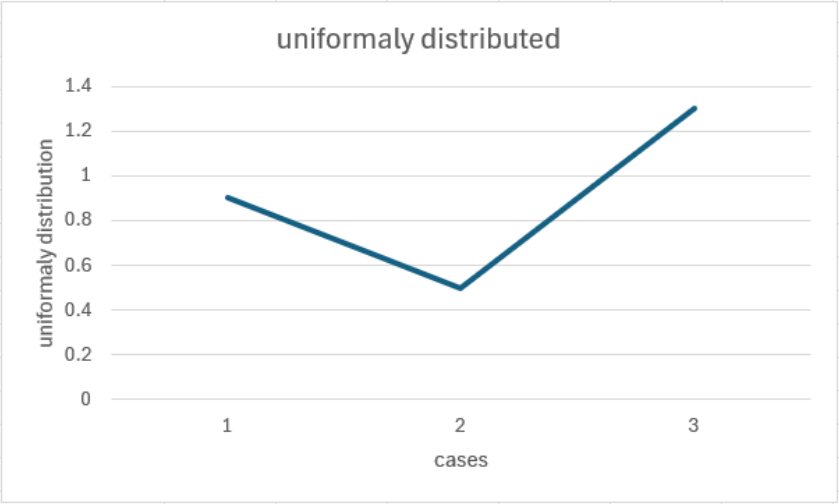
For calculating probability, we need:

1. *a: minimum value in the distribution*
2. *b: maximum value in the distribution*
3. *x_1: the minimum value you're interested in*
4. *x_2: the maximum value you're interested in*

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

A content search on x,y,and z search engine takes 15 seconds and it will show up the result every time in 25 seconds. If you will search on x , y , z search engine then what is the probability that a content will show up in 16-18 seconds ?

a	15			
b	25			
		case1	case2	case3
x1		16	17	13
x2		25	22	26
uniformal distributed		0.9	0.5	1.3



Exponential random variable

*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. The support (set of values the Random Variable can take) of an Exponential Random Variable is the set of all positive real numbers.*

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 breakdown.*

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

$$E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Here $\boldsymbol{\lambda}$ is the rate parameter and its effects on the density function .

e is a constant roughly equal to 2.718

How to Implement in excel

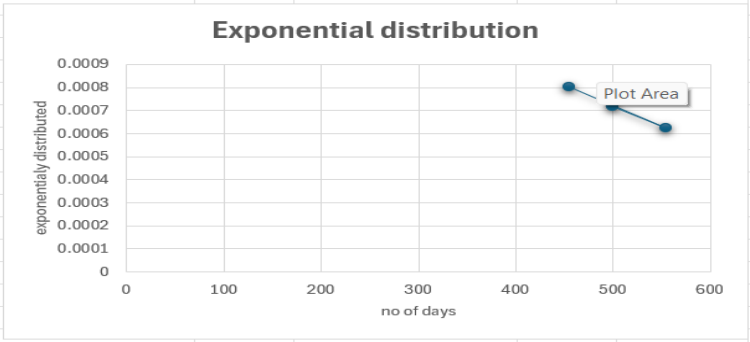
EXPON.DIST(X,lambda,cumulative)

EXPON.DIST(X,lambda,FALSE)

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

Suppose a big building constructing takes 400 days on average after an update occur find the probability that it will construct in more than 500 days and so on as in data ?

μ	400		x	455	555	500
λ	0.0025					
		Exponential distribution		0.000801549	0.000624247	0.000716262
		455	0.000802			
		555	0.000624			
		500	0.000716			



NORMAL DISTRIBUTION

We define Normal Distribution as the probability density function of any continuous random variable for any given system. Now for defining Normal Distribution suppose we take f(x) as the probability density function for any random variable X.

f(x) ≥ 0 ∀ x ϵ (−∞,+∞),

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

where,

- *x is **Random Variable***
- *μ is **Mean***
- *σ is **Standard Deviation***

Properties of Normal Distribution

- *For normal distribution of data, mean, median, and mode are equal, (i.e., Mean = Median = Mode).*
- *Total area under the normal distribution curve is equal to 1.*
- *Normally distributed curve is symmetric at the center along the mean.*
- *In a normally distributed curve, there is exactly half value to the right of the central and exactly half value to the right side of the central value.*
- *Normal distribution is defined using the values of the mean and standard deviation.*
- *Normal distribution curve is a Unimodal Curve, i.e. a curve with only one peak*

how to implement in excel

1. Input your data set into an Excel spreadsheet

2. Find the mean of your data set

=AVERAGE(cell range)

- *"cell range" is a required component and the range of cells where your data exists, such as cells A1 through A64. You can write this in the function as A1:A64.*

3. Find the standard deviation of your data set

=STDEV(cell range)

4. Select a value for the distribution

5. ***Type the NORM.DIST function and fill***

NORM.DIST(x,mean,standarddeviation,cumulative)

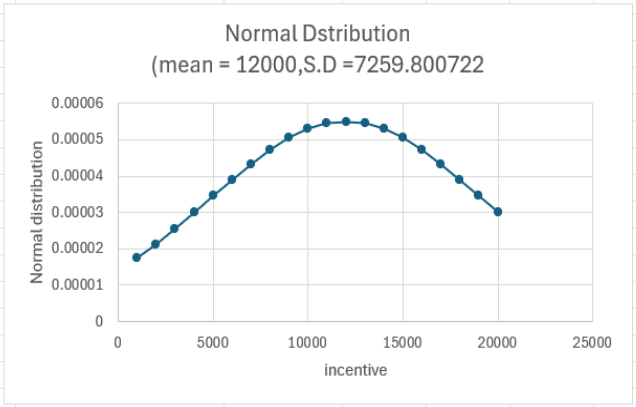
NORM.DIST(x,mean,standarddeviation,FALSE)

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

Plot the normal graph according to the data of employee and incentive as shown below .

Employe	Incentive	Normaldistribution
EMP1	1000	1.74364E-05
EMP2	2000	2.12803E-05
EPM3	3000	2.54835E-05
EMP4	4000	2.99434E-05
EMP5	5000	3.45225E-05
EMP6	6000	3.90537E-05
EMP7	7000	4.33494E-05
EMP8	8000	4.72133E-05
EMP9	9000	5.04551E-05
EMP10	10000	5.2906E-05
EMP11	11000	5.44334E-05
EMP12	12000	5.49522E-05
EMP13	13000	5.44334E-05
EMP14	14000	5.2906E-05
EMP15	15000	5.04551E-05
EMP16	16000	4.72133E-05
EMP17	17000	4.33494E-05
EMP18	18000	3.90537E-05
EMP19	19000	3.45225E-05
EMP20	20000	2.99434E-05

MEAN	12000
STANDARD DISTRIBUTION	7259.800722



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

Calculate normal graph and exponential cumulative distribution according to the data of employee ,incentive and number of days as shown below .

	Normal Distribution					EXPONENTIAL distribution			
Emplo	incentive	Normal distribution				no of days	exponential distribution		
emp1	1000	0.093224583				300	0.527633447		
emp2	2000	0.160875161				310	0.539296219		
emp3	3000	0.2544414				395	0.627493205		
emp4	4000	0.370590749	mean	5000		410	0.641203535		
emp5	5000	0.5	standard deviation	3027.65025		460	0.683363231	λ	0.0025
emp6	6000	0.629409251				520	0.727468207		
emp7	7000	0.7455586				550	0.747160404		
emp8	8000	0.839124839				580	0.765429712		
emp9	9000	0.906775417				598	0.775751395		
emp10	10000	0.950676201				600	0.77686984		

Euclidean distance

Euclidean distance is the distance between two real distinct value .It is calculate by the square root of the sum of the squared difference elements in two vectors.

$$\textbf{Euclidean Distance} = |X - Y| = \sqrt{\sum_{i=1}^{i=n} (x_i - y_i)^2}$$

X: Array or vector X

Y: Array or vector Y

x_i: Values of horizontal axis in the coordinate plane

y_i: Values of vertical axis in the coordinate plane

n: Number of observations

how to implement in excel

= SORT(SUM X MYZ(array_X,array_Y))

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

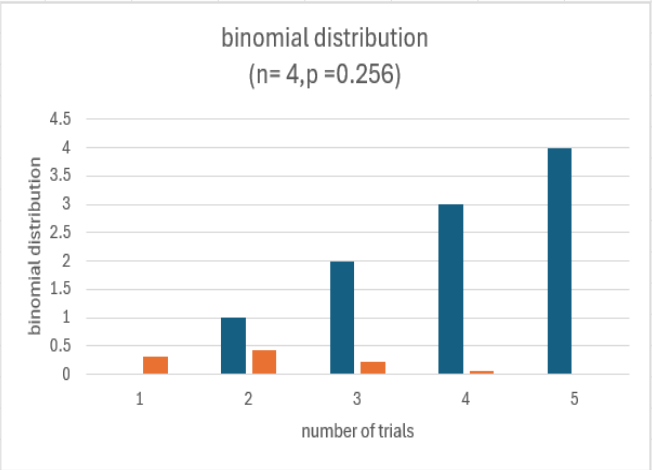
In our case , we take data from the binomial and poisson distribution to find the eucliean distance between them.

trials	bionmial distribution	poisson distribution		distance			
1	0.165894476	0.297016028		0.149195905			
2	0.269155324	0.071283847		0.132003044			
3	0.258779722	0.011405415		0.092849923		probability	0.265
4	0.163277682	0.00136865		0.031655875		k	10
5	0.070642589	0.00013139		0.005441341		λ	0.48
6	0.021224814	1.05112E-05		0.000469512			
7	0.004372848	7.2077E-07		1.9465E-05			
8	0.000591227	4.32462E-08		3.49498E-07			

10. Application problems based on the Binomial distribution.

Antibiotics occasionally cause nausea as a side effect. A major drug company has developed a new antibiotic called Phe-Mycin. The company claims that, at most, 10 percent of all patients treated with Phe-Mycin would experience nausea as a side effect of taking the drug. Suppose that we randomly select n = 7 patients and treat them with Phe-Mycin. Each patient will either experience nausea (which we arbitrarily call a success) or will not experience nausea (a failure). We will assume that p, the true probability that a patient will experience nausea as a side effect, is .0.4, the maximum value of p claimed by the drug company. Furthermore, it is reasonable to assume that patients' reactions to the drug would be independent of each other. Let x denote the number of patients among the four who will experience nausea as a side effect. It follows that x is a binomial random variable, which can take on any of the potential values 0, 1, 2, 3, or 4. That is, anywhere between none of the patients and all four of the patients could potentially experience nausea as a side effect Suppose that we wish to investigate whether p, the probability that a patient will experience nausea as a side effect of taking Phe- Mycin, is greater than0.4, the maximum value of p claimed by the drug company. This assessment will be made by assuming, for the sake of argument, that p equals .0.4, and by using sample information to weigh the evidence against this assumption and in favor of the conclusion that p is greater than 0.4. Suppose that when a sample of n=4 randomly selected patients is treated with Phe-Mycin, three of the four patients experience nausea. Because the fraction of patients in the sample that experience nausea is 3/4 = .75, which is far greater than 0.4, we have some evidence contradicting the assumption that p equals 0.4

n(number of trials)	4
p(proabbility of Success on given trials	0.256
k(number of success)	bionomial distribution
0	0.306402103
1	0.421714723
2	0.217659212
3	0.049928995
4	0.004294967

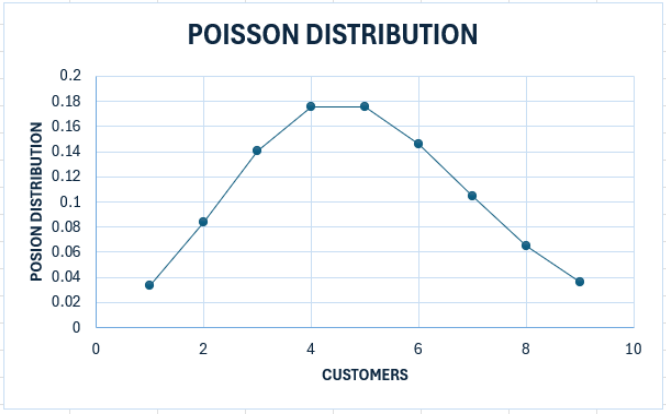


11. Application problems based on the Poisson distribution.

Ques: In a cafe, the customer arrives at a mean rate of 2 per min. Find the probability of arrival of 5 customers in 1 minute using the Poisson distribution formula.

λ	5
X	7

X	Poisson distribution
1	0.033689735
2	0.084224337
3	0.140373896
4	0.17546737
5	0.17546737
6	0.146222808
7	0.104444863
8	0.065278039
9	0.036265577

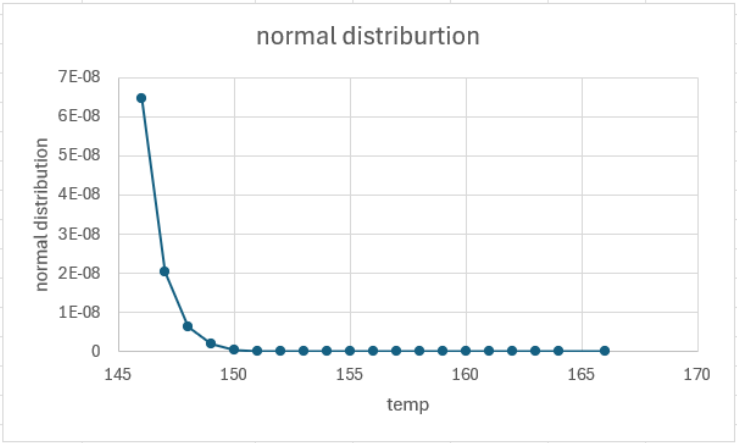


12. Application problems based on the Normal distribution.

Ques:: According to the website of the American Association for Justice, ^11 Stella Liebeck of Albuquerque, New Mexico, was severely burned by McDonald's coffee in February 1992. Liebeck, who received third-degree bums over 6 percent of her body, was awarded $160,000 in compensatory damages and $480,000 in punitive damages. A post-verdict investigation revealed that the coffee temperature at the local Albuquerque McDonald's had dropped from about 185 degree F before the trial to about 158 degree after the trial. This case concerns coffee temperatures at a fast-food restaurant. Because of the possibility of future litigation and to possibly improve the coffee's taste, the restaurant wishes to study the temperature of the coffee it serves. To do this, the restaurant personnel measure the temperature of the coffee being dispensed (in degrees Fahrenheit) at a randomly selected time during each of the 24 half-hour periods from 8 a.m. to 7:30 p.m on a given day. This is then repeated on a second day, giving the 48 coffee temperatures in excel..

Temp	normal distribution
146	6.4483E-08
147	2.04801E-08
148	6.21909E-09
149	1.80563E-09
150	5.01228E-10
151	1.3303E-10
152	3.37575E-11
153	8.19026E-12
154	1.89991E-12
155	4.21379E-13
156	8.93554E-14
157	1.81165E-14
158	3.51185E-15
159	6.50884E-16
160	1.1534E-16
161	1.95416E-17
162	3.16554E-18
163	4.90277E-19
164	7.2601E-20
166	1.39144E-21
167	1.80089E-22
169	2.63663E-24

mean	120.947
standard deviation	4.72008



Bivariate Data/ Bivariate Analysis

Bivariate analysis is one of the statistical analysis where two variables are observed. One variable here is dependent while the other is independent. These variables are usually denoted by X and Y. So, here we analyse the changes occurred between the two variables and to what extent.

The term bivariate analysis refers to as the analysis of two variables . the objective of bivariate analysis to understand the relationship between two variables. There are three common way to analysis the bivariate analysis –

1. Scatter plots

2. Correlation Coefficient

3. Simple linear Regression(SLR)

Bivariate frequency distribution

A series of statistical data showing the frequency of two variables simultaneously is called Bivariate frequency distribution. In other words, the frequency distribution of two variable is called Bivariate frequency distribution. For example: sales and advertisement expenditure , weight and height of an individual.

Why bivariate frequency distribution is significant in business research ?

1. Decision Making

2. Market-segmentation

3. Risk-assessment

4. Resource allocation

how to implement in excel

= COVARIANCE.P(array1,array2)

The COVARIANCE.P function used the following arguments array1, this is range or array of integer value. array2 is also the second range or values.

Few things to remember about argument

1. If the given array contain text or logical value then are ignore by the Covariance function in excel.

2. The data should contain numbers, names, array or references that are numeric .IF the some cell do not contain numeric data they are ignored.

3. The data set should be same size with the same number of data points.

4. The data set should not be empty nor should the standard Deviation of the value equal .

$$Cov(X,Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X and Y are the sample mean of the two set of values and n is the sample size.

5. Covariance is measure to indicate the extent to which two random variable in tandem.

6. Correlation is the measure used to represent how strongly two random variable are strongly related to each other.

7. Covariance is nothing but a measure of correlation.

8. Correlation referred to the scaled form of covariance.

9. Covariance can ary between $-\infty$ to $+\infty$ and correlation range between -1 to +1 .

10. Covariance indicate the direction of the linear relationship between variables .

11. Correlation on the other hand measure both the strength and direction of the linear relationship between two variables.

12. Covariance is affected by change in scale.

13. Correlation is not affected by the change in scale.

Pearson Correlation Coefficient formula

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

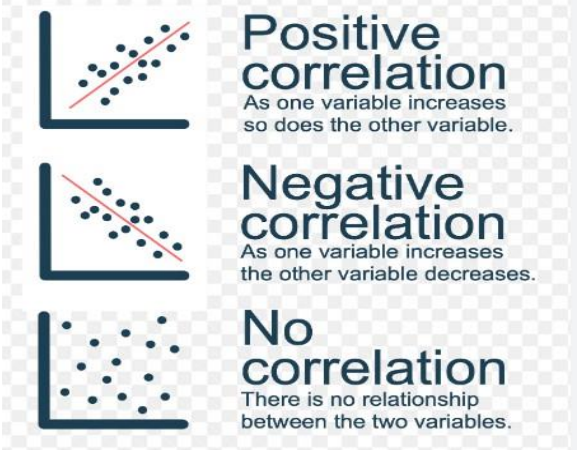
r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable



= PEARSON(arry:arry2)

Scatter plots

Scatter plots *are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a* ***Cartesian system****. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis. These plots are often called* ***scatter graphs*** *or* ***scatter diagrams****.*

Scatter plots instantly report a large volume of data. It is beneficial in the following situations –

- *For a large set of data points given*
- *Each set comprises a pair of values*
- *The given data is in numeric form*

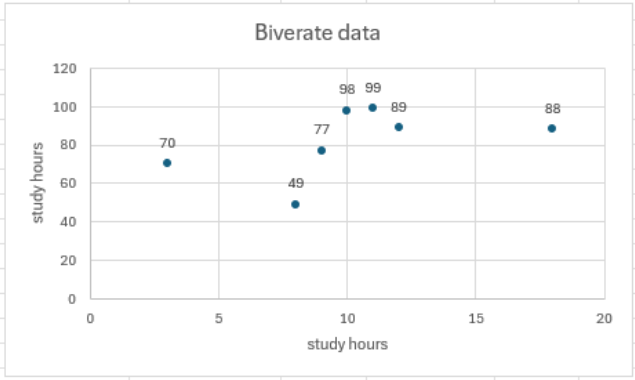
13. Presentation of bivariate data through scatter-plot diagrams and calculations of covariance.

How to perform variant analysis and experiment using the following dataset having two variables .

a) hours spent studying and

b) exam score receive by students

study hours	10	12	18	3	9	8	11
marks	98	89	88	70	77	49	99
CORRELATION	0.487345501						
COVARIANCE	33.51020408						



Click insert tab along the top ribbon then click scatter chart within chart group.

CORRELATION in excel-

= CORREL(hours,score)

COVARIANCE in excel –

COVARIANCE.P(hours,score)

14. Calculation of Karl Pearson's correlation coefficients.

X	Y1	Y2	Y3
2	40	63	10
3	60	59	35
8	80	78	15
11	95	89	22
10	56	60	10
Karl pearson's correlation	XY1	0.723339778	
	XY2	0.616424567	
	XY3	-0.22484401	

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

16_18	18_20	20-22	22_24	total
5	1	1	0	7
2	0	4	2	8
1	3	5	3	12
2	2	2	1	7
3	0	3	1	7
0	1	0	5	6
13	7	15	12	

correlation coffecient along X axis

-0.28222

corelation cofficient along y -axis

-0.02549

margerial frequency distribution of X:

marks	total
17	8
26	9
35	13
32	8
50	6
40	7

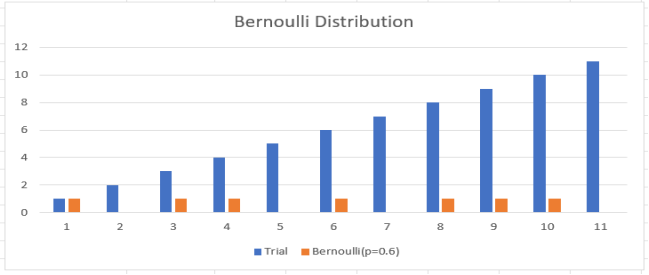
age in years	total
19	13
13	7
11	16
13	11

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

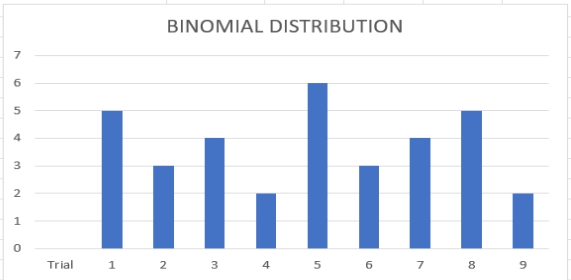
How to implement in excel-

= BINOM.INV (1,P, RAND()) will generate 1 or 0 with chance of 1 being P random number

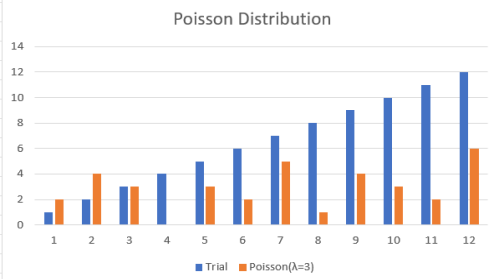
		Bernoulli Distribution
Trial	Bernoulli(p=0.6)	
1	1	p=0.6
2	0	
3	1	
4	1	
5	0	
6	1	
7	0	
8	1	
9	1	
10	1	
11	0	



		Binomial Distribution
Trial	Binomial(n=10, p=0.4)	
1	5	n=10, p=0.4
2	3	
3	4	
4	2	
5	6	
6	3	
7	4	
8	5	
9	2	



		Poisson Distribution
Trial	Poisson(λ=3)	
1	2	λ=3
2	4	
3	3	
4	0	
5	3	
6	2	
7	5	
8	1	
9	4	
10	3	
11	2	
12	6	

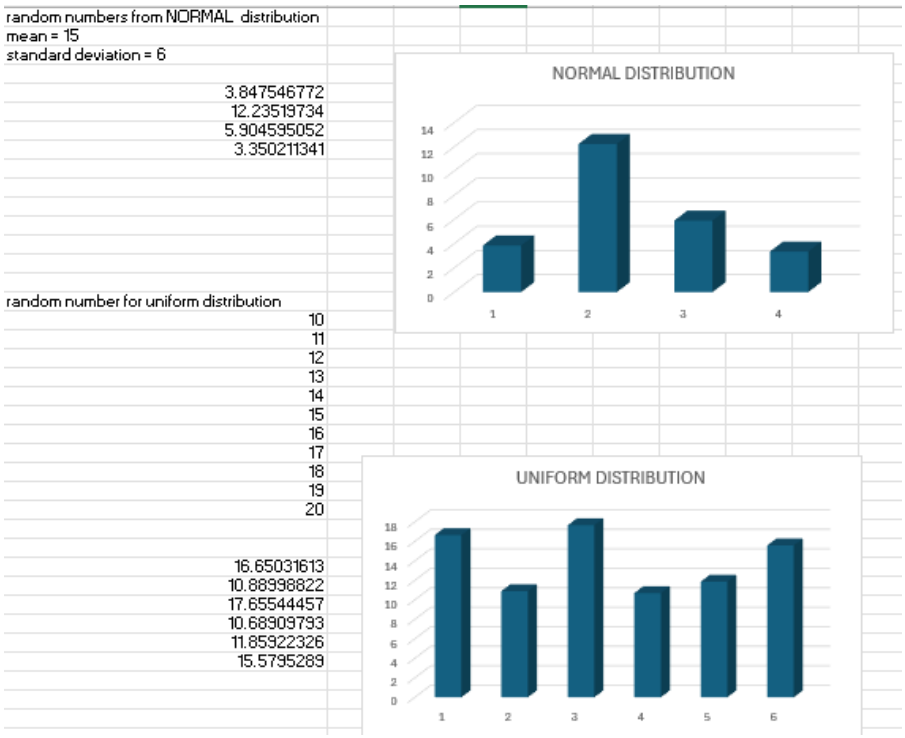


17. Generating Random numbers from continuous (Uniform, Normal) distributions.

How to implement in excel-

= NORMINV(RAND(),B2,C2)

Where this RAND() function create your probability . B2 provides you mean , C2 refers your standand deviation.



Entropy

The entropy of a random variable is the average level of information, surprise, or uncertainly inherent to the variable's possible outcomes. Given a discrete random random variable X which takes value in the alphabet x and distributed according to the P:x[0,1]. The entropy is H[X]

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

The choice of base for log varies for different applications.

Base 2 gives the unit of bits while base e gives ***natural units.***

Base o gives the units of H(X).

An equivalent definition of entropy is the expected value of the self information of a variable .

Two bits of entropy

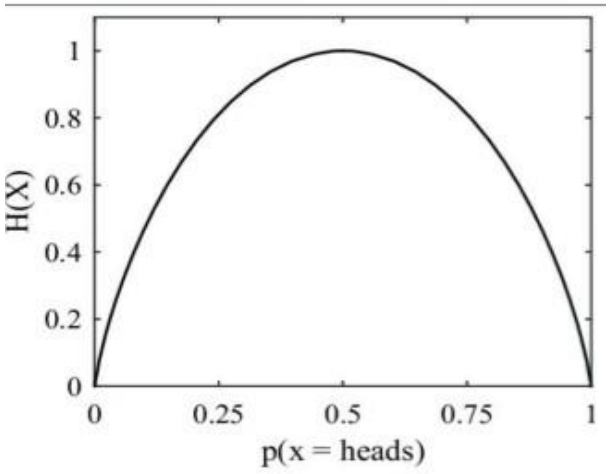
In case of a fair coin tosses, the information entropy enbits base 2 logarithm the number of possible outcomes with two coin, there are 4 possible outcomes and two bits of entropy

{HH,HT,TH,TT}

Generally information entropy is the average amount of information converged by an event, when considering all possible outcomes.

Example –

Entropy H(X) of a coin flip measured in bits, graph versus the bias of the coin where X=1 represent entropy is the result of heads.



Here the entropy is at most 1 bit and to communicate the outcome of a coin flip (2 possible values – H or T) which requires an average of 1 bit (exactly 1 bit for a fair coin)

The result of a fair dice (6 possible values) have entropy log 6 bits.

18. Find the entropy from the given data set.

HH	1	0	0
HT	0.5	-1	0.5
TH	0	0	0
TT	0.5	-1	0.5

P(X)	H(X)
1	0
0.5	1
0	0