SRI AKILANDESWARI WOMEN'S COLLEGE, WANDIWASH

TYPES OF PROBABILITY DISTRIBUTION AND THEIR APPLICATION

CLASS: I PG MATHEMATICS

MRS.P.SUDHA
ASSISTANT PROFESSOR
DEPARTMENT OF MATHEMATICS

SWAMY ABEDHANANDHA EDUCATIONAL TRUST , WANDIWASH

1.Introduction

- In the 17th century, Pascal and Fermat exchanged letters regarding a gambling problem which led to the development of the theory of probability
- In the late 17th and early 18th centuries, the Bernoulli family, particularly Jacob Bernoulli, made substantial contributions to probability theory
- Jacob Bernoulli's work "Ars Conjectandi" laid down the foundations of probability theory and introduced important concepts like the law of large numbers and the Bernoulli trials.

- De Moivre's work on the normal distribution, Baye’s theorem, and Laplace's work on probability and statistics expanded the understanding of random events and their distributions.
- The 20th century witnessed significant advancements in probability theory and distributions. Mathematicians and statisticians like Andrey Kolmogorov, Ronald Fisher, Norbert Wiener, and others contributed to the development and formalization of different probability distributions, including the binomial, Poisson, exponential, gamma, chi-square, and many more

2. Basic Preliminaries:

Probability of an event:

- If a trail results in exhaustive mutually exclusive and equally likely events are favourable to the happening of the events then the probability of happening of an event is given by.

$$P(E) = \frac{Favourable no. of .cases}{Exhaustive no. of .cases}$$

- It is defined as the likelihood that a specific event will occur denoted by a number 0 and 1.

Random Variable:

Suppose that at each point of a sample space we assign a number we then have a function defined on a sample space. This function is called a random variable.It is usually denoted by a capital letter such as X and Y. It is of two types

- Discrete Random variable
- Continuous Random Variable

Discrete Random Variable

- A random variable that takes on a finite or countably infinite number of values is called a discrete random variable

Continuous Random Variable

- A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements

3.Discrete Probability Distribution

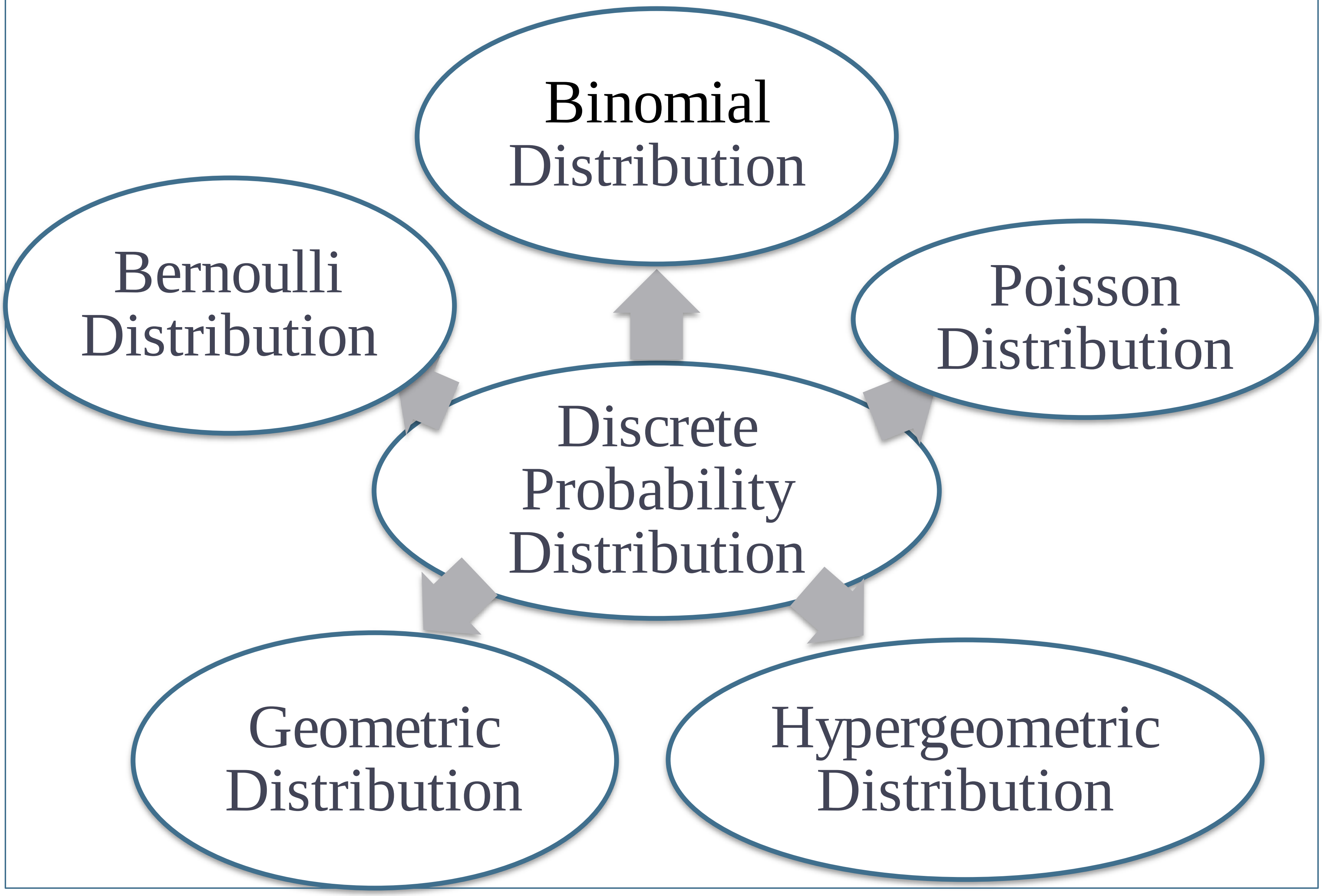
- Let X be a random variable that takes values as x_k for k =1,2,3..Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k)$$

for k=1,2,3…..

- It is convenient to introduce the probability function, also referred to as probability distribution, given by

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


Binomial Distribution
Bernoulli Distribution
Poisson Distribution
Discrete Probability Distribution
Geometric Distribution
Hypergeometric Distribution

Bernoulli Distribution

- The Bernoulli distribution is the density function of a discrete random variable having 0 and 1 as its possible values. It’s a distribution associated with trial which has two possible outcomes called success and failure. Such a trial is called Bernoulli trial. . The sample space for a Bernoulli trial is $S = \{s, f\}$ .A single toss of coin (head or tail)and the throw of die are examples.

Binomial Distribution

- The binomial distribution models the probability of observing a certain number of successes (or failures) in a fixed number of independent trials, each with the same probability of success
- If we define a random variable X on the sample space of Bernoulli trial as follows:

  X=1,if the result of the trial is a success

  X =0,if the result of the trial is a failure

- The PMF of the random variable is given by

$$P(X=x)=\binom{n}{x}p^{x}q^{n-x}=\frac{n!}{x!(n-x)!}p^{x}q^{n-x}$$

where the random variable denotes the number of successes in n trials

- The distribution is called Binomial Distribution

Real Life Problem:

- It has been stated that about 41 of 100 adult patients in a medical lab were observed with some minor curable skin allergies when a specific vaccine is injected but not any major issues. If 20 patients are randomly selected for further observation, find the probability that at most 12 of them have minor allergies but not considerable issues. How many patients can we expect with minor allergies for vaccine?

Poisson Distribution

- The Poisson distribution models the probability of a certain number of events occurring within a fixed interval of time or space, given a known average rate of occurrence and under the assumption that the events are independent of each other
- Let X be a discrete random variable that can take on the values 0,1,2,…. such that the probability function of X is given by,

$$f(x) = P(X = x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

- $x=0,1,2..$
- where is a given positive constant. This distribution is called the Poisson distribution (after S.D. Poisson, who discovered it), and a random variable having this distribution is said to be poisson distributed.

Real Life Problem

- A call center receives about 6 telephone calls between 8 a.m. and 8.30 a.m. What is the probability that they receives more than 1 call in the next 5 minutes?

Geometric Distribution

- It is named after the geometric series, which describes the sum of a sequence of terms with a common ratio.
- The probability mass function of the geometric distribution gives the probability of obtaining the first success on the trial. It is given by

$$P(X = k) = (1 - p)^{k-1} . p$$

where

- X is the random variable representing the number of trials needed to achieve the first success,
- k is the number of trials (which starts from 1 for the first success)
- p is the probability of success on each trial.

Real Life Problem

- Assume that the probability of a defective computer component is 0.02. Components are randomly selected. Find the probability that the first defect is caused by the seventh component tested. How many components do you expect to test until one is found to be defective?

Hypergeometric Distribution

- Suppose that a box contains b blue marbles and r red marbles. Let us perform trials of an experiment in which a marble is chosen at random, its color is observed, and then the marble is put back in the box. This type of experiment is often referred to as sampling with replacement.

- In such a case, if X is the random variable denoting the number of blue marbles chosen (successes) in n trials then using the binomial distribution we see that the probability of exactly x successes is

$$P(X = x) = \binom{n}{x} \frac{b^x r^{n-x}}{(b+r)^n} \qquad x = 0,1, \ldots n$$

- Since $p = b/(b+r)$, $q = 1 - p = r/(b+r)$
- If we modify the above so that sampling is without replacement.i.e., the marbles are not replaced after being chosen

$$P(X = x) = \frac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}} \qquad x = \max(0, n-r), \dots, \min(n, b)$$

- This is the hypergeometric distribution

Real Life Problem

- A school site committee is to be chosen randomly from six men and five women. If the committee consists of four members chosen randomly, what is the probability that two of them are men? How many men do you expect to be on the committee?

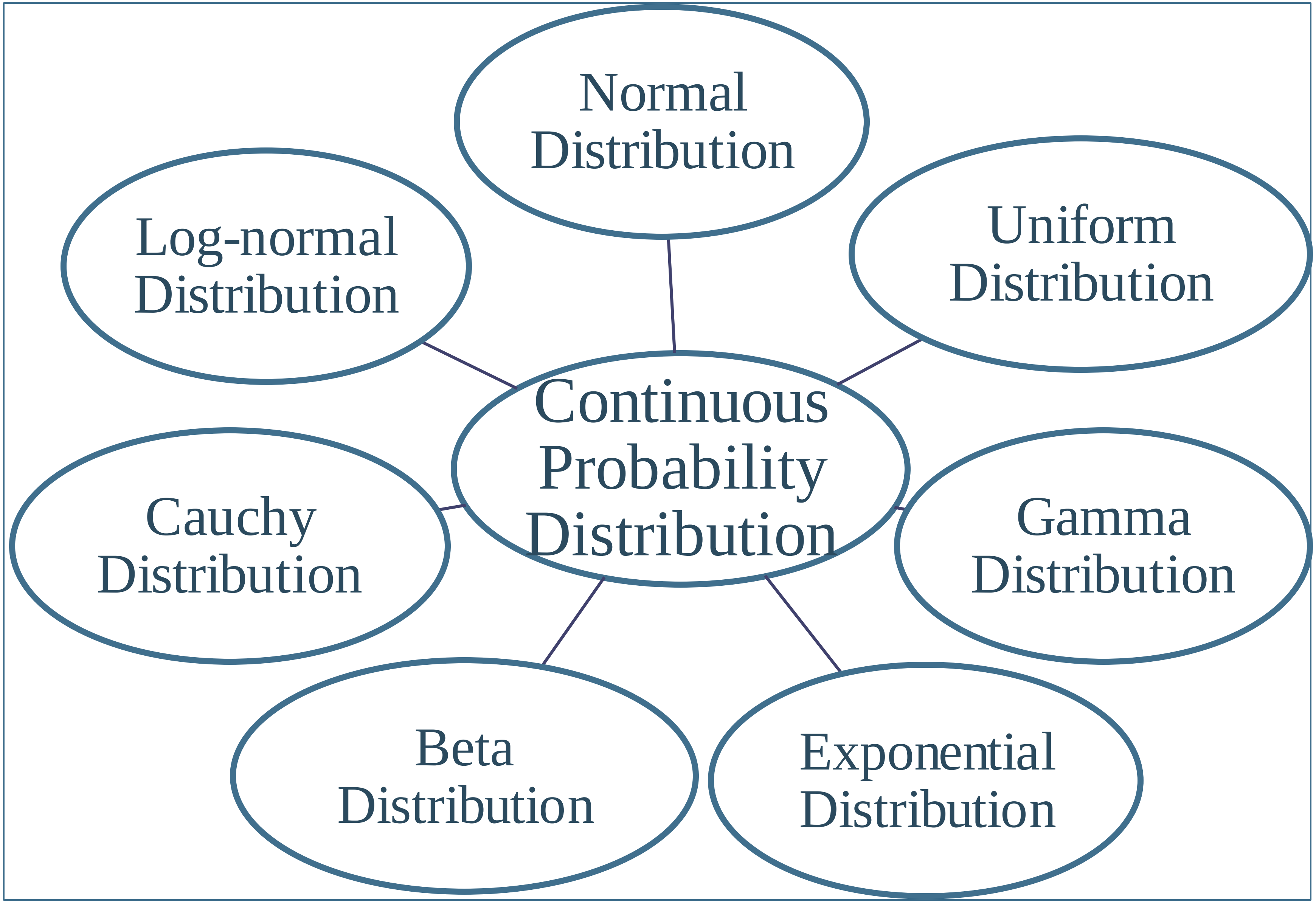
4.Continuous Probability Distribution

- A non discrete random variable is said to be absolutely continuous, or simply continuous, if its distribution function may be represented as

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du \quad (-\infty < x < \infty)$$

where the function $f(x)$ has the following properties ,

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



Normal Distribution
Log-normal Distribution
Uniform Distribution
Continuous Probability Distribution
Cauchy Distribution
Gamma Distribution
Beta Distribution
Exponential Distribution

Normal Distribution

- One of the most important examples of a continuous probability distribution is the normal distribution, sometimes called the Gaussian distribution. Its distribution function is given by,

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv$$

Z-score:

If is a normally distributed random variable and $X \sim N(\mu, \sigma)$

Then the z-score is

$$z = \frac{x-\mu}{\sigma}$$

Standardized Normal Distribution

- The Standard normal distribution is a normal distribution of standardized values called *z*-scores and is measured in units of the SD

Real Life Problem

- The weekly wages of 1000 workers are normally distributed of mean Rs.70 with a SD of Rs.5 estimate the no. of workers where weekly wages will be

i. Between Rs.69 and Rs.72

ii. More than Rs.72

iii. Less than Rs.69

Uniform Distribution

- The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur. When working out problems that have a uniform distribution, be careful to note if the data is inclusive or exclusive of end points.
- A random variable X is said to be uniformly distributed in $a \leq x \leq b$ if its density function is This distribution is called as Uniform Distribution

$$f(x) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & otherwise \end{cases}$$

Its distribution function is given by

$$F(x) = P(X \leq x)$$

$$= \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x < b \\ 1 & x \geq b \end{cases}$$

Real Life Problem

- A Bus is uniformly late between 2 and 10 minutes. How long can you expect to wait? With What Standard deviation? If it is late more than 7 minutes, you'll be late for work. What is the Probability of being late?

Exponential Distribution

- The exponential distribution is a continuous probability distribution that models the time between events in a Poisson process, where events occur continuously and independently at a constant average rate. It is characterized by a single parameter λ, which represents the rate parameter or the average number of events occurring per unit of time.

- The Probability density function of exponential distribution is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Real Life Problem

- In the last 40 years, there have been 200 earthquakes. What is the probability that there will be atleast 3 months until the next earthquake

Gamma Distribution

- The gamma distribution is a continuous probability distribution that is used to model the time until an event occurs, such as the lifetime of a product, the waiting time until a bus arrives, or the duration until a radioactive particle decays.
- A random variable X is said to have the gamma distribution, if the density function is

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

- where $\Gamma(\alpha)$ is the gamma function.
- $x>0$is the random variable representing the time until the event occurs.
- α(shape parameter) and θ(scale parameter) are both positive real numbers.
- $\Gamma(\alpha)$denotes the gamma function, which is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Real Life Problem

- If the life of one computer component (in years) has a gamma distribution with mean 6 and variance 18. Find the probability that this component has a lifetime of atleast 9 years

Beta Distribution

- The beta distribution is a continuous probability distribution defined on the interval [0, 1]
- A random variable is said to have the beta distribution if the density function is

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

- $\alpha, \beta > 0$ where $B(\alpha, \beta)$ is the beta function.

Real Life Problem

- Suppose that DVD's in certain shipment are defective with a beta distribution. Given that Find the probability that shipment has between 20% and 30% defective DVD's

Cauchy Distribution

- A random variable is said to be Cauchy distribution if the density function of is

$$f(x) = \frac{a}{\pi(x^2+a^2)} \quad a > 0, -\infty < x < \infty$$

Log-normal Distribution

- The log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed
- The probability density function is given by,

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

- $x > 0$ is random variable
- μ and σ are mean and Standard deviation of natural logarithm

5.Conclusion

- In this project, we have discussed how different distributions have unique characteristics and are suitable for modeling different types of data.
- For instance, the normal distribution is often used to describe continuous data with symmetric bell-shaped curves, while the Poisson distribution is suitable for modeling the number of rare events occurring in a fixed interval of time or space.

Reference

1. T.W. Anderson, Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, *J. Am. Stat. Assoc.* 52(1957),200-203.
2. J.D. Auble, extended tables for the Mann-Whitney statistic, *Bull. Inst. Educ. Res. 1*(1953),i-iii, 1-39
3. D. Birkes, Generalized likelihood ratio tests and uniformly most powerful tests, *Am. Stat. 44* (1990), 163-166.
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., Wiley, New York, 1968.