DEPARTMENT OF STATISTICS

ST. XAVIER'S COLLEGE (AUTONOMOUS), KOLKATA

Title:

Comparison Between Likelihood Ratio Tests and Some Non-Parametric Tests in Two Sample Problems

- Sampled Population is Normal and Exponential

Name: Shrayan Roy

Roll No: 19-300-4-07-0407

Registration No: A01-1112-0811-19

Supervisor: Prof. Pallabi Ghosh

Session: 2019-2022

DECLARATION

I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

Shrayan Roy.

Shrayan Roy

April,2022

ACKNOWLEGEMENTS

I would like to thank my supervisor Prof. Pallabi Ghosh for helping me throughout my entire research work. Without her proper guidance and resourceful insights, it was not possible to do my research work with full motivation. I have learned so many things under her guidance.

Also, I would like to thank all my other respected Professors of Department of Statistics in St. Xavier's College, Kolkata, Prof. Dr. Surabhi Dasgupta, Prof. Dr. Ayan Chandra, Prof. Dr. Surupa Chakrabarty, Prof. Dr. Durba Bhattacharya, Prof. Debjit Sengupta, Prof. Madhura Dasgupta. They all helped me to develop my interest for Statistics and love for this subject. Which motivated me do such research work. Without their support, it was not possible to complete this project work.

A big thanks to all my peers, who supported me throughout the project and shared their ideas which enlightened my knowledge. Also, I would like to thank all the online resources, videos which helped me grow my knowledge.

Last but not the least, a great thanks to my parents and my family members, without their constant support it was not possible to stay always strong to any challenges.

Shrayan Roy

SXC Kolkata

April,2022

ABSTRACT

Testing of Hypothesis is probably the most important topic in statistical inference. It has wide applications in all areas of statistics. For a particular hypothesis testing problem, we can propose different testing rules. But it is important to base our analysis on a proper testing rule. There are plenty of methods available to find good tests in parametric inference as well as in non-parametric inference. For parametric inference, a routine way to find good testing rules is the 'Likelihood Ratio Test' (LRT). These tests are based on sufficient statistics and are satisfactory in terms of power.

It is very interesting to compare the tests obtained by the Likelihood Ratio method with different non-parametric tests. We will restrict our attention to the test for equality of means for two independent populations. To be more specific, we will consider LRTs for two sample problems, when sampled populations are either normal or exponential. We will compare these LRTs with corresponding non-parametric tests of the two-sample problem for equality of means. In our discussion, we will consider – Kolmogorov-Smirnov test, Mann-Whitney U test, Wilcoxon Rank Sum test. Our main objective is to compare them in terms of power.

In this article, we will use simulations to compare the tests obtained from likelihood ratio tests with non-parametric tests in two-sample problem of equality of means when sampled populations are either normal or exponential. We will assess the penalty of using non-parametric tests instead of using the corresponding LRT (parametric test) when we have sufficient knowledge about the type of population sampled (Here, normal and exponential only).

Keywords: LRT – Likelihood Ratio Test, CDF – Cumulative Distribution Function, Two-Sample Problem, Non-Parametric, Mann Whitney, Kolmogorov-Smirnov, Wilcoxon Rank sum.

List of Diagrams and Tables

Diagrams:

Tables:

CONTENTS

1. Statistical Inference:

Statistical Inference is one of the important topics of statistics. It has wide applications. In statistical inference we have a population and we wish to know about some characteristics of the population. But the population is unknown to us. So, we can't actually say about the characteristics of population. We then wish to collect some information about the population. And for that, we take a sample, which is basically used to represent the population and based on this sample, we then try to explore different characteristics of the population. Since we have no complete knowledge about the population and we are inferring about the population based on the sample drawn from it. Some error may happen in taking conclusions about the population characteristics. In statistical inference we associate different probability models with these errors (Actually we assume that the population characteristic under consideration has some probability distribution) and hence these errors are subject to randomness. This process of exploring the unknown population based on a sample from it is known as Statistical Inference. Now, statistical inference can be classified into two broad areas.

1. Parametric Inference: Suppose, we have knowledge about the functional form of the probability distribution of the study variable. But some constants of this distribution are unknown to us, which characterises the probability distribution completely. Then we will be interested in knowing these unknown constants (called Parameters). That is, we are interested to infer about the parameters. This inference procedure based on parameters of the probability distribution is called 'Parametric Inference'.
2. Non-Parametric Inference: It may happen that, we have no knowledge about the probability distribution of the study variable. That is, we don't know any particular probability law which may be used to characterize the population characteristic. This 'Distribution Free' inference procedure is called 'Non-Parametric Inference'. Often it is called 'Distribution free procedure'.

The main difference is that, we don't assume any specific distribution in case of non-parametric inference. But in case of parametric inference, we assume a specific distribution. Thus, Non-Parametric methods has broader application in inference. It may seem that non-parametric procedures don't involve parameters. But it is not the case. We may parametrize any problem without assuming any particular probability distribution.

Any inferential procedure includes two different components. They are -

1. Estimation – Point Estimation and Interval Estimation
2. Testing of Hypothesis

In estimation we are interested to provide some value to represent the unknown population parameter and in testing of Hypothesis we are rather interested to test some conjectures.
Actually, Estimation and Testing of Hypothesis are related to each other very strongly. But we will discuss here about Testing of Hypothesis only. Although some concepts of estimation will be needed.

2. Hypothesis Testing and associated Terminologies:

2.1. Statistical Hypothesis:

A Statistical hypothesis is a hypothesis about the population distrbution. It is a conjecture or assertion about the population distrbution. Statistical hypothesis is of two different types.

1. Simple Hypothesis
2. Composite Hypothesis

A statistical hypothesis which specifies the probability distribution completely is known as 'Simple Hypothesis' and a statistical hypothesis which does not specify the probability distribution completely is known as 'Composite Hypothesis'.

2.2. Null Hypothesis and Alternative Hypothesis:

For a specific hypothesis testing problem, we have two different types of hypotheses. One is null hypothesis and the other is alternative hypothesis. Consider a population distribution with distribution function $F(x)$. The functional form of the F may be known or may be unknown to us. Suppose, we have two hypotheses of interest (say, H_1 and H_2) and these two hypotheses are mutually exclusive. We have to decide which will be our null hypothesis and which will be our alternative hypothesis. Generally, we denote null hypothesis by H_0 and alternative hypothesis by H_1.

2.3. Formulation of Statistical Test:

Formulation of statistical test includes different steps. Statistical test is basically a procedure to test whether a statistical hypothesis is correct or not. For this we have to collect some information about the population and for that we do some experimentation and collect data. Then, based on the collected data we test the hypothesis. That is make decision using some rule, whether accept or reject the hypothesis. Then, a statistical test is a procedure to test or check the validity of some hypothesis and ultimately make decision about hypothesis, whether to accept or reject it.

For example – Consider a normal population with mean μ and variance $\sigma^2 = 20$. μ is unknown to us. We wish to test $H_0: \mu = 5$. Then, we do some experimentation and collect data from the population. The data is represented by $(x_1, x_2, \ldots\ldots, x_n)$. Then, based on the data we suggest different decision rules for accepting(rejecting) the hypothesis H. If we have data on n observations. Then we may suggest different testing rules.

Rule 1: Reject H_0 if and only if, $\bar{x} > 5.5$, $\bar{x}$ is sample mean.

Rule 2: Reject H_0 if and only if, $\bar{x} > 5.5 + \frac{1.96}{\sqrt{n}}$, $\bar{x}$ is sample mean.

Rule 3: Reject H_0 if and only if, $\tilde{x} > 5 + \frac{1.96}{\sqrt{n}}$, $\tilde{x}$ is sample median.

Rule 4: Reject H_0 if and only if, $x_{(n)} > 6.5$, $x_{(n)}$ is the maximum value in the sample.

One may suggest infinitely many tests. But our task is to choose an appropriate test for testing the hypothesis H_0.

It is clear that in the light of data a test either reject H_0 or accept H_0. So, for each possible sample there are two possible decisions. Hence, the test divides the all-possible sample space into two mutually exclusive or exhaustive parts or regions. the samples for which a test will reject H_0 on the basis of its rule is called Critical Region of that test. It is denoted by w and the complementary region is called Acceptance Region of that test.

Similarly, under H_0 we have a part of the parameter of space (Ω). This is hypothesized parameter space under H_0. It is denoted by Ω_0 and the parameter space under H_1 is denoted by Ω_1. Note that, it may happen that $\Omega_0 \cup \Omega_\perp = \Omega \subseteq \Omega^*$. Here Ω is the parameter space of interest.

2.4. Finding Optimum Test:

Forming a hypothesis testing rule may seem very easy but forming a good testing rule is difficult. Now, the question is, why we are looking for 'good' testing rule?

In the previous example we have suggested four different testing rules. But we cannot just use any testing rule. When we wish to collect data, then we do some experimentation and based on it collect data. We call it a random sample from the probability distribution of the study variable under consideration. Based on the data and the testing rule, we accept or reject our hypothesis. Now, if we again conduct the same experiment and collect data. Then it is very natural to get a different data and it is also possible that we may get different conclusion based on the same testing rule. So, for the same problem we may get different conclusion. So, it is possible to conduct error. We should use that testing rule, for which we have less error. That is, less probable to conduct error.

In testing of hypothesis context, we may conduct two types of error.

1. Type - 1 Error
2. Type - 2 Error

Type - 1 Error: This type of error occurs when we wrongly reject a true null hypothesis. Since, rejecting a true null hypothesis is serious. So, in testing of hypothesis context Type - 1 - Error is more serious.

Type - 2 Error: This type of error occurs when we reject a true alternative hypothesis.

We use probabilities to measure type - 1 error and type - 2 error. The measures are defined as –

$$P(Type\ -1\ -\ Error)\ =\ P_r(Reject\ H_0\ ,When\ it\ is\ actually\ true)$$
$$=\ P_r(Sample\ is\ in\ critical\ region(W)\ |\ H_0\ actually\ true) =\ \alpha$$

$$P(Type\ -2\ -\ Error) = P_r(Reject\ H_1, When\ it\ is\ actually\ true)$$
$$= P_r(Sample\ is\ in\ Acceptance\ region\ \ |\ H_1\ actually\ true)$$
$$=\ 1 - \beta$$

Where, β is the power of the test. In testing of hypothesis power is more frequently used than probability of type -2 error.

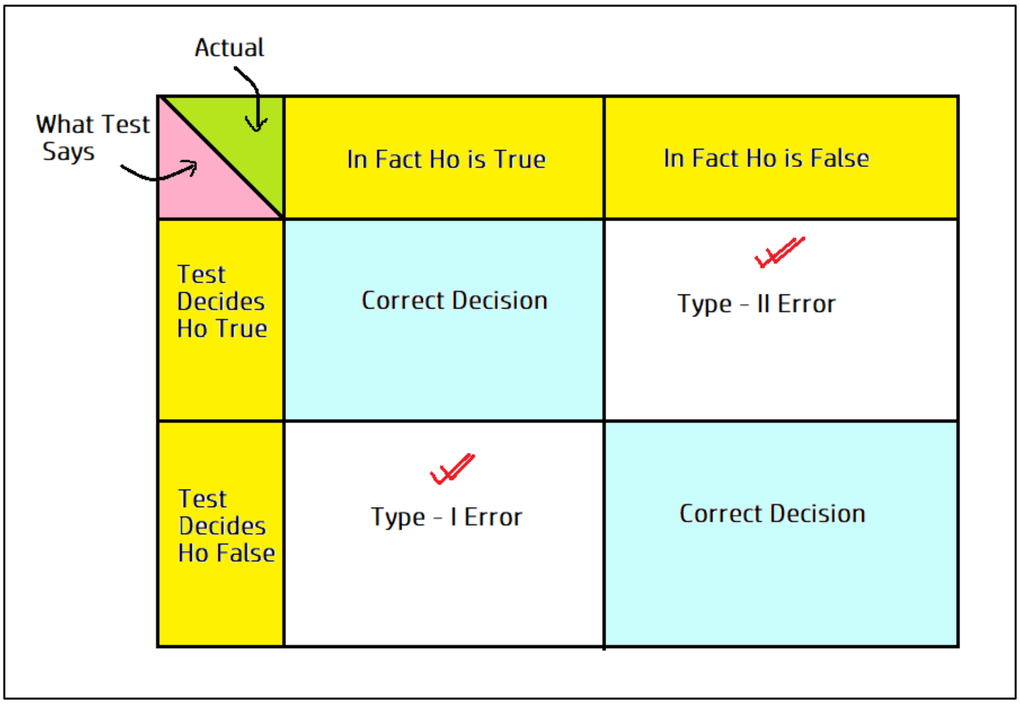


Diagram 1: Table of Correct Decisions and Wrong Decisions in a hypothesis testing problem

2.5. Level of Significance:

It is clear that a good test control type 1 error and type 2 error simultaneously. This simultaneous minimization of error probabilities is difficult. The usual test procedure is to restrict the error probability which is more serious. Thus, we first restrict probability of type 1 error at a certain level and then we minimize probability of type 2 error. We choose a number in [0,1] and then we consider those tests for which probability of type 1 error is less than that chosen value. This fixed number is called Level of Significance. It is denoted by α. (It may be confusing to use α to denote both Type – 1 Error probability and Level of Significance. But it is mostly used notation. Later we will mention explicitly, where α denotes type 1 error probability and where level of significance). Generally low value is chosen as level of significance (For example- 0.10, 0.05, 0.01 etc)

2.6. Power Function & Size of test:

We have defined power of a test. In terms of probability, it is the probability of rejecting the null hypothesis, when it is actually false (This is one of the correct decisions arising from a testing problem). So, obviously high value of power indicates good test. Power is calculated under H_1 (that is H_1 is true). But power function is more general. Considering a function of $\theta \in \Omega$ (parameter space), it is defined as –

$$\beta(\theta) = P_r(Reject\ H_0 |\ True\ Parameter\ value\ is\ \theta), \theta \in \Omega$$

For, θ being parameter value under null hypothesis, $\beta(\theta)$ gives type 1 error probability and for θ being parameter value under alternative hypothesis, $\beta(\theta)$ gives power.

Size of a test is the maximum possible probability of committing type 1 error. Hence it is defined as –

Size of a test = $Supremum\ of\ \ \beta(\theta)\ over\ hypothesized\ value\ of\ \theta\ under\ H_0$

$$= \sup_{\theta \in \Omega_0} \beta(\theta)$$

Size of a test is very important concept in testing of hypothesis. Even size is frequently used than type – 1 error. It is very important to note that θ is parameter. It may be population mean, median, quantile, difference between two population mean, ratio of variances etc. It depends upon problem.

2.7. Graphical Meaning of Power:

Up to now we have discussed different types of error in a hypothesis testing problem, the different criterion for a good test, power of a test, size of test, power function of a test.

Theoretically we have understood the meaning of power, size. But for a better understanding we will use of diagrams.

Consider a testing problem, suppose for a normal distribution with unknown parameter μ and known variance $\sigma^2 = 25$, we want to test $H_0 : \mu \leq 10$ against $H_1 : \mu > 10$. For that we draw sample of size 20 from the population. A test rule to reject H_0 maybe –

$$\bar{X} > 10 + \frac{8.22426}{\sqrt{20}}$$

It can be shown that, the power function of the above testing rule is increasing in μ.

Here, we will try to understand using probability density function what is actually power of a test and why it is increasing as μ increases. This will be needed later in our main discussion.

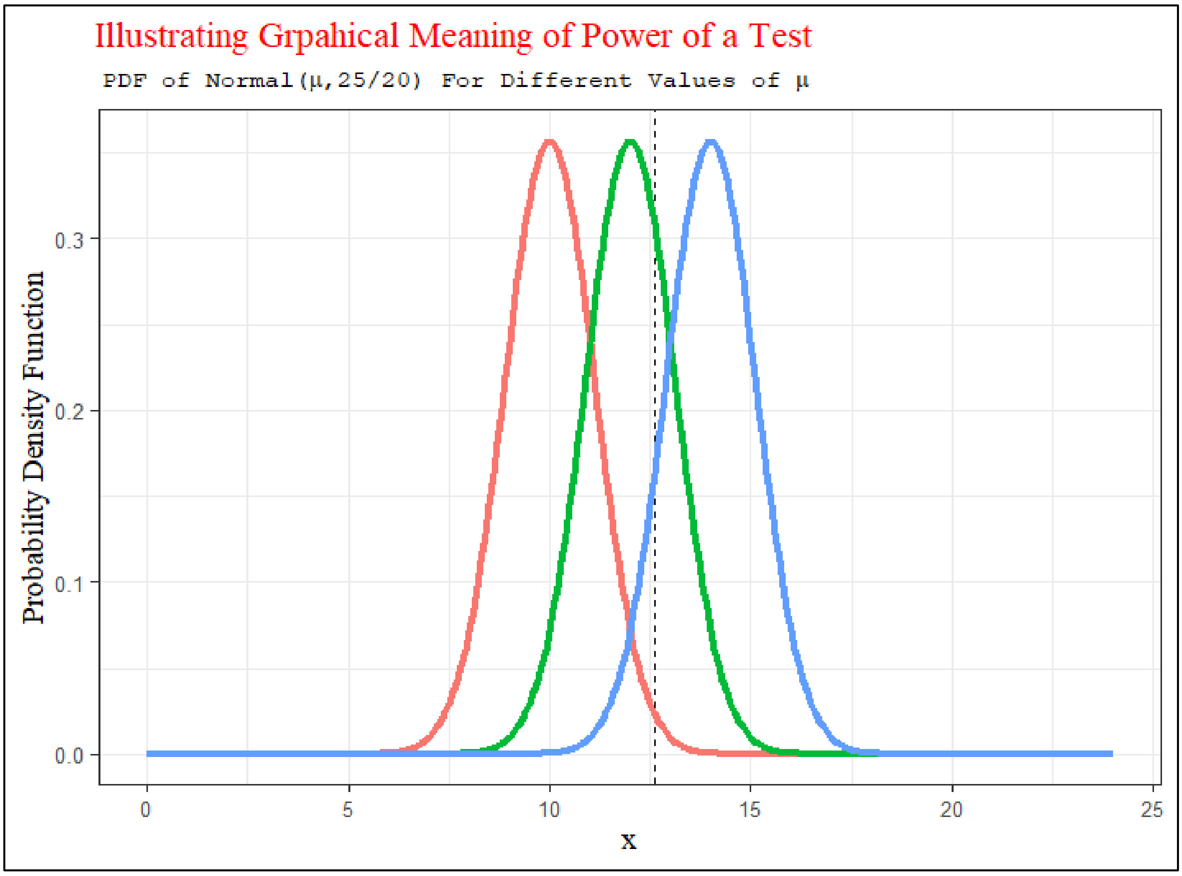


Diagram 2: Graphical meaning of power of a test (testing rule)

To graphically understand meaning of power, we have plotted the pdf of sample mean for different values of μ. We take $\mu = 10,12,14$. The vertical line is drawn at $x = 10 + \frac{8.22426}{\sqrt{20}}$. For any particular value of μ, the area under the pdf to right of the vertical line indicates the probability that $\bar{X} > 10 + \frac{8.22426}{\sqrt{20}}$, for that value of μ. That is for any particular value of μ, area to the right of the vertical line under pdf indicates value of power function for that value of μ.

It is clear that as μ increases, the pdf shifts to the right. Thus, area under the pdf to the right of vertical line increases, for which power function increases. Thus, power of the test in this case totally depends upon the probability distribution of $\bar{X}$.

2.8. Test Statistic:

Every hypothesis test is formed using a statistic. Under H_0 whose distribution is completely known to us. In hypothesis testing this statistic is often called as test statistic. Since, test statistic is actually a statistic in nature. It does not involve any parameter.

2.9. Level - α - Class:

For a testing problem we can form many tests (testing rule). But we will restrict our attention to those tests, whose level of significance is a particular value (say, α), which is specified by the statistician or decision maker depending upon the seriousness of wrongly rejecting H_0. Depending upon different test statistic we will get different tests with level of significance as α. Thus, all tests with level of significance α forms a class of tests, called Level - α- Class. Now, in this level - α- class we will find the good test.

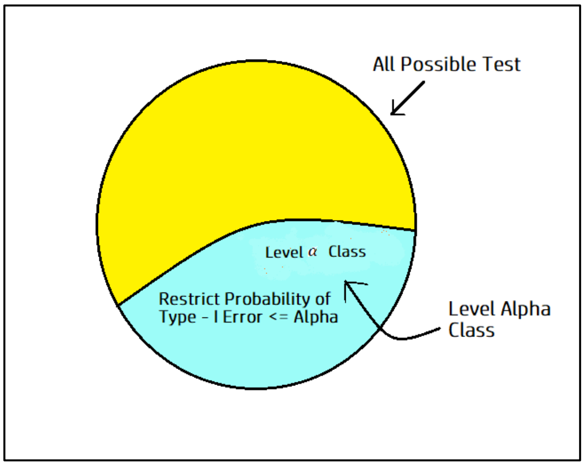


Diagram 3: Level α Class

2.10. One Sided and Two-Sided Tests:

In Testing problem depending on the alternative hypothesis if the test uses the left tail (right tail) of the curve of sampling distribution of the test statistic (Under H_0) as critical region. Then the test is known as *Left-Tailed* (*Right-Tailed*) test. Together they are called *One-Sided Tests*. On the other hand, if the test uses both the tails of the curve of sampling distribution of the test statistic (Under H_0) as critical region. Then the test is called *Two-Sided Test*.

2.11. P Value:

The choice of specific value of level of significance (α) is completely arbitrary and is determined by non-statistical considerations such as possible implication of false rejection of H_0 and cost. However, there is another value associated with a test, called P-value of the test. The P-value associated with a test is the probability that the test statistic T takes the observed value T_{obs} and more extreme value in the direction indicated by the alternative under H_1.
The smaller the P-value, the more extreme the outcome and the stronger the evidence against H_0. If α is our level of significance, then for a testing problem we reject H_0 at level α iff P-Value $\leq \alpha$.

3. Methods of Finding tests:

3.1. Introduction:

Up to this we have discussed different components of a hypothesis testing problem. But how to exactly form a test? that is how to form such decision rule for a hypothesis problem? we have not discussed about it till now.
In Parametric inference we generally use special types of methods to find appropriate tests for a particular hypothesis testing problem. Those methods are very useful in order to form a good decision rule. Some methods of finding good tests are –

1. Exact tests (A heuristic approach of finding test)
2. Likelihood Ratio Test (LRT)
3. Most powerful test (MP test)
4. Union – Intersection test   and etc.

Exact tests are suggested by R.A. Fisher and it is a heuristic approach for finding good test. Likelihood ratio tests are very useful in testing composite type of hypothesis. Most powerful tests are useful in finding tests which have maximum possible power for a particular testing problem. It is to be noted that, there are many more methods for forming a good test. The methods that we have just discussed are mainly used for findings tests in parametric inference. In case of non-parametric inference most tests are based on linear rank statistics, number of runs etc (we will discuss about them later). Those tests are flexible and can be used for any population distribution. Thus, comparison between these tests is important.
Our motive is to compare parametric tests and non-parametric tests for a particular type of hypothesis problem. But before that, we will discuss about Likelihood Ratio test for finding a good tests in the context of parametric inference.

3.2. Likelihood Ratio Tests:

Consider the population distribution $\{f(x;\theta): \theta \in \Omega^*\}$ where $f(x;\theta)$ is the PDF or PMF of the population distribution and $\theta \in \Omega^* \subseteq \mathbb{R}^k$.

We want to test, $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega_1$. Where $\Omega_0 \cap \Omega_1 = \Omega \subseteq \Omega^*$. Here, Ω is the parameter space of interest.

To test the hypothesis, we draw a random sample $(X_1, X_2, \ldots, X_n)$ of size n from the population distribution. Let, $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ be an observed sample(data). Then $\prod_{i=1}^{n} f(x_i;\theta)$ is the PDF or PMF of $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$.

Then, for the given data $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, $\prod_{i=1}^{n} f(x_i;\theta) = L\left(\underset{\sim}{x};\theta\right)$ is the likelihood function of the data $\underset{\sim}{x}$, as function of $\theta \in \Omega^*$.

Note that, $\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right)$ gives the best possible likeliness or explanation of the data $\underset{\sim}{x}$ over Ω_0.

Similarly, $\sup_{\theta\in\Omega_1} L\left(\underset{\sim}{x};\theta\right)$ gives the best possible likeliness or explanation of the data $\underset{\sim}{x}$ over Ω_1.

The basic idea is to compare $\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right)$ and $\sup_{\theta\in\Omega_1} L\left(\underset{\sim}{x};\theta\right)$. Which is large for a given data.

Since, value of likelihood is generally small. It is logical to compute the ratio $\frac{\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right)}{\sup_{\theta\in\Omega_1} L\left(\underset{\sim}{x};\theta\right)}$. If the value of the ratio is large, then the data is best explained by Ω_0. While if the value is small, then the data is best explained by Ω_1. But the problem with this measure is that, it is unbounded. In order to get a bounded measure, we consider the ratio –

$$\lambda(\underset{\sim}{x}) = \frac{\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right)}{\sup_{\theta\in\Omega} L\left(\underset{\sim}{x};\theta\right)}$$

As, $\Omega_0 \subset \Omega$, $\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right) \le \sup_{\theta\in\Omega} L\left(\underset{\sim}{x};\theta\right)$. Which implies $0 \le \lambda(\underset{\sim}{x}) \le 1$. Here, $\lambda(\underset{\sim}{x})$ is a statistic. Because, it is independent of the parameter. So, the test will be carried out by likelihood ratio (LR) $\lambda(\underset{\sim}{x})$.

Note that, a small value of $\lambda(\underset{\sim}{x})$ near zero indicates that $\sup_{\theta\in\Omega_0} L\left(\underset{\sim}{x};\theta\right)$ is quite small in comparison to $\sup_{\theta\in\Omega} L\left(\underset{\sim}{x};\theta\right)$. Which implies that there is much better explanation over $\Omega =$

$\Omega_0 \cup \Omega_1$ than the best possible explanation provided by Ω_0. That is the data supports Ω_1 than Ω_0.

Hence, a small value of $\lambda(\underset{\sim}{x})$ near zero supports $H_1 : \theta \in \Omega_1$ and suspects $H_0 : \theta \in \Omega_0$.

Definition: For testing $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega_1$, a test of the form: "Reject H_0 iff $\lambda(\underset{\sim}{x}) < c$". Where, c is so chosen that $\sup_{\theta \in \Omega_0} P_\theta(\lambda(\underset{\sim}{x}) < c) = \alpha$ is called Likelihood Ratio test at size α.

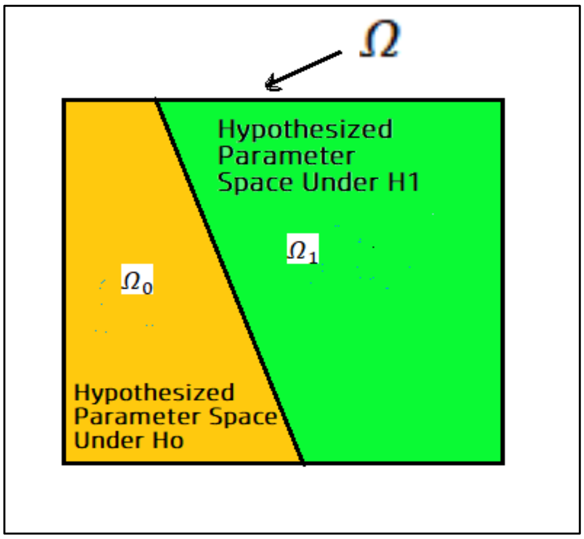


Diagram 4: Dividing parameter space in LRT

3.3. Properties of Likelihood Ratio Tests:

Now, we will discuss some important properties of likelihood ratio tests.

1. The Likelihood Ratio Test is based on sufficient statistics.
2. Under some Regularity conditions on the pdf, the random variable $-2ln(\lambda(\underset{\sim}{x}))$ under H_0 is asymptotically distributed as a Chi-Square (χ^2) Random Variable with degrees of freedom k. Where, k equals to the difference between the number of independent parameters in Ω and the number in Ω_0.

4. Family of Probability Distributions:

The family of probability distributions is the collection of different probability distributions. i.e., collection of DFs, PDFs, PMFs. A family of distributions can be characterized by parameters. Here we will discuss about two families. 1. Location Family 2. Scale Family

1. **Location Family**: Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$. Is called the location family with standard pdf $f(x)$ and μ is called the location parameter of the family. For example - For $Normal(\mu, \sigma^2)$ the location parameter is μ. If we keep $\sigma^2 = \sigma_0^2$ fixed. Then, for different values of μ, we get *Normal Location family.*

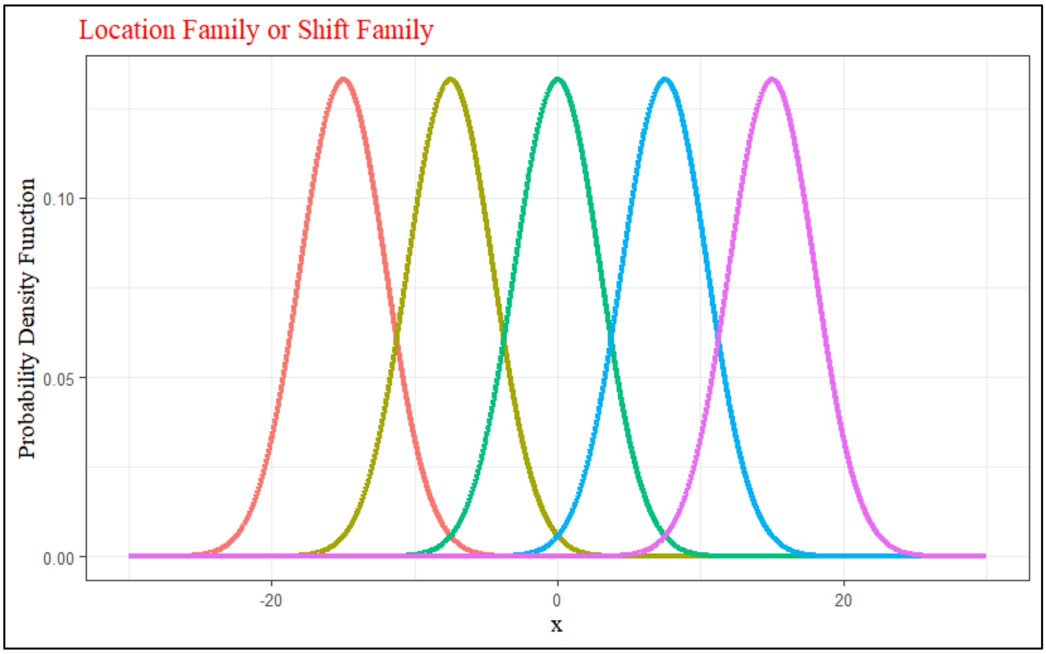


Diagram 5: Location family of Shift family

From the above graph we can clearly see that PDFs of location family are simply obtained by shift in location of the standard pdf (PDF of Normal (0, σ_0^2) distribution), keeping shape of the standard pdf intact. However, one should not think that location parameter is always mean or, median of the distribution.

2. **Scale Family**: Let $f(x)$ be any pdf. Then the family of pdfs $(1/\sigma)f(x/\sigma)$indexed by the parameter σ, $0 < \sigma < \infty$. Is called the scale family with standard pdf $f(x)$ and σ is called the scale parameter of the family. For example – For Exponential (mean = μ) the scale parameter is μ. Then, for different values of μ, we get *Exponential Scale family.*

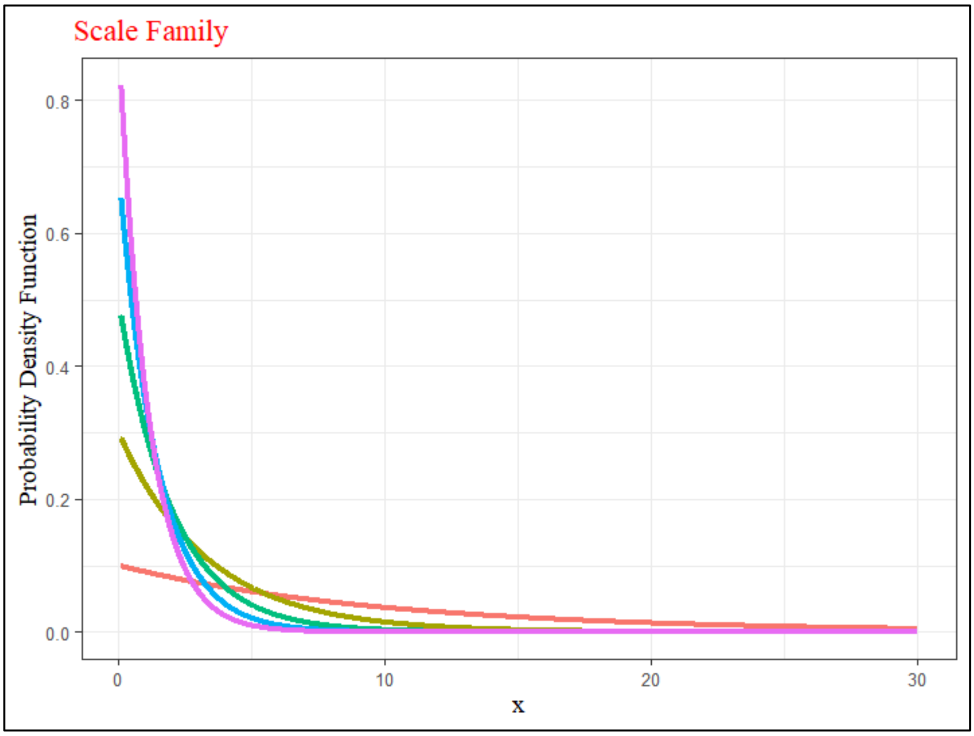


Diagram 6: Scale family

From the above graph we can clearly see that PDFs of scale family are simply obtained by either stretch or by contract of the graph of standard pdf (Here, PDF of Exponential Distribution with mean 1) while maintaining the same basic shape of the graph.

The use of Location families and scale families in hypothesis testing is important. We will discuss about them in the next section.

5. Different Types of Hypothesis Problem:

Depending upon how sampling is done from the population, we can have different types of hypothesis problems.

1. One Sample Problem
2. Paired Sample Problem
3. Two Sample Problem
4. Multi Sample Problem etc.

Here we will discuss about Two sample problem only.

5.1. Two Sample Problems:

The two sample problems are different from paired sample problems. In paired sample problem the data is consisted of two samples, but each element in one sample is related with a particular element of the other sample by some unit of association. The sampled data can be considered as two dependent sample or alternatively single sample of pairs from a bivariate population. But in case of two sample problem, we are interested in data related to two independent population only i.e., random sample drawn independently from two independent population. Here not only the elements within a sample are independent but also every element in the first sample is independent of every element in the second sample. Some real-life examples are –

1. A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 27 men and 21 women were employed in the experiment. Their reaction time to a particular stimulus is noted. Now, on the basis of the given data we will compare the mean reaction time of men and women. Clearly, it is two sample problem related to location parameter.
2. The manager of a dairy is in the market for a new bottle-filling machine and is considering machines manufactured by companies A and B. If ruggedness, cost, and convenience are comparable in the two machines, the deciding factor will be the variability of fills. The machine producing fills with the smaller variance being preferable. Thus, he/she will randomly select bottles from the two machines manufactured by companies A and B and will take necessary measurements to check which has smaller variance. Clearly, it is two sample problem related to scale parameter.

5.2. Hypothesis of Interest in Two Sample Problems:

We have already discussed about two sample problems in a brief way. Two sample problems are our main concern, so we will discuss more about them in details. As we have said that the fundamental difference between two sample problems and paired sample problems is that the samples drawn from two different population are independent.

The statistical universe consists of two populations, which we can call X population and Y population, with cumulative distribution functions denoted by F_X and F_Y respectively. Then We draw a random sample of size n_1 from the X population and random sample of size n_2 from Y population. They are denoted by – $(X_1, X_2, \dots, X_{n_1})$ and $(Y_1, Y_2, \dots, Y_{n_2})$. Usually, the hypothesis of interest in two sample problems are related to location and scale parameter. But to make it more general we usually define the hypothesis problems in terms of CDFs. A very commonly used hypothesis problem is that the two samples are drawn from identical population. i.e.,

$$H_0: F_Y(x) = F_X(x) \ \forall x \ \dots. (*)$$

But if we are willing to make parametric assumptions concerning the forms of the underlying populations and assume that the differences between the two populations occur only with respect to some parameters, such as mean or the variance. For example, if it is assumed that Population X has the probability distribution $Normal(\mu_1, \sigma^2)$ and Y has the probability distribution $Normal(\mu_2, \sigma^2)$. Then the hypothesis (*) boils down to $H_0: \mu_1 = \mu_2$. Then, a testing rule (e.g., LRT) related to equality of means of two normal population with equal variances can be used to test the hypothesis. The performance of those tests depends upon the parametric assumptions. But instead, if we use non-parametric tests for the same hypothesis, there is no such parametric assumptions related to form of distribution.

5.3. Concept of Stochastic Ordering of Two Random Variables:

We have discussed the choice of Null hypothesis in general two sample problems. Now, it is important to know different types of alternatives hypotheses also. The concept of stochastic ordering will be needed here. If X and Y be two random variables with distribution functions F_X and F_Y respectively. Then, X is said to be stochastically larger than Y if $F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x. This simply means that on an average X takes larger value than Y. In terms of probability, we can say that $P(X > Y) > P(X < Y)$.

Again, X is said to be stochastically smaller than Y if $F_Y(x) \leq F_X(x)$ for all x and $F_Y(x) < F_X(x)$ for some x. This simply means that on an average Y takes larger value than X. In terms of probability, we can say that $P(X < Y) > P(X > Y)$.

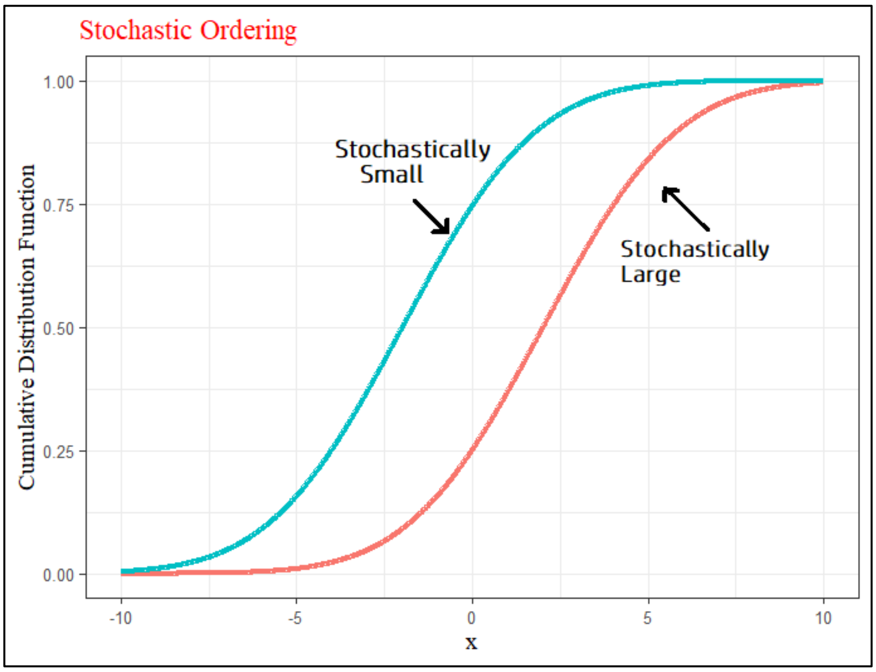


Diagram 7: Concept of Stochastic Ordering

5.4. Different Types of Alternatives Under Two Sample Problem:

We will discuss different choices of alternative hypothesis under two sample problem. If our null hypothesis is the two samples are drawn from common identical population with distribution function unspecified, then under null hypothesis, it is expected that in the combined ordered (increasing) sample the orderings of sample values from population X and Y are totally random. Because, if actually X is stochastically larger than Y, then it is expected that X value will have larger orderings than Y values. While if X is stochastically smaller than Y, then it is expected that X values will have lower orderings then Y values. Thus, orderings in the combined sample can be used to test whether null hypothesis is true or not. It is the basis of non-parametric two sample tests. While in case of parametric tests we can consider the difference between location measures (e.g., sample mean, median etc.) as a tool for testing the hypothesis. Thus, alternative hypotheses can be formed using this concept of stochastic orderings. The different types of alternatives are –

1. $H_1 : F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x
2. $H_1 : F_Y(x) \leq F_X(x)$ for all x and $F_Y(x) < F_X(x)$ for some x
3. $H_1 : F_Y(x) \neq F_X(x)$ for some x

It is to be noted that in parametric set up under H_0 the common CDF may be unspecified but its functional form is known to us. But in case of non-parametric set up the functional form is also unknown to us. Thus, in case of parametric set up the null and alternative hypothesis boil down to hypothesis formed using parameters only. We will see this equivalence later in our discussion.

Sometimes, in some non-parametric hypothesis problems, the hypotheses are written in terms of some probability. For example, in Mann Whitney U test (To be discussed later) the null hypothesis is $H_0 : P(X > Y) = 0.5$ and alternative hypothesis may be any of the following –

1. $H_1 : P(X > Y) > 0.5$
2. $H_1 : P(X > Y) < 0.5$
3. $H_1 : P(X > Y) \neq 0.5$

Thus, there are variety of ways to form null hypothesis and alternative hypothesis.

6. Study of Two Sample Problems:

6.1. Introduction:

In two sample problems we have different testing rules. Under parametric set up we have parametric tests and under non-parametric set up we have non-parametric tests. Parametric tests are based on some parametric assumptions (e.g., Probability distribution, Homogeneity of variances). If the parametric assumptions hold, then the tests are well and good. While violation of parametric assumptions has significant impact on the performance of tests. Although some parametric tests perform well even if parametric assumptions are not valid. On the other hand, non-parametric tests have no such assumptions, except the probability distribution of the study variable must be continuous (for some tests).

In this article our focus is the two-sample problem related to equality of mean of the two independent populations. We will mainly study the parametric tests and non-parametric tests, when our sample data is generated from Normal Distribution and Exponential Distribution. Our main aim is to compare different non-parametric tests with corresponding parametric tests. For example – If the probability distribution of study variable is normal distribution. then, we will compare between the parametric test of normal and different non-parametric tests in terms of level and power.

The choice of Normal Distribution and Exponential distribution is objective. Since, our aim is to compare different parametric tests for the hypothesis of equality of mean under different situations with corresponding non-parametric tests. The mean of normal distribution is its Location Parameter, while the mean of exponential distribution is its Scale Parameter. So, we can perform a further study of what is the effect on the non-parametric tests when the mean is location parameter and when the mean is scale parameter.

6.2. Our Hypothesis Problems:

We have already mentioned that our objective is to study the hypothesis of equality of means of two independent population under two different situations. When both the population are normal and when both the population are exponential. We state our hypothesis problem.

1. Suppose, we have two statistical population denoted by X and Y. The probability distribution of X is $Normal(\mu_1, \sigma_1^2)$ and that of Y is $Normal(\mu_2, \sigma_2^2)$. Now, we want to test the hypothesis $H_0: \mu_1 = \mu_2$.
2. Suppose, we have two statistical population denoted by X and Y. The probability distribution of X is $Exponential(Mean = \mu_1)$ and that of Y is $Exponential(Mean = \mu_2)$. Now, we want to test the hypothesis $H_0: \mu_1 = \mu_2$.

Now, we will define our alternative hypothesis. For both the null hypothesis we have three different choices of alternatives.

- $H_1: \mu_1 > \mu_2$
- $H_1: \mu_1 < \mu_2$
- $H_1: \mu_1 \neq \mu_2$

The first two alternatives are one-sided alternatives and the last alternative is both-sided alternative. We will choose $H_1: \mu_1 > \mu_2$ as our alternative hypothesis. Because, our objective is to study parametric tests and different non-parametric tests under different situations. So, first two hypotheses are basically equivalent with respect to our study. And the last hypothesis (both-sided) is basically taking the first two together.

6.3. Parametric Tests for Two Sample Problems:

We will now derive parametric tests for the testing problem $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$, when both the population distribution is Normal and when both the population distribution is Exponential. We will derive LRT for both cases.

6.3.1. Population Distribution is Normal Distribution:

6.3.1.1. Derivation of the test:

When both X and Y population is Normal population. i.e., The probability distribution of X is $Normal(\mu_1, \sigma_1^2)$ and that of Y is $Normal(\mu_2, \sigma_2^2)$. We want to test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$. We draw a random sample $(X_1, X_2, \ldots, X_{n_1})$ of size n_1 and a random sample $(Y_1, Y_2, \ldots, Y_{n_2})$ of size n_2 independently from the distribution of X and Y respectively. Let us define the following quantities-

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ be respectively the sample means of X and Y.

$S_x^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$ and $S_y^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ respectively be the sampling variance of X and Y.

Now, the pdf of X is given by –

$$f_x(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2};\ -\infty < x < \infty$$

And that of Y is given by –

$$f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2};\ -\infty < y < \infty$$

Then the joint pdf of $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ is given by –

$$f(\underset{\sim}{x}, \underset{\sim}{y}\,;\ \mu_1, \mu_2, \sigma_1^2, \sigma_1^2) \ = \ \prod_{i\,=\,1}^{n_1} f_x(x_i)\ x\ \prod_{i\,=\,1}^{n_2} f_Y(y_i)$$

$$= \ \frac{1}{(\sigma_1\sqrt{2\pi})^{n_1}}\ e^{-\frac{1}{2}\sum_{i\,=\,1}^{n_1}\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2}\ \frac{1}{(\sigma_2\sqrt{2\pi})^{n_2}}\ e^{-\frac{1}{2}\sum_{i\,=\,1}^{n_2}\left(\frac{y_i-\mu_2}{\sigma_2}\right)^2}$$

We assume that, $\sigma_1^2 \ = \ \sigma_2^2 \ = \ \sigma^2$. This assumption is needed to get a simplified form of the test statistic whose distribution is known under H_0. We will discuss about it in more detail. Under this assumption the simplified form of the joint distribution is –

$$f\left(\underset{\sim}{x}, \underset{\sim}{y}\,; \mu_1, \mu_2, \sigma^2\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{n_1\,+\,n_2}}\, e^{-\frac{1}{2\sigma^2}\{\sum_{i\,=\,1}^{n_1}(x_i-\mu_1)^2+\sum_{i\,=\,1}^{n_2}(y_i-\mu_2)^2\}}$$

Thus, likelihood function for the observed data $\underset{\sim}{x}, \underset{\sim}{y}$ is given by –

$$L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right) \ = f\left(\underset{\sim}{x}, \underset{\sim}{y}\,; \mu_1, \mu_2, \sigma^2\right)$$

Now, in this hypothesis problem-

Parameter space under H_0 is $\Omega_0 \ = \ \{(\mu_1, \mu_2, \sigma^2) : \mu_1 = \ \mu_2 = \mu \ \in \ \mathbb{R}\,, 0 \ < \ \sigma \ < \infty\}$

Parameter space under H_1 is $\Omega_1 = \ \{(\mu_1, \mu_2, \sigma^2) : \mu_1 \ > \ \mu_2 \ \in \ \mathbb{R}, 0 \ < \ \sigma \ < \infty\}$

Thus, the parameter space of interest is –

$\Omega = \ \{(\mu_1, \mu_2, \sigma^2) : \mu_1 \ \geq \ \mu_2 \ \in \ \mathbb{R}, 0 \ < \ \sigma \ < \infty\}$

Now, we have to compute the ratio –

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) \ = \frac{\sup\limits_{\Omega_0} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)}{\sup\limits_{\Omega} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

Thus, to find the Likelihood Ratio Test we have to maximize the likelihood $L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)$ under Ω_0 and Ω. But the likelihood maximization problem under Ω is difficult, because it is not an open rectangle type parameter space. It would be easier if Ω is an open rectangle.

Actually, the test statistic for One-Sided Hypothesis and Two-Sided Hypothesis are exactly same (under same parametric condition). Hence, if we find LRT for the alternative $H_1: \mu_1 \neq \mu_2$ and then use the test statistic for $H_1: \mu_1 > \mu_2$. It will suffice our purpose. So, we will do this.

For the alternative $H_1: \mu_1 \neq \mu_2$, parameter space under H_1 is given by -

$\Omega_1 = \ \{(\mu_1, \mu_2, \sigma^2) : \mu_1 \neq \ \mu_2 \ \in \ \mathbb{R}, 0 \ < \ \sigma \ < \infty\}$

Thus, parameter space of interest is $\Omega = \ \{(\mu_1, \mu_2, \sigma^2) : \mu_1\,, \mu_2 \ \in \ \mathbb{R}, 0 \ < \ \sigma \ < \infty\}$

To find $\sup_{\Omega_0} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)$, we need to find MLEs of μ_1, μ_2, σ^2 over Ω_0.

Under H_0, $\mu_1 = \mu_2 = \mu$.

Hence, $\hat{\mu}_{MLE} = \frac{n_1\bar{X}+n_2\bar{Y}}{n_1+n_2}$ and

$\hat{\sigma}_0{}^2{}_{MLE} = \frac{\sum_{i=1}^{n_1}(X_i - \hat{\mu}_{MLE})^2 + \sum_{i=1}^{n_2}(Y_i - \hat{\mu}_{MLE})^2}{n_1+n_2}$

Thus, $\sup_{\Omega_0} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right) = L\left(\hat{\mu}_{MLE}, \hat{\mu}_{MLE}, \hat{\sigma}_0{}^2{}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)$

Now to find $\sup_{\Omega} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)$, we need to find MLEs of μ_1, μ_2, σ^2 over Ω. Now, the MLEs are given by –

$$\hat{\mu}_{1MLE} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i = \bar{X}$$

$$\hat{\mu}_{2MLE} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i = \bar{Y}$$

$$\hat{\sigma}^2{}_{MLE} = \frac{\sum_{i=1}^{n_1}(X_i - \hat{\mu}_{1MLE})^2 + \sum_{i=1}^{n_2}(Y_i - \hat{\mu}_{2MLE})^2}{n_1+n_2} = \frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2}$$

Notice that, $\hat{\sigma}_0{}^2{}_{MLE} = \hat{\sigma}^2{}_{MLE} + \frac{n_1 n_2}{(n_1+n_2)^2}(\bar{X}-\bar{Y})^2$

Thus, $\sup_{\Omega} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right) = L\left(\hat{\mu}_{1MLE}, \hat{\mu}_{2MLE}, \hat{\sigma}^2{}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)$

Hence, the likelihood ratio is given by –

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\sup_{\Omega_0} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)}{\sup_{\Omega} L\left(\mu_1, \mu_2, \sigma^2; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

$$= \frac{L\left(\hat{\mu}_{MLE}, \hat{\mu}_{MLE}, \hat{\sigma}_0{}^2{}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)}{L\left(\hat{\mu}_{1MLE}, \hat{\mu}_{2MLE}, \hat{\sigma}^2{}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

On simplification we get, $\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \left(\frac{\hat{\sigma}^2{}_{MLE}}{\hat{\sigma}_0{}^2{}_{MLE}}\right)^{\frac{n_1+n_2}{2}} = \left\{\frac{1}{1+\frac{n_1 n_2}{(n_1+n_2)^2}\left(\frac{\bar{x}-\bar{y}}{\hat{\sigma}_{MLE}}\right)^2}\right\}^{\frac{n_1+n_2}{2}}$

Define, $s^2 = \frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2} = \frac{n_1+n_2}{n_1+n_2-2}\hat{\sigma}^2{}_{MLE}$ and $t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$. Then,

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \{\frac{1}{1+\frac{t^2}{n_1+n_2-2}}\}^{\frac{n_1+n_2}{2}}$$

It is clear that as $|t|$ increases $\lambda(\underset{\sim}{x}, \underset{\sim}{y})$ decreases. Thus, $\lambda(\underset{\sim}{x}, \underset{\sim}{y}) < c \Rightarrow |t| > k$.

Under H_0, $T = \frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ i.e., t distribution with degrees of freedom equal to $n_1 + n_2 - 2$. To find a size α LRT we use the critical point of $t_{n_1+n_2-2}$.

Hence, a size α LRT is given by- 'Reject H_0 iff $|T| = |\frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}| > t_{\alpha/2;n_1+n_2-2}$'. Where, $t_{\alpha/2;n_1+n_2-2}$ is the upper $100\left(1-\frac{\alpha}{2}\right)\%$ point of $t_{n_1+n_2-2}$ distribution.

Hence, we have derived LRT for the alternative $H_1: \mu_1 \neq \mu_2$. Now, we can say that the LRT for the alternative $H_1: \mu_1 > \mu_2$ is based on the test statistic $T = \frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$. If the alternative is true, it is expected that value of T will be positive and large. On the basis of this intuition the size α is given by – 'Reject H_0 iff $T = \frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha;n_1+n_2-2}$'. Where, $t_{\alpha;n_1+n_2-2}$ is the upper $100(1-\alpha)\%$ point of $t_{n_1+n_2-2}$ distribution.

The above test (testing rule) is known as 'Two Sample t-test'.

6.3.1.2. Assumptions and Properties:

We will discuss some important properties of this test (testing rule). The following test is based on certain assumptions –

1. It assumes that the variances of the two-population distribution are equal. i.e., assumption of homogeneity of variance. Which is indeed an important assumption.
2. Here, the test assumes that the probability distribution of the two population is normal distribution.

It is interesting to see the effect on the performance of the test, when at least one of the assumptions is violated.

Now, we will consider example to illustrate how to use this test.

6.3.1.3. Example:

Under normal conditions, is the average body temperature same for men and women?

Medical researchers interested in this question collected data from a large number of men and women, and random samples from that data are presented in the accompanying table (Body Temperature in Degree Fahrenheit).

Men	96.9	97.4	97.5	97.4	97.7	97.8	97.9	98.6	98.5
Women	98.8	97.9	98.5	98.2	98.4	98.8	97.8	98.7	99.3

Is there sufficient evidence to indicate that mean body temperature of women is greater than body temperature of men.

Before solving the problem, we prefer to draw boxplot of the given data. We draw two boxplots for men and women separately. From the plot it seems that the given sets of data come from a symmetric distribution. Also, the variability in temperatures of men and women are more or less same.

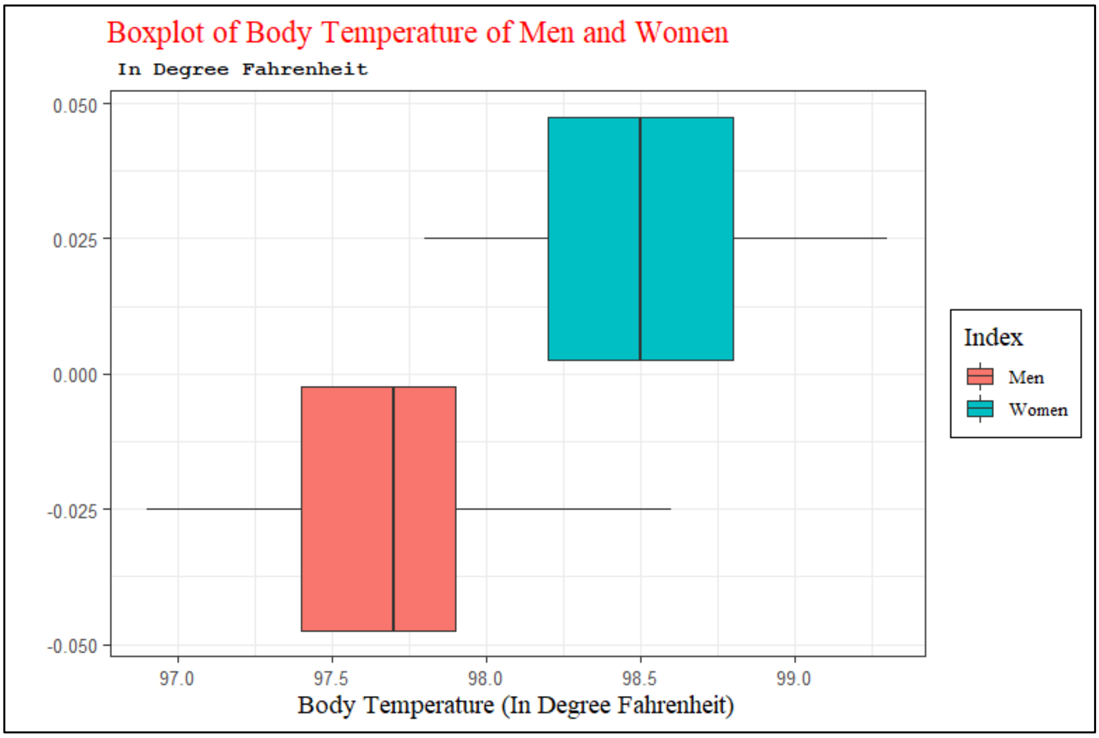


Diagram 8: Boxplot of Body temperature (In Fahrenheit) of men and women for given data

Let, X be a random variable denoting the temperature of a randomly selected women and Y be another random variable denoting the temperature of a randomly selected men. Where, temperatures are measured in degree Fahrenheit.

We assume that $X \sim Normal(\mu_1, {\sigma_1}^2)$ and $Y \sim Normal(\mu_2, {\sigma_2}^2)$ independently. (${\sigma_1}^2$ and ${\sigma_2}^2$ are unknown.)

We want to test H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 > \mu_2$

We draw a random sample $(X_1, X_2, \dots, X_{n_1})$ of size n_1 and a random sample $(Y_1, Y_2, \dots, Y_{n_2})$ of size n_2 independently from the distribution of X and Y respectively. Let us define the following quantities-

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ be respectively the sample means of X and Y.

$S_x^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$ and $S_y^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ respectively be the sampling variance of X and Y.

We further assume that the variances of the two population are equal. i.e., ${\sigma_1}^2 = {\sigma_2}^2$.

To apply the 'Two Sample t-test' we need to calculate the test statistic -

$$T = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Under H_0, $T = \frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

We Reject H_0 iff $T = \frac{\bar{X}-\bar{Y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha;n_1+n_2-2}$. Where, $t_{\alpha;n_1+n_2-2}$ is the upper $100(1-\alpha)\%$ point of $t_{n_1+n_2-2}$ distribution.

In our hypothesis problem, $n_1 = n_2 = 9$

From the given data –

Women (X)	Men(Y)
$\bar{x} = 98.48889$	$\bar{y} = 97.74444$
$s_x^2 = 0.2261111$	$s_y^2 = 0.2927778$

Thus, $s^2 = 0.2594444$, $T_{obs} = 3.10039$.

We choose $\alpha = 0.05$. So, $t_{\alpha;n_1+n_2-2} = t_{0.05;16} = 1.745884$.

As, $T_{obs} > t_{0.05;16}$. We reject H_0.

Hence, in light of the given data it seems that the body temperature of women is greater than body temperature of men.

6.3.2. Population Distribution is Exponential Distribution:

6.3.2.1. Derivation of the test:

When both X and Y population is Exponential population. i.e., The probability distribution of X is $Exponential(Mean = \mu_1)$ and that of Y is $Exponential(Mean = \mu_2)$. We want to test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$. We draw a random sample $(X_1, X_2, \ldots, X_{n_1})$ of size n_1 and a random sample $(Y_1, Y_2, \ldots, Y_{n_2})$ of size n_2 independently from the distribution of X and Y respectively. Let us define the following quantities-

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ be respectively the sample means of X and Y.

Now, the pdf of X is given by –

$$f_X(x) = \frac{1}{\mu_1} e^{-\frac{x}{\mu_1}};\ 0 < x < \infty;\ \mu_1 > 0$$

And the pdf of Y is given by –

$$f_Y(y) = \frac{1}{\mu_2} e^{-\frac{y}{\mu_2}};\ 0 < y < \infty;\ \mu_2 > 0$$

Then the joint pdf of $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ is given by –

$$f(\underset{\sim}{x}, \underset{\sim}{y};\ \mu_1, \mu_2) = \prod_{i=1}^{n_1} f_x(x_i)\ x \prod_{i=1}^{n_2} f_Y(y_i)$$

$$= \frac{1}{\mu_1^{n_1}} e^{-\frac{\sum_{i=1}^{n_1} x_i}{\mu_1}} \frac{1}{\mu_2^{n_2}} e^{-\frac{\sum_{i=1}^{n_2} y_i}{\mu_2}}$$

Thus, likelihood function for the observed data $\underset{\sim}{x}, \underset{\sim}{y}$ is given by –

$$L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right) = f\left(\underset{\sim}{x}, \underset{\sim}{y}; \mu_1, \mu_2\right)$$

Now, in this hypothesis problem-

Parameter space under H_0 is $\Omega_0 = \{(\mu_1, \mu_2) : \mu_1 = \mu_2 = \mu \in \mathbb{R}^+\}$

Parameter space under H_1 is $\Omega_1 = \{(\mu_1, \mu_2) : \mu_1 > \mu_2 \in \mathbb{R}^+\}$

Thus, the parameter space of interest is –

$\Omega = \{(\mu_1, \mu_2) : \mu_1 \geq \mu_2 \in \mathbb{R}^+\}$

Now, we have to compute the ratio –

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\sup_{\Omega_0} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)}{\sup_{\Omega} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

Thus, to find the Likelihood Ratio Test we have to maximize the likelihood $L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)$ under Ω_0 and Ω. But the likelihood maximization problem under Ω is difficult, because it is not an open rectangle type parameter space. It would be easier if Ω is an open rectangle.

Thus, here also we will find LRT for the alternative $H_1 : \mu_1 \neq \mu_2$ and then use the test statistic for $H_1 : \mu_1 > \mu_2$. It will suffice our purpose. So, we will do this.

For the alternative $H_1 : \mu_1 \neq \mu_2$, parameter space under H_1 is given by -

$\Omega_1 = \{(\mu_1, \mu_2) : \mu_1 \neq \mu_2 \in \mathbb{R}^+\}$

Thus, parameter space of interest is $\Omega = \{(\mu_1, \mu_2) : \mu_1, \mu_2 \in \mathbb{R}^+\}$

To find $\sup_{\Omega_0} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)$, we need to find MLEs of μ_1, μ_2 over Ω_0.

Under H_0, $\mu_1 = \mu_2 = \mu$. Hence, $\hat{\mu}_{MLE} = \frac{n_1\bar{X} + n_2\bar{Y}}{n_1 + n_2}$

Thus, $\sup_{\Omega_0} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right) = L\left(\hat{\mu}_{MLE}, \hat{\mu}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)$

Now to find $\sup_{\Omega} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)$, we need to find MLEs of μ_1, μ_2 over Ω. Now, the MLEs are given by –

$$\hat{\mu}_{1_{MLE}} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i = \bar{X}$$

$$\hat{\mu}_{2_{MLE}} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i = \bar{Y}$$

Thus, $\sup_{\Omega} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right) = L\left(\hat{\mu}_{1_{MLE}}, \hat{\mu}_{2_{MLE}}; \underset{\sim}{x}, \underset{\sim}{y}\right)$

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\sup_{\Omega_0} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)}{\sup_{\Omega} L\left(\mu_1, \mu_2; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

$$= \frac{L\left(\hat{\mu}_{MLE}, \hat{\mu}_{MLE}; \underset{\sim}{x}, \underset{\sim}{y}\right)}{L\left(\hat{\mu}_{1_{MLE}}, \hat{\mu}_{2_{MLE}}; \underset{\sim}{x}, \underset{\sim}{y}\right)}$$

On simplification, $\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = Constant\ (\frac{\hat{\mu}_{1_{MLE}}}{\hat{\mu}_{MLE}})^{n_1} (\frac{\hat{\mu}_{2_{MLE}}}{\hat{\mu}_{MLE}})^{n_2}$

$$= Constant\ \left(\frac{n_1\bar{x}}{n_1\bar{x} + n_2\bar{y}}\right)^{n_1} \left(\frac{n_2\bar{y}}{n_1\bar{x} + n_2\bar{y}}\right)^{n_2}$$

$$= Constant \left(\frac{n_1\bar{x}}{n_1\bar{x} + n_2\bar{y}}\right)^{n_1} \left(1 - \frac{n_1\bar{x}}{n_1\bar{x} + n_2\bar{y}}\right)^{n_2}$$

Now, Define t = $\frac{n_1\bar{x}}{n_1\bar{x} + n_2\bar{y}}$

Then, $\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = Constant\ t^{n_1}(1 - t)^{n_2}$ (The Constant is Positive)

Now, $h(t) = t^{n_1}(1 - t)^{n_2}$

We will study the nature of these functions through derivatives.

$$h'(t) = t^{n_1-1}(1 - t)^{n_2-1}(n_1 + n_2)\left(\frac{n_1}{n_1 + n_2} - t\right)$$

So, $h'(t) = \begin{cases} > 0\ if\ t < \frac{n_1}{n_1+n_2} \\ < 0\ if\ t > \frac{n_1}{n_1+n_2} \end{cases}$

Thus, for $t < \frac{n_1}{n_1+n_2}$, $h(t)$ increases and for $t > \frac{n_1}{n_1+n_2}$, $h(t)$ decreases.

So, $g(t) < C \Rightarrow t < C_1\ or, t > C_2$

We will further proceed. Note that, $t = \frac{\frac{n_1\bar{x}}{n_2\bar{y}}}{1 + \frac{n_1\bar{x}}{n_2\bar{y}}} = \frac{t_1}{t_1+1}$

Now, $\frac{dt}{dt_1} = \frac{(t_1+1)-t_1}{(1+t_1)^2} = \frac{1}{(1+t_1)^2} > 0 \Rightarrow t$ is increasing in t_1

Thus, from the above study of the function $g(t)$, we can say that –

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) < C \Rightarrow t < C_1\ or, t > C_2 \Rightarrow t_1 < K_1\ or, t_1 > K_2$$

That is, $\frac{\bar{x}}{\bar{y}} < k_1\ or,\ \frac{\bar{x}}{\bar{y}} > k_2$

Now, we have to find the distribution of the test statistic. By *Additive Property* of Exponential Distribution, we have -

$\frac{2}{\mu_1}\sum_{i=1}^{n_1} X_i \sim Gamma(\frac{1}{2}, \frac{2n_1}{2}) \equiv \chi^2_{2n_1}$ and $\frac{2}{\mu_2}\sum_{i=1}^{n_2} Y_i \sim Gamma(\frac{1}{2}, \frac{2n_2}{2}) \equiv \chi^2_{2n_2}$ independently. Thus, $\frac{\bar{X}}{\bar{Y}}\frac{\mu_2}{\mu_1} \sim F_{2n_1,2n_2}$.

Under $H_0: \mu_1 = \mu_2$. Hence, $\frac{\bar{X}}{\bar{Y}} \sim F_{2n_1,2n_2}$ i.e., F distribution with degrees of freedom $2n_1, 2n_2$.

Thus, a size α LRT is given by – 'Reject H_0 iff $\frac{\bar{X}}{\bar{Y}} < k_1\ or, \frac{\bar{X}}{\bar{Y}} > k_2$'. Where, k_1 and k_2 is so chosen that $P_{H_0}(\frac{\bar{X}}{\bar{Y}} < k_1) = \alpha_1\ and\ P_{H_0}(\frac{\bar{X}}{\bar{Y}} > k_2) = \alpha_2. Where, \alpha_1 + \alpha_2 = \alpha.$

We choose $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. Using critical points of $F_{2n_1,2n_2}$ distribution, we determine k_1 and k_2. Thus, $k_1 = F_{1-\frac{\alpha}{2};2n_1,2n_2}$ and $k_2 = F_{\frac{\alpha}{2};2n_1,2n_2}$.

Hence, we have derived LRT for the alternative $H_1: \mu_1 \neq \mu_2$. Now, we can say that the LRT for the alternative $H_1: \mu_1 > \mu_2$ is based on the test statistic $T = \frac{\bar{X}}{\bar{Y}}$. If the alternative is true, it is expected that value of T will be large. On the basis of this intuition the size α is given by –

‘Reject H_0 iff $\frac{\bar{X}}{\bar{Y}} > F_{\alpha;2n_1,2n_2}$’. Where, $F_{\alpha;2n_1,2n_2}$ is the upper $100(1-\alpha)\%$ point of $F_{2n_1,2n_2}$ distribution.

6.3.2.2. Assumptions and Properties:

This test (testing rule) is based on the assumption that the probability distribution of the two population is exponential distribution. It is interesting to see the effect on the performance of the test, if the population distribution is not exponential distribution.

Now, we will consider example to illustrate how to use this test.

6.3.2.3. Example:

Two company A and B produce bulbs. Company A claims that their produced bulbs have more lifetime than the bulbs produced by company B. We want to know whether the claim of company A is valid. For that a sample of bulbs are drawn from the bulbs produced by the two companies independently and under essentially similar condition their lifetimes (in hours) are noted. The data is presented in the table below.

Company A	25.7	13.6	15.1	30.2	8.8	X	X
Company B	39.3	14.2	24.9	7.5	10.4	12.2	20.6

Is there sufficient evidence to indicate that the claim of company A is justified?

Before solving the problem, we prefer to draw boxplot of the given data. We draw two boxplots for company A and company B separately. From the plot it seems that the given sets of data come from a positively skewed distribution.

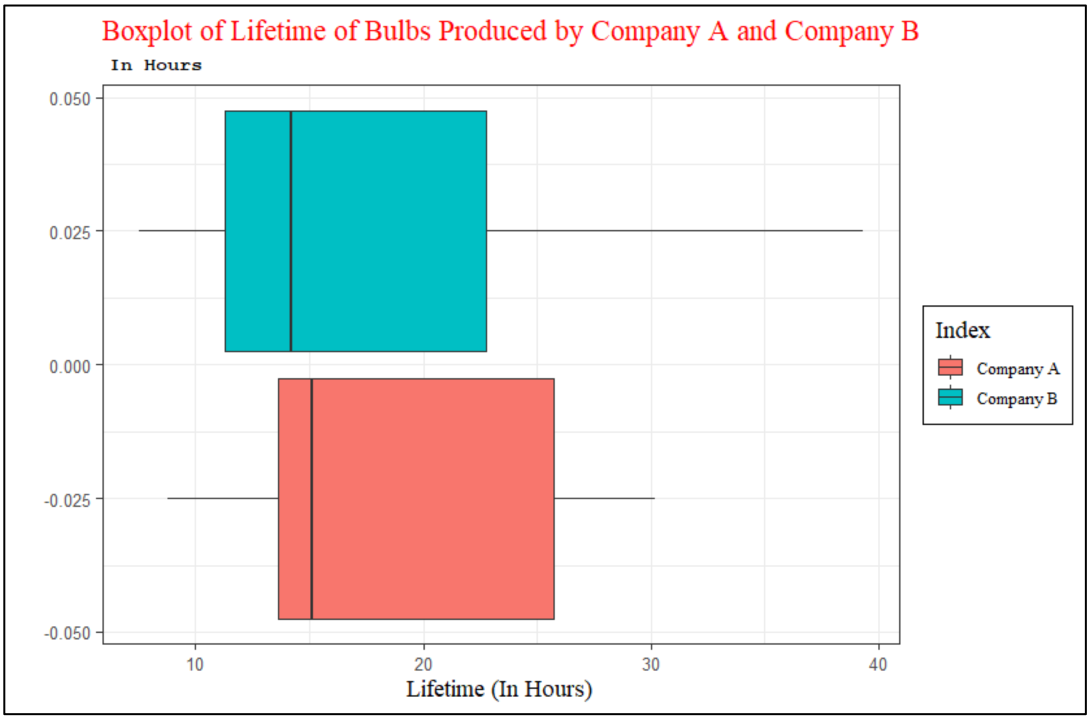


Diagram 9: Boxplot of Lifetime of bulbs produced by company A and B for given data

Let, X be a random variable denoting the lifetime (in hours) of a randomly selected Bulb produced by company A and Y be another random variable denoting the lifetime (in hours) of a randomly selected Bulb produced by company B.

We assume that $X \sim Exponential(Mean = \mu_1)$ and $Y \sim Exponential(Mean = \mu_2)$ independently.

We want to test H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 > \mu_2$

We draw a random sample $(X_1, X_2, \ldots, X_{n_1})$ of size n_1 and a random sample $(Y_1, Y_2, \ldots, Y_{n_2})$ of size n_2 independently from the distribution of X and Y respectively. Here, the given datasets are our sample.

Let us define the following quantities-

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ be respectively the sample means of X and Y.

We will use the likelihood test, that we have derived just. We will calculate the test-statistic.

$$T = \frac{\bar{X}}{\bar{Y}}$$

Under H_0: $\mu_1 = \mu_2$. Hence, $\frac{\bar{X}}{\bar{Y}} \sim F_{2n_1, 2n_2}$ i.e., F distribution with degrees of freedom $2n_1, 2n_2$.

In our hypothesis problem, $n_1 = 5, n_2 = 7$

On the basis of the given data, $\bar{x} = 18.68$, $\bar{y} = 18.44286$.

Thus,$T_{obs} = 1.012858$

We choose $\alpha = 0.05$. So, $F_{\alpha;2n_1,2n_2} = F_{0.05;10,14} = 2.602155$.

As, $T_{obs} < F_{0.05;10,14}$. We accept H_0.

Hence, in the light of the given data it seems that the lifetime of the bulbs produced by company A and company B are same. Hence, the claim of the company A is not justified.

6.4. Non-Parametric Tests for Two Sample Problems:

6.4.1. Introduction:

Here, we will consider non-parametric tests for two-sample problems. A variety of tests are available. But we will restrict our attention to –

- Kolmogorov-Smirnov Test
- Mann Whitney U Test
- Wilcoxon Rank Sum Test

Since, Non-parametric tests are distribution free tests i.e., no assumption of probability distribution is required (e.g., population distribution Normal).

Here, we have two statistical populations. We call them as X population and Y population, with cumulative distribution functions denoted by F_X and F_Y respectively. Then we draw a random sample of size n_1 from the X population and random sample of size n_2 from Y population. They are denoted by – $(X_1, X_2, \dots, X_{n_1})$ and $(Y_1, Y_2, \dots, Y_{n_2})$. In non-parametric tests we consider the following null hypothesis- H_0: $F_x(x) = F_y(x)$.

6.4.2. Equivalence Between Hypothesis in parametric setup and non-parametric setup:

Before we start our discussion of different non-parametric tests, it is important to understand the equivalence between hypothesis in parametric setup and non-parametric setup.

In our discussion, we have only considered two probability distributions.

- Normal
- Exponential

We want to test the hypothesis concerning equality of mean of the two independent population distributions against specific alternative hypothesis under two different situations. When, both the populations are normally distributed and when both the populations are exponentially distributed.

However, in non-parametric setup we have no such parameters. Here, we define the hypothesis in terms of distribution function (CDF). H_0: $F_x(x) = F_y(x)$.

We will establish the equivalence between these two hypotheses.

Suppose, the populations under consideration are normally distributed. Then, $X \sim Normal(\mu_1, {\sigma_1}^2)$ and $Y \sim Normal(\mu_2, {\sigma_2}^2)$ independently. (${\sigma_1}^2$ and ${\sigma_2}^2$ are unknown.) We further assume that ${\sigma_1}^2$ and ${\sigma_2}^2$ are equal (${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$). Now, the CDFs of X and Y are given by –

$$F_X(x) = \Phi\left(\frac{x-\mu_1}{\sigma}\right) \forall x \in \mathbb{R}$$

$$F_Y(y) = \Phi\left(\frac{y-\mu_2}{\sigma}\right) \forall y \in \mathbb{R}$$

Thus, $F_x(x) = F_y(x) \, \forall x \in \mathbb{R}$, essentially means that $\mu_1 = \mu_2$. And if we consider the alternative $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x. Then, the alternative implies that -

$\Rightarrow \Phi\left(\frac{x-\mu_2}{\sigma}\right) > \Phi\left(\frac{x-\mu_1}{\sigma}\right) \forall x \in \mathbb{R}$

$\Rightarrow \frac{x-\mu_2}{\sigma} > \frac{x-\mu_1}{\sigma} \forall x \in \mathbb{R}$

$\Rightarrow \mu_1 > \mu_2$

Thus, the hypothesis in non-parametric set up boils down to hypothesis in parametric setup. Hence, we have established the equivalence of parametric hypothesis and non-parametric hypothesis, when the population distributions are normal. Similarly, we can establish the equivalence, when the population distributions are exponential.

So, we conclude that, in our discussion, the hypothesis problem $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$ in parametric setup is equivalent to the hypothesis problem $H_0: F_x(x) = F_y(x)$ against $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x.

Now, we will discuss about different non-parametric tests. The first test that we will discuss is Kolmogorov-Smirnov Test.

6.4.3. Kolmogorov-Smirnov Test:

6.4.3.1. Introduction:

Kolmogorov Smirnov test is an example of non-parametric test for two sample problems. The main assumption of the test is that the populations under consideration must have continuous distribution functions.

6.4.3.2. Test Statistic:

We are interested in comparing the empirical distribution functions of the two samples drawn. For that we will define the order statistics. Let the order statistics corresponding to two samples are denoted by –

$$(X_{(1)}, X_{(2)}, \ldots, X_{(n_1)}) \text{ and } (Y_{(1)}, Y_{(2)}, \ldots, Y_{(n_2)})$$

For the random sample of size n_1 from the population X, the empirical distribution function is given by –

$$S_{n_1}(x) = \begin{cases} 0 \; if \; x < X_{(1)} \\ k/n_1 \; if \; X_{(k)} \le x < X_{(k+1)} \\ 1 \; if \; x \ge X_{(n_1)} \end{cases} \qquad for \; k = 1,2,3,\ldots,n_1 - 1$$

For the random sample of size n_2 from the population Y, the empirical distribution function is given by –

$$S_{n_2}(x) = \begin{cases} 0 \; if \; x < Y_{(1)} \\ k/n_2 \; if \; Y_{(k)} \le x < Y_{(k+1)} \\ 1 \; if \; x \ge Y_{(n_2)} \end{cases} \qquad for \; k = 1,2,3,\ldots,n_2 - 1$$

In the combined ordered arrangement of $n_1 + n_2$ samples observations $S_{n_1}(x)$ and $S_{n_2}(x)$ are respective proportion of X and Y observations which don't exceed the value x. If H_0 is true, then we have two samples come from same population. Then, there must be a reasonable agreement between the two empirical CDFs. i.e., the two empirical CDFs must be more or less identical. If the empirical CDFs are very much different, then it is an indication that the null hypothesis is not true. This is the intuitive logic behind the Kolmogorov-Smirnov Test. Thus, the test statistic based on the empirical CDFs must be formed in such a way that, it reflects the above facts.

Here, our alternative hypothesis is H_1: $F_Y(x) \ge F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x. The Kolmogorov-Smirnov test statistic for this one-sided alternative is given by –

$$D_{n_1,n_2} = \max_x (S_{n_2}(x) - S_{n_1}(x))$$

Some authors use the notation $D^+_{n_1,n_2}$ to denote D_{n_1,n_2}. It is to be noted that the test statistic changes depending upon our alternative hypothesis.

If the alternative is true, it expected that the value of D_{n_1,n_2} will be large. Thus, we reject H_0 iff $D_{n_1,n_2} \ge c_\alpha$. Here, $P_r(D_{n_1,n_2} \ge c_\alpha \,|H_0) \le \alpha$. ($\alpha$ is our chosen level of significance)

The P-Value of is given by –

$$P_r(D_{n_1,n_2} \ge D_0 \,|H_0)$$

Where D_0 is the observed value of the two sample Kolmogorov-Smirnov statistic D_{n_1,n_2}. Under H_0 the distribution of D_{n_1,n_2} is independent of the population distribution. Hence it can be attributed as a distribution free method.

In order to obtain level and power of the test we need to know the distribution of D_{n_1,n_2}. It is very difficult to obtain the distribution D_{n_1,n_2} under H_0. A recursive relation may be used to obtain a relationship between different probabilities. Graphical methods are also used to obtain the distribution of D_{n_1,n_2}under H_0. The derivation of the exact null probability distribution of $D_{n_1,n_2}$is usually attributed to $Gnedenko$ (1954) $and\ Korolyuk$ (1961), papers by $Massey$ (1951b, 1952) are important. $Hodges$ (1958) summarized several approaches for finding the null distribution of D_{n_1,n_2}. The derivation of exact distribution of D_{n_1,n_2}, when null hypothesis is not true is difficult. Simulation techniques may be used to study the distribution D_{n_1,n_2} in those cases. We will use Tables to compute the P-Value of the Test.

6.4.3.3. Computation of the test statistic and decision:

To compute the test statistic D_{n_1,n_2} for observed sample ($x_1, x_2, \ldots, x_{n_1}$) and $(y_1, y_2, \ldots, y_{n_2})$, we need to find the empirical cumulative distribution for both the samples, which are denoted by $S_{n_1}(x), S_{n_2}(x)$. Then, we will arrange the combined sample (z) in increasing order. After that we will compute the differences $S_{n_2}(z) - S_{n_1}(z)$ for each value of z in the ordered combined sample. Then we will calculate $\max_z(S_{n_2}(z) - S_{n_1}(z))$. This will give us the value of the test-statistic D_{n_1,n_2}. Which is actually the observed value (D_0) of the Kolmogorov-Smirnov test-statistic for the observed sample. From the Tables of null distribution D_{n_1,n_2}, we calculate $P_r(D_{n_1,n_2} \geq D_0 \mid H_0)$. If the probability is less than α. (α is our chosen level of significance). Then we reject H_0, otherwise we accept the H_0.

6.4.3.4. Asymptotic Distribution of D_{n_1,n_2} under H_0:

The asymptotic distribution of the test statistic may be obtained under H_0. Due to Gilvenko-Cantelli Theorem the asymptotic distribution of $\sqrt{\frac{n_1 n_2}{n_1+n_2}} D_{n_1,n_2}$ is given by –

$$\lim_{n_1,n_2\to\infty} P\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}} D_{n_1,n_2} \leq d\right) = 1 - e^{-2d^2}$$

This approximation is valid for large values of n_1, n_2.

6.4.3.5. Problem of Ties:

Ties within and across samples are handled by considering r distinct values only. However, when the assumption of continuous distribution is valid, it is hardly to get same values. But if the assumption of continuous distribution is not valid, the use of the test statistic may reduce the performance of the Kolmogorov-Smirnov test.

6.4.3.6. Properties:

This test is very easy to apply. Using tables, we can perform the test very easily and for large sample sizes we can use asymptotic distribution. This test can be used to test general types of alternatives -

- $H_1 : F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x
- $H_1 : F_Y(x) \leq F_X(x)$ for all x and $F_Y(x) < F_X(x)$ for some x
- $H_1 : F_Y(x) \neq F_X(x)$ for some x

Since, the test is sensitive to any type of difference in cumulative distribution functions of the two population under consideration, we mainly use this test for preliminary studies of the data. The Kolmogorov test is more powerful than Runs test (It is another example of non-parametric test for two sample problem) when compared against *Lehman alternative*. It is to be noted that the large sample performance of the test against specific location or scale alternative varies considerably according to the population sampled.

6.4.3.7. Example:

Consider the example we have given in case of LRT of exponential distribution. If we want to use Kolmogorov-Smirnov test for testing the null hypothesis of equality of means, then we have to first calculate the empirical CDFs for the two samples. We will use the same notations explained in the previous example (Exponential LRT).

We plot the empirical CDFs on the same graph, to get an understanding.

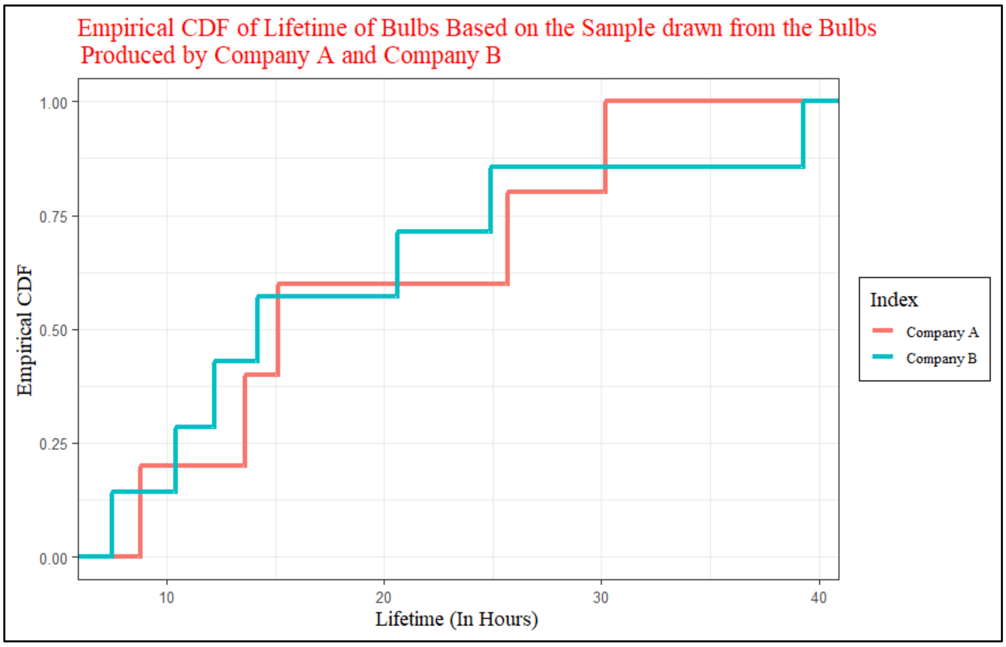


Diagram 10: Empirical CDF of lifetime of Bulbs based on the sample drawn from the bulbs produced by company A and B

From the graph, we can see there is an agreement between the two empirical CDFs, which is somehow an indication of identical distribution. Let's see what does the test says. Notice the assumption of no ties is justified here.

Let, $S_{n_1}(x), S_{n_2}(x)$ respectively denote the empirical CDF of lifetime of bulbs of the random samples drawn from the population of bulbs produced by company A and company B. Here, $n_1 = 5, n_2 = 7$. Now, we find the ordered arrangement of the combined sample. We denote it by z. Then, to calculate the Kolmogorov-Smirnov test statistic, we have to calculate $\max_z(S_{n_2}(z) - S_{n_1}(z))$. For that we prepare a table as below.

Table 1: Calculation Related to Kolmogorov-Smirnov Statistic

z Combined Sample	$S_{n_1}(z)$ Value of Empirical CDF of X Sample for Combined Sample	$S_{n_2}(z)$ Value of Empirical CDF of Y Sample for Combined Sample	$S_{n_2}(z) - S_{n_1}(z)$ Difference Between Empirical CDFs
7.5	0.0	0.143	0.143
8.8	0.2	0.143	-0.057

10.4	0.2	0.286	0.086
12.2	0.2	0.429	0.229
13.6	0.4	0.429	0.029
14.2	0.4	0.571	0.171
15.1	0.6	0.571	-0.029
20.6	0.6	0.714	0.114
24.9	0.6	0.857	0.257
25.7	0.8	0.857	0.057
30.2	1.0	0.857	-0.143
39.3	1.0	1.000	0.000

From the above table we can clearly see that, the value of the test statistic $D_{n_1,n_2} = \max_z(S_{n_2}(z) - S_{n_1}(z))$ is given by for $z = 24.9$. So, $D_{n_1,n_2} = 0.257$.

From the table of critical values of Kolmogorov-Smirnov test, we find that the critical value for the above test with $n_1 = 5, n_2 = 7$ for level of significance 0.05 is given by $23/35 = 0.6571429$. Hence, $D_{n_1,n_2} = 0.257 < 0.6571429$. Thus, we accept H_0 at level of significance 0.05.

Hence, in the light of the given data it seems that the lifetime of the bulbs produced by company A and company B are same. Hence, the claim of the company A is not justified.

6.4.4. Mann Whitney U Test:

6.4.4.1. Introduction:

Mann Whitney U test is different from Kolmogorov-Smirnov Test. Mann Whitney U test is based on the pattern exhibited in the ordered arrangement of combined sample. If H_0 is true, then it is expected that in the combined ordered sample X and Y values are distributed in random order. But if H_0 is not true, then there must be some pattern exhibited in the ordered arrangement of X and Y sample values. However, Mann Whitney test not only captures this pattern exhibited in the sample, it also takes into account the magnitude of say, the Y's in relation to the X's, i.e., the position of the Y's in the combined ordered sequence. If the sample Y values are mostly greater than sample X values or vice versa, or both, then it is evidence against the null hypothesis of identical distributions.

6.4.4.2. Test Statistic:

The Mann Whitney U test statistic is defined as the number of times a Y precedes an X in the combined ordered arrangement of the two independent random samples. We have already mentioned that our random samples are denoted by $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$. Like in case of Kolmogorov-Smirnov Test, here also we assume that the samples are drawn from continuous distribution. So, the possibility of $X_j = Y_i$ (for some i and j) need not be considered. We define $n_1 n_2$ indicator variables as –

$$D_{ij} = \begin{cases} 1 & if\ Y_j < X_i \\ 0 & if\ Y_j > X_i \end{cases} \quad for\ i = 1,2,\ldots,n_1\ and\ j = 1,2,\ldots,n_2$$

Then, Mann Whitney U statistic is defined as –

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_{ij}$$

Our alternative hypothesis is $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x. Then a logical rejection would be 'Reject H_0 iff $U > c_\alpha$'. Where, c_α is so chosen that- $P_r(U \geq c_\alpha | H_0) \leq \alpha$. ($\alpha$ is our chosen level of significance).

The P-Value is given by –

$$P_r(U \geq U_0 | H_0)$$

Where, U_0 is the observed value of the two sample Mann Whitney U statistic.

If H_1 is true, then it is expected that in the sample most of X values are larger than Y values. That is, value of U will be larger in that case. The fact that it is consistent test criterion can be shown by investigating the convergence of $\frac{U}{n_1 n_2}$ to a certain parameter. Actually, we can redefine the null hypothesis in terms of that parameter.

We define,

$$p = P(Y < X)$$

Then, our null hypothesis can be written as $H_0: p = 0.5$. Because, under H_0 X and Y random variables are identically and independently distributed. Thus under H_0, $P(Y < X)$ can be written as –

$$P(Y < X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_y(y) f_x(x) dy\, dx$$

$$= \int_{-\infty}^{\infty} F_X(x) dF_Y(x) = \frac{1}{2}$$

In our hypothesis testing problem, the alternative hypothesis is $H_1: p > 0.5$.

Now, we will try to find the moments of the distribution of U in general situation. Here, D_{ij}'s are Bernoulli random variables. The first and second order raw moments are defined as –

$$E(D_{ij}) = E(D_{ij}^2) = p$$

$$Var(D_{ij}) = p(1-p)$$

To find covariances we note that Dij's are not independent whenever the X subscripts or Y subscripts are common. Thus,

$Cov(D_{ij}, D_{hk}) = 0 \ for \ all \ i \neq h \ and \ j \neq k$

$Cov(D_{ij}, D_{ik}) = p_1 - p^2 \ for \ all \ j \neq k$

$Cov(D_{ij}, D_{hj}) = p_2 - p^2 \ for \ all \ i \neq h$

Where, p_1 and p_2 are defined as –

$$p_1 = P(Y_j < X_i \cap Y_k < X_i)$$

$$p_2 = P(Y_j < X_i \cap Y_j < X_k)$$

Since, U is defined as sum of D_{ij}'s, hence substituting the expressions obtained above we have –

$$E(U) = n_1 n_2 p$$

$$Var(U) = Var\left(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} D_{ij}\right)$$

$$= n_1 n_2 [p - p^2(n_1 + n_2 - 1) + (n_2 - 1)p_1 + (n_1 - 1)p_2]$$

Since, $E(U/n_1 n_2) = p$ and $Var(U/n_1 n_2) \to 0$ as $n_1, n_2 \to \infty$. Thus, $U/n_1 n_2$ is a consistent estimator of p. Thus, test based on U is obviously good.

Now, to obtain level α critical region of the test, we need to know the distribution of U under H_0. Now, derivation of the general distribution of U is difficult. Under H_0, each of the $\binom{n_1+n_2}{n_1}$ possible ordered arrangements of the random variables in the combined sample are equally likely. Hence,

$$P(U = u) = \frac{r_{n_1,n_2}^u}{\binom{n_1+n_2}{n_1}}$$

Where, r_{n_1,n_2}^u is the number of distinguishable arrangements of n_1 $X's$ and n_2 $Y's$ such that in each arrangement, the number of times Y precedes a X is exactly u. For small sample sizes the probabilities may be obtained using recursion relation. If the sample with fewer observations is always labelled the X sample, tables are needed only for $n_1 \leq n_2$ and left-tail critical points. Such tables are widely available, for example in Auble (1953) or Mann and Whitney (1947).

Clearly, U can take every integer value between 0 and $n_1 n_2$. Hence, Large value of U indicates X takes on an average large value than Y, while small values indicate Y takes on an average large value than X. It can be shown that, under H_0 the distribution of U is symmetric about $\frac{n_1 n_2}{2}$. Thus, $P(U - \frac{n_1 n_2}{2} = u) = P(U - \frac{n_1 n_2}{2} = -u)$ for all u.

For large sample sizes we can use asymptotic probability distribution of U. But for that we need to obtain mean and variance of U under H_0. We have already obtained a general expression for mean and variance. Under H_0, $p = \frac{1}{2}, p_1 = \frac{1}{3}, p_2 = \frac{1}{3}$. so, we get-

$$E(U) = \frac{n_1 n_2}{2} \text{ and } Var(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Using Central Limit Theorem, it can be shown that under H_0-

$$T = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \xrightarrow{L} Z \sim Normal(0,1)$$

This approximation has been found good for small sample sizes also. The reason for such good approximation is that the distribution of U is symmetric under H_0. Since, U can take only integer values, a continuity correction of 0.5 is required to calculate any probability using this approximation.

For the alternative hypothesis $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x, using asymptotic distribution the testing rule will be 'Reject H_0 iff T > τ_α'. Where, τ_α is the upper 100(1- α)% point of $Normal(0,1)$ distribution.

6.4.4.3. Problem of Ties:

When a large number of ties are present in the sample, we need to reformulate the Mann-Whitney U statistic. Since, in our discussion that part is not needed, we are omitting that part. However, one should understand that the Mann-Whitney U statistic we have defined does not takes into account ties. Hence, if the assumption of continuous population distribution is not valid, the use of the statistic in that case may reduce the performance of the test.

6.4.4.4. Properties:

Mann Whitney U test is the most frequently used test for two sample problem. Only independence and continuous distribution need to be assumed to apply the test successfully. This test is mainly used for the null hypothesis of identical distributions. The test is simple to use for any sample size and tables of exact null distributions are widely available. Even the asymptotic distribution can be used for small sample sizes also. The large sample approximation is quite useful and in case of presence of ties, the corrections may be incorporated to find the test statistic. The test is good for the particular tests like – equality of means, equality of medians etc. It has been observed that, the test performs well when the population distributions under consideration are from same location family i.e., the two populations differ with respect to location only. Hence, we can compare this test with corresponding parametric tests of locations also.

6.4.4.5. Example:

Consider the example, that we have given in case of LRT of exponential distribution. we will use Mann Whitney U test for testing the null hypothesis of equality of means. We will use the same notations explained in the example of Exponential LRT. Here, $n_1 = 5, n_2 = 7$. Notice, the assumption of no ties is valid here.

First, we will consider the combined ordered arrangement of the two samples. We will keep in mind which are X sample values and which are Y sample values. Then, we will calculate the number of times a Y precedes a X in the combined ordered arrangement of the two independent random samples. For that we construct the following table.

Table 2: Calculation related to Mann Whitney U Test

X Sample Value	Y that precedes X	Number of Y's that precedes X
8.8	7.5	1
13.6	7.5, 10.4, 12.2	3
15.1	7.5, 10.4, 12.2, 14.2	4
25.7	7.5, 10.4, 12.2, 14.2, 20.6, 24.9	6
30.2	7.5, 10.4, 12.2, 14.2, 20.6, 24.9	6

From the above table we can see that, the value of Mann Whitney U statistic is simply the sum of the last column values. Thus, $U = 20$.

From the table of critical values of Mann Whitney U test, we find that the critical value for the above test with $n_1 = 5, n_2 = 7$ for level of significance 0.05 is given by 28. Hence, $U = 20 < 28$. Thus, we accept H_0 at level of significance 0.05.

Now, we will apply large sample approximation to conduct the test. We will calculate the statistic -

$$T = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

For the given data, we get $T_{obs} = 0.405999$. Now, the critical point for level 0.05 test is $\tau_\alpha = \tau_{0.05} = 1.644854$. Which is quite large than T_{obs}. Hence, here also we reject H_0, as $T_{obs} < \tau_{0.05}$.

So, in both cases we get the same decision. Hence, in the light of the given data it seems that the lifetime of the bulbs produced by company A and company B are same. Hence, the claim of the company A is not justified.

6.4.5. Wilcoxon Rank-sum Test:

6.4.5.1. Introduction:

Wilcoxon Rank-sum Test is another non-parametric test and the last non-parametric test, we will consider in our discussion. It is very much related with Mann-Whitney U test. Like in case of Kolmogorov-Smirnov Test and Mann Whitney U Test, here also we assume that the samples are drawn from continuous distribution. So, the possibility of $X_j = Y_i$ (for some i and j) need not be considered. Wilcoxon Rank-sum test is defined as function of *Linear-rank statistics*. We will give a short note on linear rank statistics here.

6.4.5.2. Test Statistic

Many two sample non-parametric tests are based on linear combination of certain indicator variables for the combined ordered sample. Such functions are called *Linear Rank Statistics*.

Suppose we have two independent random samples denoted by – $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ from two independent population X and Y with CDFs F_X and F_Y respectively. Under the null hypothesis of identical distributions, we have $N = n_1 + n_2$ observations from the common but unknown population. We order them in increasing order.

We define the following indicator random variables –

$$Z_i = \begin{cases} 1 & if\ ith\ observation\ in\ the\ ordered\ sample\ is\ from\ X\ population \\ 0 & if\ ith\ observation\ in\ the\ ordered\ sample\ is\ from\ Y\ population \end{cases}$$

$for\ all\ i = 1\ to\ N$

Then, we define the following statistic $T = \sum_{i=1}^{N} a_i Z_i$. Where, a_i are given constants, weights or scores. For example, we can choose $a_i = i$, which is basically rank of i^{th} observation in the combined ordered sample.

For different choice of a_i, we get different statistics. These all choices of statistics together create a class of statistics, called Linear-Rank Statistics.

Wilcoxon Rank-Sum Statistic is a linear rank statistic and as the name suggests, it is sum of ranks of observations. The Wilcoxon-Rank sum statistic is defined as -

$$W_N = \sum_{i=1}^{N} iZ_i$$

Where, Z_i is the indicator variable as defined earlier. Higher values of W_N indicates that most of large values are from X population and lower values of W_N indicates that most of the large values are from Y population. Thus, for testing the null hypothesis $H_0: F_x(x) = F_y(x)$ against the alternative hypothesis $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x. Then a logical rejection would be 'Reject H_0 iff $W_N > c_\alpha$'. Where, c_α is so chosen that- $P_r(W_N \geq c_\alpha \,|H_0) \leq \alpha$. ($\alpha$ is our chosen level of significance).

The P-Value of is given by –

$$P_r(U \geq W_N^0 \,|H_0)$$

Where, W_N^0 is the observed value of the two sample Wilcoxon Rank-sum statistic.

If there are no ties, the exact mean and variance of the statistic W_N under null hypothesis can be easily found. Under null hypothesis it can be shown that –

$$E(Z_i) = P(Z_i = 1) = \frac{n_1}{N}$$

$$Var(Z_i) = \frac{n_1}{N}\left(1 - \frac{n_1}{N}\right)$$

$$Cov(Z_i, Z_j) = -\frac{n_1 n_2}{N^2(N-1)} \;\forall i \neq j$$

Then, we have –

$$E(W_N) = \sum_{i=1}^{N} iE(Z_i) = \frac{n_1}{N}\sum_{i=1}^{N} i = \frac{n_1(N+1)}{2}$$

$$Var(W_N) = Var\left(\sum_{i=1}^{N} iZ_i\right) = \sum_{i=1}^{N} i^2 Var(Z_i) + \sum_{1\leq i\neq j\leq N} Cov(Z_i, Z_j)$$

$$= \frac{n_1 n_2(N+1)}{12}$$

It can be shown that under H_0 the distribution of W_N is symmetric about its mean. Thus, $P(W_N - \frac{n_1(N+1)}{2} = w) = P(W_N - \frac{n_1(N+1)}{2} = -w)$ for all w. The statistic W_N takes every integer value between $\sum_{i=1}^{n_1} i = \frac{n_1(1+n_1)}{2}$ and $\sum_{i=n_2+1}^{n_1+n_2} i = \frac{2N-n_1+1}{2}$. W_N takes the value $\sum_{i=1}^{n_1} i = \frac{n_1(1+n_1)}{2}$, when all $X's$ are small than all $Y's$ and W_N takes the value $\sum_{i=n_2+1}^{n_1+n_2} i = \frac{2N-n_1+1}{2}$, when all $Y's$ are small than all $X's$. The exact null distribution of W_N can be found using enumeration and recursive formulas. Tables of Null distribution of W_N are widely available for different values of n_1, n_2. However, asymptotic probability distribution of W_N can be used for large values of n_1, n_2. We will discuss about it. But before that, we will establish the equivalence between Mann Whitney U statistic and Wilcoxon Rank-Sum Statistic. Mann Whitney U statistic is given by –

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_{ij}$$

Where,

$$D_{ij} = \begin{cases} 1 & if\ Y_j < X_i \\ 0 & if\ Y_j > X_i \end{cases} \quad for\ i = 1,2,\ldots,n_1\ and\ j = 1,2,\ldots,n_2$$

Now, U can be written as –

$$U = \sum_{i=1}^{n_1} (D_{i1} + D_{i2} + \ldots D_{in_2})$$

Now, $(D_{i1} + D_{i2} + \ldots D_{in_2})$ gives the number of values for j for which $Y_j < X_i$ or the rank of X_i reduced by m_i, the number of $X's$ which are less than or equal to X_i. Then, we can write –

$$U = \sum_{i=1}^{n_1} [rank(X_i) - m_i] = \sum_{i=1}^{n_1} rank(X_i) - (m_1 + m_2 + \ldots + m_{n_1})$$

$$= \sum_{i=1}^{N} iZ_i - \frac{n_1(1+n_1)}{2}$$

$$= W_N - \frac{n_1(1+n_1)}{2}$$

Thus, Mann Whitney U statistic and Wilcoxon Rank Sum Statistic differs only by a constant. Hence, we can use the properties derived of Mann Whitney U statistic for Wilcoxon Rank-sum statistic. Using the notations described earlier in Mann Whitney U statistic, we can write –

$$E(W_N) = n_1 n_2 p + \frac{n_1(1+n_1)}{2}$$

$$Var(W_N) = n_1 n_2 [p - p^2(n_1 + n_2 - 1) + (n_2 - 1)p_1 + (n_1 - 1)p_2]$$

These properties will be helpful in further discussion.

It is clear that the tables of null distribution Mann Whitney U statistic are equivalent to the tables of null distribution of Wilcoxon Rank-sum statistic. However, to get critical points for Wilcoxon Rank-sum Test, we need to add some constant ($\frac{n_1(1+n_1)}{2}$) to the critical point of Mann Whitney U Test.

That is, $P_r(U \geq c_\alpha | H_0) \leq \alpha$. ($\alpha$ is our chosen level of significance).

⇨ $P_r(U + \frac{n_1(1+n_1)}{2} \geq c_\alpha + \frac{n_1(1+n_1)}{2} | H_0) \leq \alpha$

⇨ $P_r(W_N \geq c_\alpha + \frac{n_1(1+n_1)}{2} | H_0) \leq \alpha$

This result is very much helpful. Later we will see that Mann Whitney U Test and Wilcoxon Rank-sum Test are equivalent in terms of every respect (e.g., Level, Power, Consistency).

For large sample sizes we can use asymptotic probability distribution of W_N. But for that we need to obtain mean and variance of W_N under H_0. We have already obtained a general expression for mean and variance. Under H_0, $p = \frac{1}{2}, p_1 = \frac{1}{3}, p_2 = \frac{1}{3}$. so, we get-

$$E(W_N) = \frac{n_1 n_2}{2} \text{ and } Var(W_N) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Using Central Limit Theorem, it can be shown that under H_0 -

$$T = \frac{W_N - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}} \xrightarrow{L} Z \sim Normal(0,1)$$

This approximation has been found good for small sample sizes also. The reason for such good approximation is that the distribution of W_N is symmetric under H_0. Since, W_N can take only integer values, a continuity correction of 0.5 is required.

For the alternative hypothesis $H_1: F_Y(x) \geq F_X(x)$ for all x and $F_Y(x) > F_X(x)$ for some x, using asymptotic distribution the testing rule will be 'Reject H_0 iff T > τ_α'. Where, τ_α is the upper 100(1- α)% point of $Normal(0,1)$ distribution.

6.4.5.3. Problem of Ties:

When a large number of ties are present in the sample, we need to reformulate the Wilcoxon Rank-sum statistic. Since, in our discussion that part is not needed, we are omitting that part. However, one should understand that the Wilcoxon Rank-sum statistic we have defined does not takes into account ties. Hence, if the assumption of continuous population distribution is not valid, the use of the statistic in that case may reduce the performance of the test.

6.4.5.4. Properties:

Wilcoxon Rank-sum test is widely used non-parametric test in two sample problem. Its properties are similar to Mann Whitney U test. This test is mainly used for the null hypothesis of identical distributions. The test is simple to use for any sample size and tables of exact null distributions are widely available. Even the asymptotic distribution can be used for small sample sizes also. The large sample approximation is quite useful and in case of presence of ties, the corrections may be incorporated to find the test statistic. The test is good for the particular tests like – equality of means, equality of medians etc. Like Mann Whitney U test, it is also sensitive to location alternatives.

Thus, we can compare the performance of this test with corresponding parametric tests of location also.

6.4.5.5. Example:

Consider the example, that we have given in case of LRT of exponential distribution. we will use Wilcoxon Rank sum test for testing the null hypothesis of equality of means. We will use the same notations explained in the example of Exponential LRT. Here, $n_1 = 5, n_2 = 7$. So, $N = n_1 + n_2 = 12.$ Notice, the assumption of no ties is valid here.

First, we will consider the combined ordered arrangement of the two samples. We will keep in mind which are X sample values and which are Y sample values. Then, we will assign rank to them. (The convention is to assign rank 1 to smallest observation in the combined sample). We prepare the following table –

Table 3: Calculation Related to Wilcoxon Rank Sum Test

X Sample Value	Number of Y Observations less than X	Number of X observations less than equal to X	Rank of X observation in the combined sample
8.8	1	1	2
13.6	3	2	5
15.1	4	3	7
25.7	6	4	10
30.2	6	5	11

From the above table we can see that, the value of Wilcoxon Rank Sum statistic is simply the sum of the last column values. Thus, $W_N = 35$.

From the table of critical values of Wilcoxon Rank Sum test, we find that the critical value for the above test with $n_1 = 5, n_2 = 7$ for level of significance 0.05 is given by $28 + \frac{n_1(1+n_1)}{2} = 43$. (We have already discussed how we can use the equivalence between Mann Whitney U test and Wilcoxon Rank sum test to find critical value for Wilcoxon Rank sum test using critical value of Mann Whitney U Test for same level of significance). Hence, $W_N = 35 < 43$. Thus, we accept H_0 at level of significance 0.05.

Now, we will apply large sample approximation to conduct the test. We will calculate the statistic -

$$T = \frac{W_N - \frac{n_1(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}$$

For the given data, we get $T_{obs} = 0.405999$. (Same as in case of Mann Whitney U Test, which is quite natural). Now, the critical point for level 0.05 test is $\tau_\alpha = \tau_{0.05} = 1.644854$. Which is quite large than T_{obs}. Hence, here also we reject H_0, as $T_{obs} < \tau_{0.05}$.

So, in both cases we get the same decision. Hence, in the light of the given data it seems that the lifetime of the bulbs produced by company A and company B are same. Hence, the claim of the company A is not justified.

7. Difficulty in Comparison of Tests in terms of Power:

We have completed our general discussion of different parametric tests and non-parametric for our concerned hypothesis problem. Now, we will discuss how to obtain power of these tests. It is clear that for finding power, we have to find the distribution of the test statistic under the alternative hypothesis. In other words, we wish to obtain power function. But the question is - power function will be function of which quantity? In case of One sample test related mean of distribution, power curve is drawn against different values mean (μ). But in case of two sample problem of equality of mean, we have two population means, which are unknown under null hypothesis and under alternative hypothesis. So, it is not possible to make power function as function of the two population means. Here, one may consider the difference between the population means. As our null hypothesis is equality of population means, a good test must detect any difference between the population means. So, here we will consider power functions of the test as a function of difference between the two population means. But how to obtain power functions. In case of discussion of non-parametric tests, we have seen that even the null distributions of the test statistics are difficult to find, so it is not possible to obtain a closed mathematical expression of power functions in case of non-parametric tests. Even in case parametric tests it is not very easy to find. For example – In case of testing equality of mean of two independent normal distributions, the distribution of the test statistic –

$$T = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Under H_1 is not very simple. The power function is a function of CDF of non-centralized t distribution, which is very complicated. Thus, if one wishes to compare the different tests in terms of power, it is difficult. Statistician uses an alternative approach, which is simulation.

In the next section we will give a brief introduction to what is simulation and how we can use this to find empirical power of different tests.

8. What is Simulation and how to use it:

8.1. A general discussion on simulation and its use in Statistics:

In simulation one uses models to create specific situations, in order to study the response of the model to them. Thus, one gets an idea about what would happen to the system (a portion of our interest) in real world. So, simulations can be used to carryout experiments which are dangerous, difficult, expensive to do in the real world. For example - What would happen if two engines of a plane stop in the mid-air?

However, in statistics we will consider simulation where things are subject to randomness or chance factor. Models for such systems involve random variables and modelling such systems is called Probabilistic Modelling or Stochastic Modelling. Here comes the role of Stochastic Simulation.

In probability theory we have studied different types of models, in statistics we have studied different modelling techniques. But sometimes probabilistic and statistics models are so complex that mathematical treatment is not possible. Thus, tools of mathematical analysis are not sufficient to answer all the relevant questions about them. Stochastic Simulation is an alternative approach, which can be used to answer such questions. In stochastic simulation we generate random variables and then insert them into a model of our interest. Thus, mimicking outcome for the whole system. In our discussion we will only consider stochastic simulations.

In statistics when we use certain models, we have some assumptions regarding the applicability of those models. For example – One of the important assumptions of classical linear model is that the errors are independently and identically normally distributed with mean zero and constant variance. But what happens when these assumptions are violated? Exact analytical evaluation may not be possible always. Although large sample approximations to properties may be possible, but what would happen if sample size is not large is still a question. In these situations, simulation may be a way out. Simulation is a numerical technique for conducting experiments on computer.

By Simulation statisticians mean *Monte Carlo Simulation*. It is a computer experiment involving random sampling from probability distributions. There are various uses of Monte Carlo Simulation in statistics. Such as –

1. One may wish to know different properties of an estimator. Is the estimator biased in finite samples? Is it consistent? How does the sampling variance change as parameter value changes?
2. How does a given test (testing rule) perform if the assumptions are violated?
3. Does a procedure for constructing a confidence interval for a parameter achieve the advertised confidence interval?
4. What is penalty of using non-parametric tests instead of using parametric tests, if one knows the probability law (e.g., whether it is normal or double exponential) of the study variable?

To answer such questions, one can use Monte Carlo simulation.

We will use R software to carry out our simulations. It is very easy to use and user friendly also.

8.2. Simulation to get idea about Sampling Distribution of Estimator and Test-Statistic:

An estimator or test-statistic has a true sampling distribution under a set of conditions (e.g., Parameter value, sample size, true distribution of population etc). Sometimes the sampling distribution cannot be obtained mathematically. But using simulation we can get an idea about it. For example – In case of testing equality of means of two independent normal distributions, under H_0 the sampling distribution of the test statistic is t distribution. However, under H_1 the sampling distribution of the test statistic is not known. We can use simulation to get idea about it. We will write down the general steps of such problems.

Step-1: Generate R independent datasets under the conditions of interest. The conditions may be sample size, value of true parameters, true probability distribution of the population.

Step-2: Based on the data generated in step-1, we will calculate the numerical value of the estimator/test-statistic(T) for each data set. We denote those values by - $T_1, T_2, \ldots\ldots, T_R$.

Step-3: If R is large enough, then the values of the estimator/test-statistic generated can be considered as a good representative of the true sampling distribution of the estimator/test-statistic under the given set of conditions. Thus, from that one can get idea about the mean, variability and other related properties of the estimator/test-statistic.

Now, we will use the above steps to generate R = 1000 observations of the test statistic T, that we have derived in case of LRT of equality of two independent normal populations. We choose, $n_1 = 7, n_2 = 8, \mu_1 = 20, \mu_2 = 17$ and $\sigma^2 = 16$. We present the histogram of the test-statistic T under these conditions.

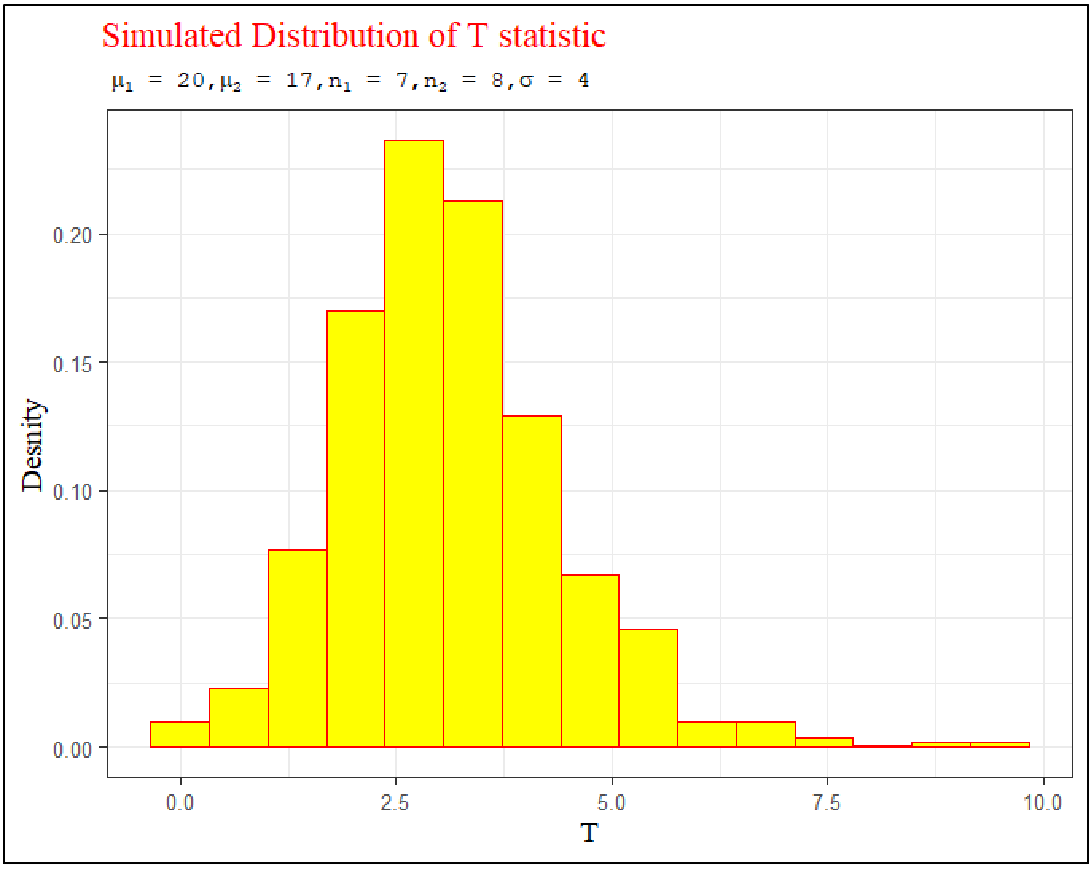


Diagram 11: Simulated Distrbution of T statistic for $\mu_1 = 20, \mu_2 = 17, \sigma = 4$.

From the above histogram we can clearly see that it is very much deviated from the null distribution, which is symmetric about zero.

We can get an idea about the mean and variability of the distribution of the estimator/test-statistic from the values of statistic T generated through simulation. In step-2 we get R values of estimator/test-statistic $T_1, T_2, \ldots\ldots, T_R$. Then, simulated mean and variance of the estimator/test-statistic are given by –

$$\widehat{E(T)} = \frac{1}{R}\sum_{i=1}^{R} T_i$$

$$\widehat{Var(T)} = \frac{1}{R}\sum_{i=1}^{R}(T_i - \widehat{E(T)})^2$$

These simulated values give a good idea about the mean and variance of the true sampling distribution of estimator/test-statistic under the set of conditions.

In our example, the simulated mean and variance is given by –

$$\widehat{E(T)} = 6.230714 \; and \; \widehat{Var(T)} = 3.153801$$

So, the actual mean and variance of the sampling distribution would be near to the simulated mean and variance.

Thus, one can find mean of a probability distribution or a function of random variable using simulation. We will use this concept to find the value of an integral (over finite range).

8.3. A Beautiful Example:

Suppose, we want to evaluate the integral $\int_{-5}^{5} xe^x \, dx$. This is very simple integral and we can calculate this by hand also. The value of the integral is 593.6931. Now, we will use simulation to calculate this integral.

Consider, $X \sim Uniform(-5,5)$. Consider the function $g(x) = xe^x$. Then, Expectation of $g(X)$ is given by –

$$E\big(g(X)\big) = \int_{-5}^{5} g(x) \, \frac{1}{10} dx = \frac{1}{10}\int_{-5}^{5} xe^x \, dx$$

Which implies, $\int_{-5}^{5} xe^x \, dx = 10 * E\big(g(X)\big)$.

Thus, to calculate the integral by simulation, we have to calculate an approximate value $E\big(g(X)\big)$ by simulation. Which can be calculated by drawing a random sample of size R = 1000 from $Uniform(-5,5)$ and then calculating the value of the function for each value of the sample and ultimately, we take mean of those values. Setting set. Seed = 987654321 in R we get the simulated mean 58.96545. we then multiply it by 10 to get simulated value of the integral. The value is 589.6545, which is near to 593.6931.

8.4. Simulation to Calculate Empirical Size and Power of given Test:

Sometimes it is not possible to calculate size and power of a test (testing rule) explicitly. The main difficulty is that the true sampling distribution of the test-statistic is difficult to obtain mathematically. We have discussed that how simulation can be used to get an idea of the sampling distribution of the test-statistic under given set of conditions. We can add some additional steps to get the size and power of the test.

Suppose, $X \sim F(x;\theta)$, θ is the unknown parameter ($\theta \in \Omega$) and $F(x,\theta)$ is probability law of X. We consider the following testing problem $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ a testing rule rejects H_0 iff the test statistic $T > c_\alpha$. We wish to calculate size and power of the test through simulation.

Step-1: Generate R independent datasets under the conditions of interest. The conditions may be sample size, value of true parameters, true probability distribution of the population. If we want to calculate level, then true parameter value is $\theta = \theta_0$ (If we want to calculate power, the true parameter value $\theta \in \Omega - \{\theta_0\}$).

Step-2: Based on the data generated in step-1, we will calculate the numerical value of the test-statistic(T) for each data set. We denote those values by - $T_1, T_2, \ldots\ldots, T_R$.

Step-3: If R is large enough, then the values of the test-statistic generated can be considered as a good representative of the true sampling distribution of the test-statistic under the given set of conditions. Now, we calculate the proportion of cases, where the value of the test statistic exceeds c_α. Which is empirical size (empirical power for given parameter θ).

If the critical point c_α is chosen so that level of the test is α. Then, it is expected that the empirical size will be near α. If theoretical results are known, we can verify whether simulated results are near to theoretical results.

We will not give examples here. Because, we will consider a large number of simulations related to calculation of empirical size and power of the parametric and non-parametric tests related to two sample problem which we have discussed in the previous section.

9. Simulations, Results and Observations:

9.1. Introduction:

In this section we will discuss different simulation studies related to size and power of different parametric and non-parametric tests in two sample problem. Here, our null hypothesis is equality of means and alternative is mean of X is greater than mean of Y. We wish to study the penalty (if any) of using non-parametric test instead of using corresponding parametric test (LRT), when we know the type of sampled population (i.e., Normal Population or Exponential Population).

Here we will consider different situations. For example- we will take different pairs of sample sizes, different level of significance, different value of R (discussed in previous section). For small sample sizes we take R = 3000 and for large sample size we take R = 2000. For small sample sizes empirical size and power are not very stable, it changes very much from one simulation to other simulation. Hence, we will take large value of R here. While in large sample sizes the empirical size and power are comparatively stable. Hence, R = 2000 will work. We will consider level of significance $\alpha = 0.10, 0.05$. Sample size is an important factor. For small samples sizes some tests perform very bad, while some perform well. Thus, we will consider different pair of combination of sample sizes. We will take large sample sizes to examine whether the test is consistent or not.

For simulation study we will use R software. For each simulation discussed here, we have taken set. Seed = 987654321 for uniformity of result. To calculate power, we have to consider the parameter values (In our case Mean) under $H_1: \mu_1 > \mu_2$. To draw the empirical power curve, we will plot the simulated power values against difference between the population means ($\mu_1 - \mu_2$). But the problem is that the power of all tests will not just depend on difference between population means, it depends on magnitude of $\mu_1 \; and \; \mu_2$ also. Hence, it is important to see the study power for different values of μ_1 and μ_2. For that we will use an alternative approach. To calculate power, we will first fix $\mu_2 = \mu$ and then take $\mu_1 = \mu + d$. In our case we will consider different values of $d > 0$ (Note that, $d = 0$ for level).

To calculate the simulated size and power in different situations, we will use some user defined functions in R. Using those user-defined functions we can calculate empirical level and power under different situations.

9.2. SIMULATION 01:

Comparison of Empirical Size and Power of Two Sample T test and Kolmogorov-Smirnov Test When Sampled Population is Normal

9.2.1. Motivation:

Kolmogorov-Smirnov is one of the popular tests for two sample problems. The applicability of the test is wide. It can detect any type of difference between the CDFs of two populations. On the other hand, Two Sample T test is another well-known test for testing equality of means of two independent normal populations. The main assumption of this test is that the two normal population differs with respect to mean only, that is variance is same. It is very interesting to study how the Kolmogorov-Smirnov test will perform for testing the equality of two means of two independent normal population.

9.2.2. Objective:

We wish to compare empirical size and power of T test and Kolmogorov-Smirnov Test when both the sampled population is Normal. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.2.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.T1$ The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

mu: Mean of Y population.

d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $mu + d$. Where, mu is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.

$alpha$: It is level of significance of the test.

$exact.crit$: If it is TRUE then, the function calculates empirical size and power using exact critical value. We need to give exact critical point of the test (For the given sample size and level of significance) as input. If it is FALSE, then it will calculate empirical size and power using large sample approximation of the test-statistic. By default, it is FALSE.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.T1$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.2.4. Discussion:

For the discussion we will take variance of the two population as 1. We will consider $d = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using table and graphs. We have used exact critical values from the table of Kolmogorov-Smirnov Statistic.

For $n_1 = 4, n_2 = 5$, we will construct the table of empirical size and power of Kolmogorov-Smirnov test and T test side by side for $mu = 5,10$ and $alpha = 0.10,0.05$.

Table 4: Empirical size and Power of Kolmogorov-Smirnov Test and T test for $mu = 5, n_1 = 4, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	T Test	Kolmogorov – Smirnov Test	T Test
0.0	0.08033	0.10767	0.04733	0.05767
0.5	0.19533	0.27467	0.12500	0.15500
1.0	0.41100	0.54533	0.28167	0.37600
1.5	0.65300	0.79467	0.49400	0.62833
2.0	0.85233	0.93333	0.70833	0.84967
2.5	0.94700	0.98833	0.85833	0.95400
3.0	0.98667	0.99833	0.94967	0.98933
3.5	0.99767	1.00000	0.98400	0.99933
4.0	1.00000	1.00000	0.99433	0.99967

Table 5: Empirical size and Power of Kolmogorov-Smirnov Test and T test for $mu = 10, n_1 = 4, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	T Test	Kolmogorov – Smirnov Test	T Test
0.0	0.08033	0.10767	0.04733	0.05767
0.5	0.19533	0.27467	0.12500	0.15500
1.0	0.41100	0.54533	0.28167	0.37600
1.5	0.65300	0.79467	0.49400	0.62833

2.0	0.85233	0.93333	0.70833	0.84967
2.5	0.94700	0.98833	0.85833	0.95400
3.0	0.98667	0.99833	0.94967	0.98933
3.5	0.99767	1.00000	0.98400	0.99933
4.0	1.00000	1.00000	0.99433	0.99967

Notice, for both values of mu we get same table of empirical size and power. Which is very much intuitive. Even we will get same table for all possible values of mu. Instead of mu, if we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)

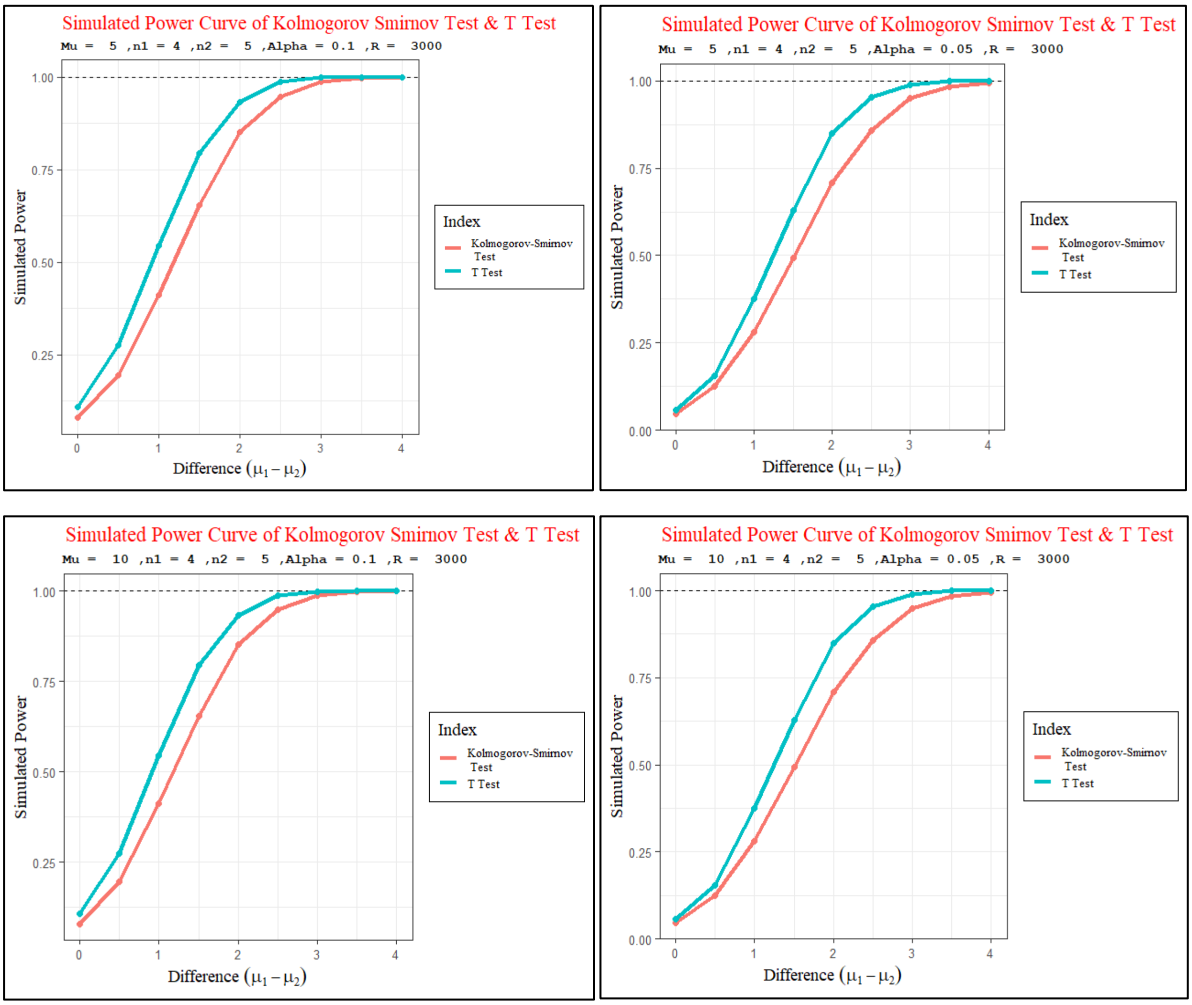


Diagram 12: Simulated Power Curve of Kolmogorov Smirnov test & T test for $mu = 5{,}10, n_1 = 4, n_2 = 5, R = 3000$ and $\alpha = 0.10{,}0.05$

One thing to notice is that, for level of significance 0.10, the power of both tests reaches 1. While for level of significance 0.05, none of the tests reaches power 1. Also, the power of T-test more quickly reaches 1 than Kolmogorov-Smirnov Test and over all values of difference (d) T-test is uniformly powerful than Kolmogorov-Smirnov Test.

Now, if we take $n_1 = 5, n_2 = 4$ (that is, we interchange value of n_1 and n_2), then we get a different table. We have only presented the table for $mu = 10$.

Table 6: Empirical size and Power of Kolmogorov-Smirnov Test and T test for $mu = 10$, $n_1 = 5, n_2 = 4$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	T Test	Kolmogorov – Smirnov Test	T Test
0.0	0.08100	0.12033	0.04667	0.06500
0.5	0.19667	0.26667	0.11533	0.15700
1.0	0.42833	0.56633	0.27767	0.38933
1.5	0.65200	0.79367	0.49933	0.64133
2.0	0.84367	0.92600	0.71467	0.84067
2.5	0.95067	0.98733	0.86900	0.95767
3.0	0.98667	0.99867	0.94567	0.99067
3.5	0.99767	1.00000	0.98133	0.99933
4.0	0.99933	1.00000	0.99467	1.00000

We see that, if we interchange the value of n_1 and n_2, then empirical power of both the tests changes. But it doesn't mean that power of two tests is not symmetric in sample sizes (n_1,n_2). It may happen due to sample fluctuations. Until we know the exact analytical form of power functions, we cannot say this explicitly.

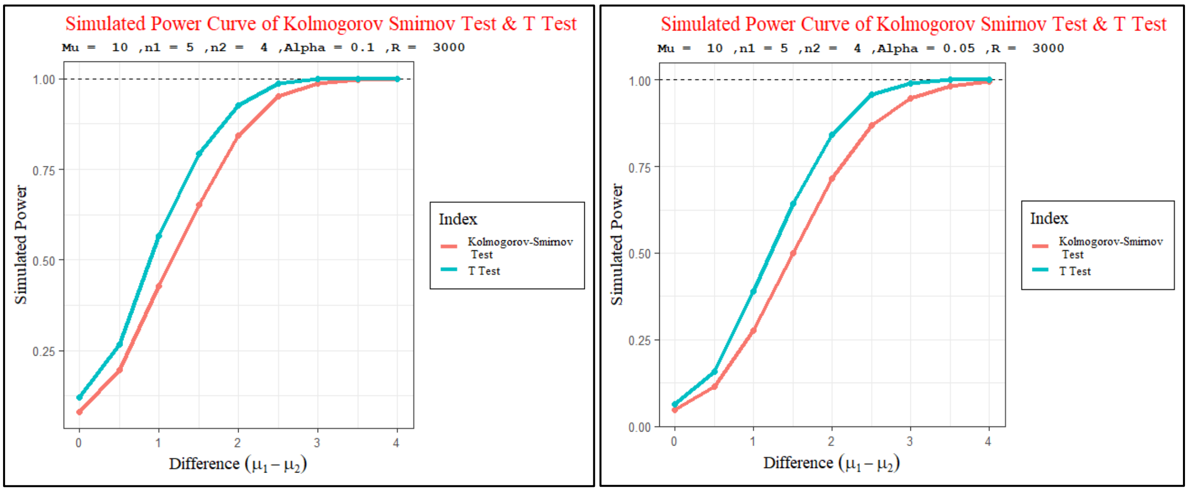


Diagram 13: Simulated Power Curve of Kolmogorov Smirnov test & T test for $mu = 10, n_1 = 5, n_2 = 4, R = 3000$ and $\alpha = 0.10,0.05$

Now, we will take $n_1 = 10, n_2 = 10$, we will construct the table of empirical size and power of Kolmogorov-Smirnov test and T test side by side for $mu = 10$ and $alpha = 0.10,0.05$.

Table 7: Empirical size and Power of Kolmogorov-Smirnov Test and T test for $mu = 10$, $n_1 = 10, n_2 = 10$ and $alpha = 0.10,0.05$, $R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	T Test	Kolmogorov – Smirnov Test	T Test
0.0	0.08750	0.10600	0.02700	0.05400
0.5	0.31250	0.39900	0.14250	0.25750
1.0	0.70750	0.83050	0.47600	0.71050
1.5	0.93350	0.97850	0.81100	0.94300
2.0	0.99650	1.00000	0.95950	0.99900
2.5	0.99950	1.00000	0.99650	1.00000
3.0	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000
4.0	1.00000	1.00000	1.00000	1.00000

Here, the power reaches 1 more quickly than previous cases. Thus, as we increase the sample size power of both the test reaches 1 more quickly.

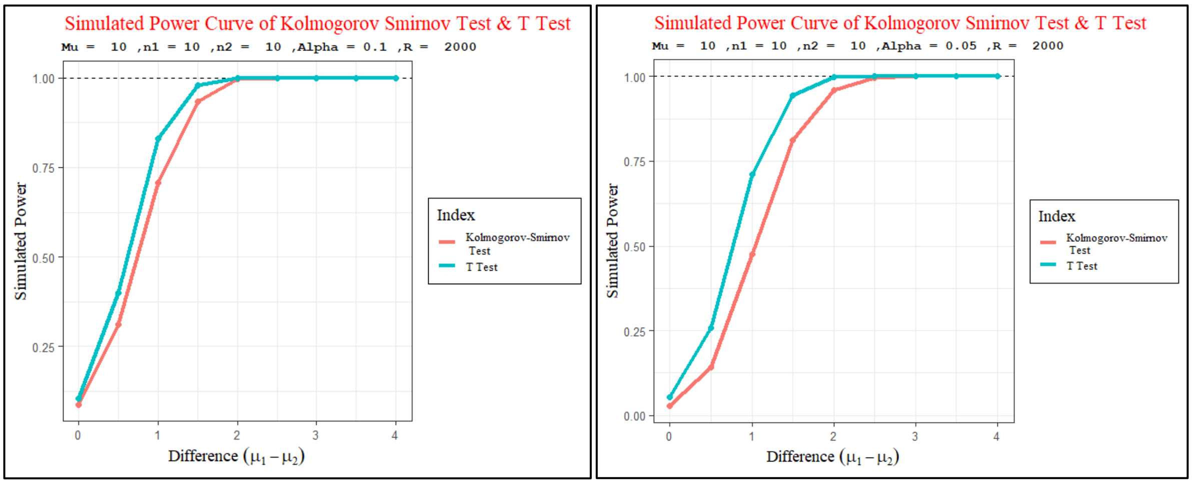


Diagram 14: Simulated Power Curve of Kolmogorov Smirnov test & T test for $mu = 10, n_1 = 10, n_2 = 10, R = 2000$ and $\alpha = 0.10, 0.05$

Finally, we consider another case, large value of n_1 and n_2. We take, $n_1 = 16$ and $n_2 = 20$.

Table 8: Empirical size and Power of Kolmogorov-Smirnov Test and T test for $mu = 10$, $n_1 = 16, n_2 = 20$ and $alpha = 0.10, 0.05$, $R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	T Test	Kolmogorov – Smirnov Test	T Test
0.0	0.08950	0.10400	0.03700	0.05000
0.5	0.48150	0.58850	0.33900	0.43300
1.0	0.89300	0.95100	0.79900	0.89600
1.5	0.99450	0.99850	0.98100	0.99550
2.0	1.00000	1.00000	0.99950	1.00000
2.5	1.00000	1.00000	1.00000	1.00000
3.0	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000
4.0	1.00000	1.00000	1.00000	1.00000

Here, the power reaches 1 more quickly than all the previous cases.

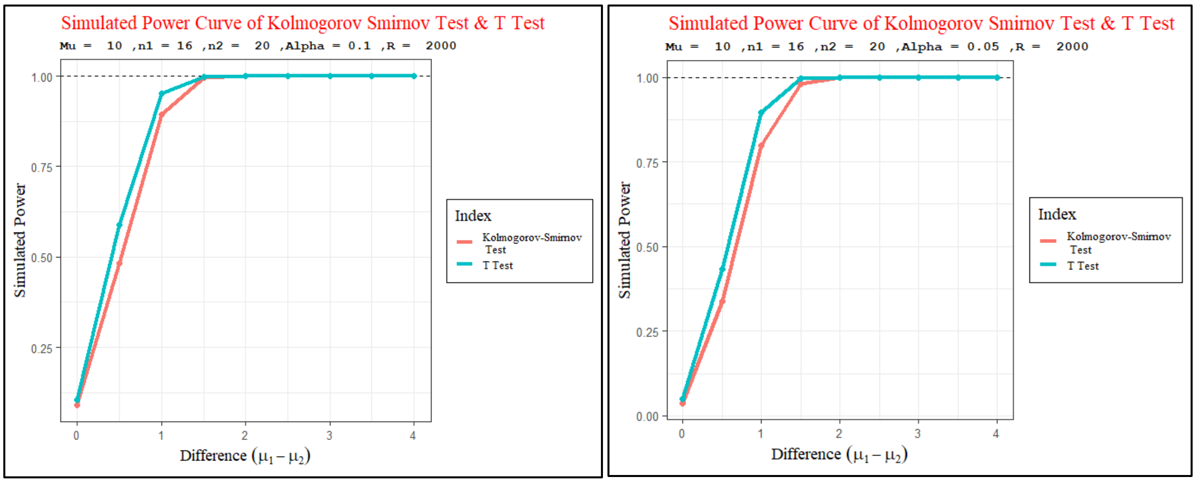


Diagram 15: Simulated Power Curve of Kolmogorov Smirnov test & T test for $mu = 10, n_1 = 16, n_2 = 20, R = 2000\ and\ \alpha = 0.10, 0.05$

9.2.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is normal, T-test performs better than Kolmogorov-Smirnov test in terms of power. i.e., the T-test is able to detect the same difference between the means of two independent normal population with same variance more frequently than Kolmogorov-Smirnov test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are not very high. As sample sizes increase the difference between the power of two test decreases. It is clear from the graph also.
2. For small sample sizes (here, $n_1 = 4\ and\ n_2 = 5$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance(alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.
3. The power of both the tests do not depend upon the value of mu (Y population mean). Other parameters remain fixed, if mu changes, it doesn't change the value of power. Instead of mu, if we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)

4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 50, n_2 = 50$ and $mu = 10$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

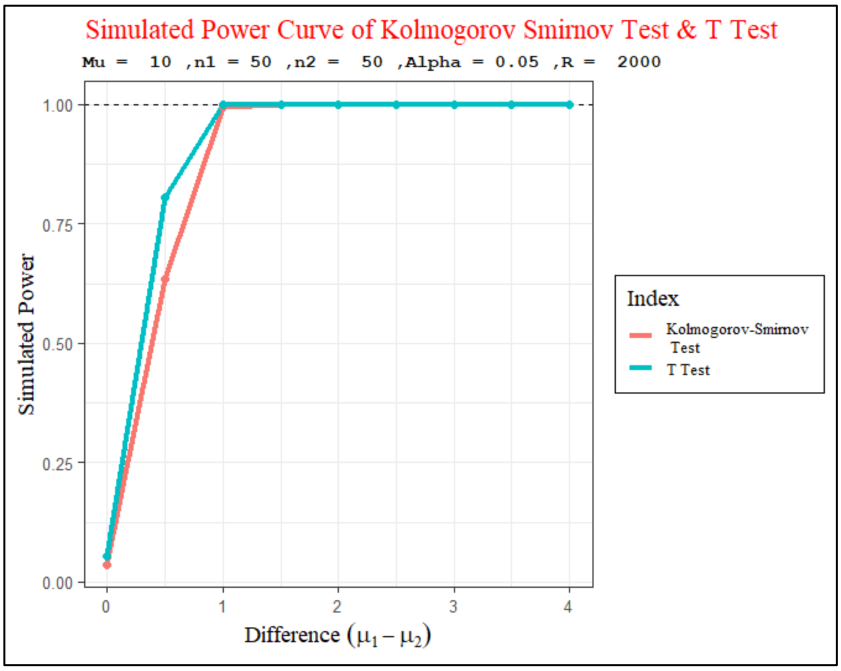


Diagram 16: Simulated Power Curve of Kolmogorov Smirnov test & T test for $mu = 10, n_1 = 50, n_2 = 50, R = 2000$ and $\alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except for $d = 0.5$. (Here, we have used large sample approximation for Kolmogorov-Smirnov test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.3. SIMULATION 02:

Comparison of Empirical Size and Power of Two sample Exponential LRT and Kolmogorov-Smirnov Test When Sampled Population is Exponential

9.3.1. Motivation:

Kolmogorov-Smirnov is one of the popular tests for two sample problems. The applicability of the test is wide. It can detect any type of difference between the CDFs of two populations. On the other hand, two sample Exponential LRT is another well-known test for testing equality of means of independent exponential population. It is very interesting to study how the Kolmogorov-Smirnov test will perform for testing the equality of two means of two independent exponential population.

9.3.2. Objective:

We wish to compare empirical size and power of two sample Exponential LRT and Kolmogorov-Smirnov Test when both the sampled population is Exponential. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.3.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.LRT1$. The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

$lamda$: Mean of Y population.

d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $lamda + d$. Where, $lamda$ is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.

$alpha$: It is level of significance of the test.

$exact.crit$: If it is TRUE then, the function calculates empirical size and power using exact critical value. We need to give exact critical point of the test (For the given sample size and level of significance) as input. If it is FALSE, then it will calculate empirical size and power using large sample approximation of the test-statistic. By default, it is FALSE.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.LRT1$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.3.4. Discussion:

For the discussion we will consider $d = 0, 1, 2, 3, \ldots\ldots, 30$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using tables and graphs. We have used exact critical values from the table of Kolmogorov-Smirnov Statistic.

For $n_1 = 5, n_2 = 7$, we will construct the table of empirical size and power of Kolmogorov-Smirnov test and two sample Exponential LRT side by side for $lamda = 1,3$ and $alpha = 0.10, 0.05$.

Table 9: Empirical size and Power of Kolmogorov-Smirnov Test and two sample Exponential LRT for $lamda = 1$, $n_1 = 5, n_2 = 7$ and $alpha = 0.10,0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	Two sample Exponential LRT	Kolmogorov – Smirnov Test	Two sample Exponential LRT
0	0.09067	0.10667	0.03367	0.05233
1	0.32733	0.44567	0.14667	0.31300
2	0.54500	0.71233	0.30400	0.58567
3	0.69433	0.84500	0.41333	0.75600
4	0.77167	0.90500	0.51000	0.84133
5	0.84067	0.94933	0.60833	0.90033
6	0.87967	0.97100	0.66133	0.94367
7	0.91133	0.98267	0.72033	0.96033
8	0.93333	0.99067	0.75767	0.97733
9	0.94467	0.99067	0.78867	0.98133
10	0.95833	0.99667	0.81400	0.98633
11	0.96467	0.99700	0.83567	0.99267
…..	……	…..	……	…..
17	0.98833	0.99900	0.91267	0.99767
18	0.98900	0.99933	0.90633	0.99800
…..	……	…..	……	…..
27	0.99533	1.00000	0.95933	0.99900
28	0.99633	1.00000	0.96467	1.00000
29	0.99567	1.00000	0.96400	0.99967
30	0.99667	0.99967	0.95867	0.99967

Table 10: Empirical size and Power of Kolmogorov-Smirnov Test and two sample Exponential LRT for $lamda = 3$, $n_1 = 5, n_2 = 7$ and $alpha = 0.10,0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)	
	0.10	0.05

	Kolmogorov – Smirnov Test	Two sample Exponential LRT	Kolmogorov – Smirnov Test	Two sample Exponential LRT
0	0.09067	0.10667	0.03367	0.05233
1	0.15933	0.20867	0.06367	0.12167
2	0.24633	0.33567	0.10300	0.21200
3	0.33267	0.46300	0.16167	0.31800
4	0.40033	0.54767	0.19467	0.39800
5	0.49067	0.62733	0.26367	0.49900
6	0.55400	0.71200	0.30133	0.57800
7	0.62367	0.76367	0.37233	0.64733
8	0.65267	0.81367	0.40133	0.70333
9	0.68300	0.84500	0.42233	0.74367
10	0.72667	0.88900	0.46967	0.79667
…..	……	…..	……	…..
17	0.86867	0.96600	0.64000	0.93467
18	0.87100	0.96867	0.64000	0.93200
…..	……	…..	……	…..
27	0.94333	0.99267	0.78400	0.98100
28	0.94633	0.99400	0.78967	0.98367
29	0.95000	0.99367	0.80433	0.98233
30	0.95500	0.99367	0.81433	0.98467

Notice, for different values of $lamda$ we get different table of empirical size and power. Which is not in case, when sampled population is normal. Actually, exponential distribution is characterized by mean of the distribution. The variance of the distribution also depends upon the value of mean of the distribution (Which is not in case of Normal Distribution.). That is why, as value of $lamda$ changes, power of both the test changes. It is empirically observed that as value of $lamda$ increases or decreases, the power of both the test decreases or increases.

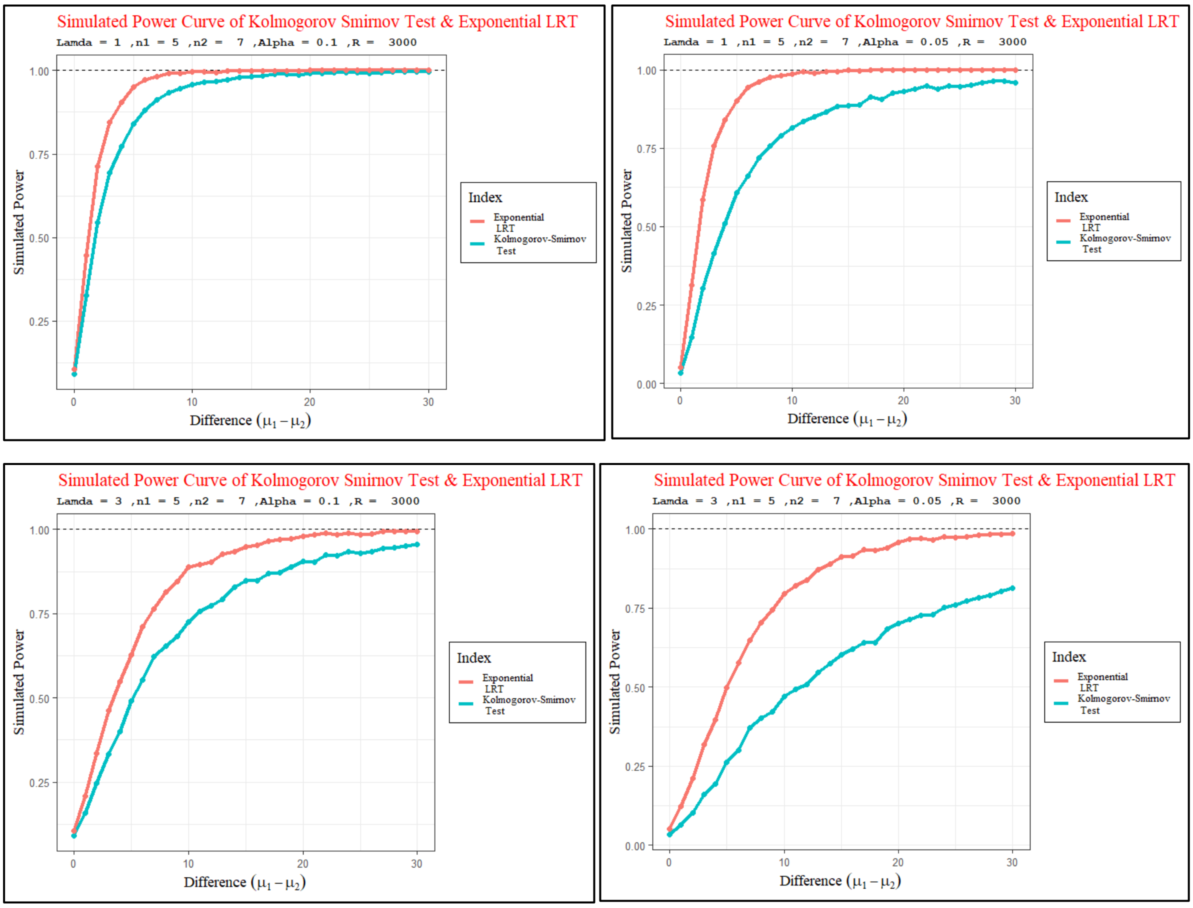


Diagram 17: Simulated Power Curve of Kolmogorov Smirnov test & Exponential LRT for $lamda = 1{,}3,\ n_1 = 5, n_2 = 7, R = 3000\ and\ \alpha = 0.10{,}0.05$

One thing to notice is that, for level of significance 0.10, the power of both tests reaches 1. While for level of significance 0.05, none of the tests reaches power 1. Also, the power of two sample Exponential LRT reaches more quickly 1 than Kolmogorov-Smirnov Test and over all values of difference (d) two sample Exponential LRT is uniformly powerful than Kolmogorov-Smirnov Test. If you closely observe the table of values of powers above, you will observe that the empirical power function is not monotone. It may happen due to sampling fluctuations.

Here also if we interchange sample sizes (n_1, n_2), we will see that, empirical power of both the tests changes. But it doesn't mean that power of two tests is not symmetric in sample sizes (n_1, n_2). It may happen due to sample fluctuations. Until we know the exact analytical form of power functions, we cannot say this explicitly.

For $lamda = 1$ and $n_1 = 10, n_2 = 10$, we get these set of values for power.

Table 11: Empirical size and Power of Kolmogorov-Smirnov Test and two sample Exponential LRT for $lamda = 1$, $n_1 = 10, n_2 = 10$ and $alpha = 0.10, 0.05, R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	Two sample Exponential LRT	Kolmogorov – Smirnov Test	Two sample Exponential LRT
0	0.08000	0.10400	0.03400	0.04900
1	0.44350	0.59250	0.22900	0.45300
2	0.70350	0.88500	0.47350	0.79300
3	0.85250	0.96650	0.64600	0.92250
4	0.90950	0.98600	0.76000	0.96650
5	0.95450	0.99500	0.83800	0.98650
6	0.97550	0.99850	0.91300	0.99450
7	0.97750	0.99900	0.92550	0.99700
8	0.99150	0.99850	0.95300	0.99750
9	0.99300	1.00000	0.96200	0.99900
10	0.99650	1.00000	0.97450	1.00000
……	…….	…….	……	…….
20	1.00000	1.00000	0.99650	1.00000
21	1.00000	1.00000	0.99800	1.00000
…..	…..	…..	…..	…..
29	1.00000	1.00000	1.00000	1.00000
30	1.00000	1.00000	0.99950	1.00000

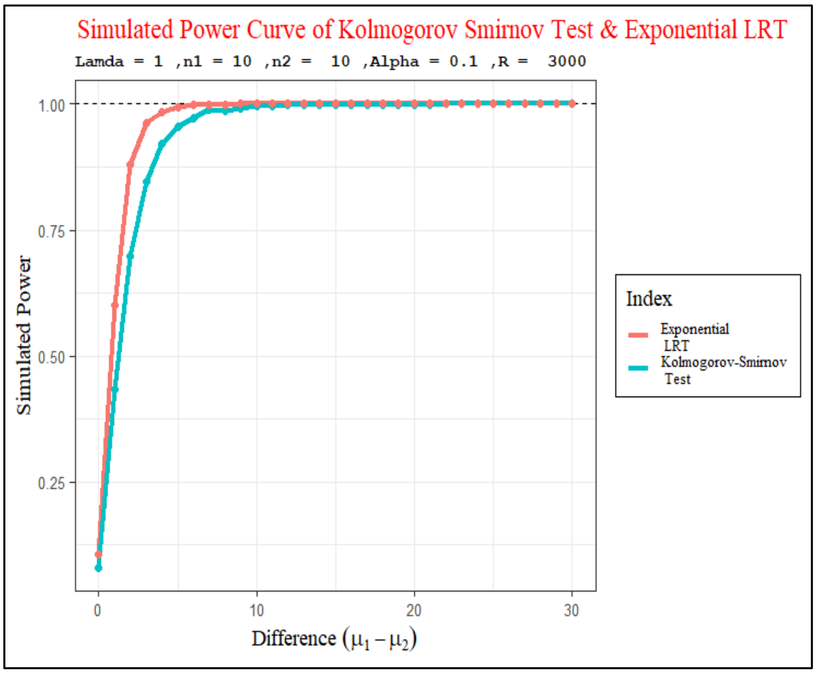


Diagram 18: Simulated Power Curve of Kolmogorov Smirnov test & Exponential LRT for $lamda = 1, n_1 = 10, n_2 = 10, R = 3000 \text{ } and \text{ } \alpha = 0.10$

For $lamda = 3$ and $n_1 = 18, n_2 = 12$, we get these set of values for power.

Table 12: Empirical size and Power of Kolmogorov-Smirnov Test and two sample Exponential LRT for $lamda = 3$, $n_1 = 18, n_2 = 12$ and $alpha = 0.10, 0.05$, $R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Kolmogorov – Smirnov Test	Two sample Exponential LRT	Kolmogorov – Smirnov Test	Two sample Exponential LRT
0	0.10200	0.09400	0.04750	0.04200
1	0.26300	0.30800	0.15200	0.18550
2	0.43700	0.53200	0.28400	0.36950
3	0.57450	0.69600	0.41550	0.55750
4	0.71100	0.83250	0.56200	0.70650
5	0.79750	0.90650	0.65300	0.82800
6	0.85350	0.94400	0.74200	0.89050
7	0.91750	0.97350	0.82000	0.94850
8	0.94750	0.98350	0.86950	0.96350
9	0.95350	0.99100	0.89150	0.97500

10	0.97550	0.99700	0.92800	0.99050
…..	…..	…..	…..	…..
20	0.99800	1.00000	0.99550	0.99950
21	0.99900	1.00000	0.99600	1.00000
…..	…..	…..	…..	…..
29	1.00000	1.00000	1.00000	1.00000
30	1.00000	1.00000	1.00000	1.00000

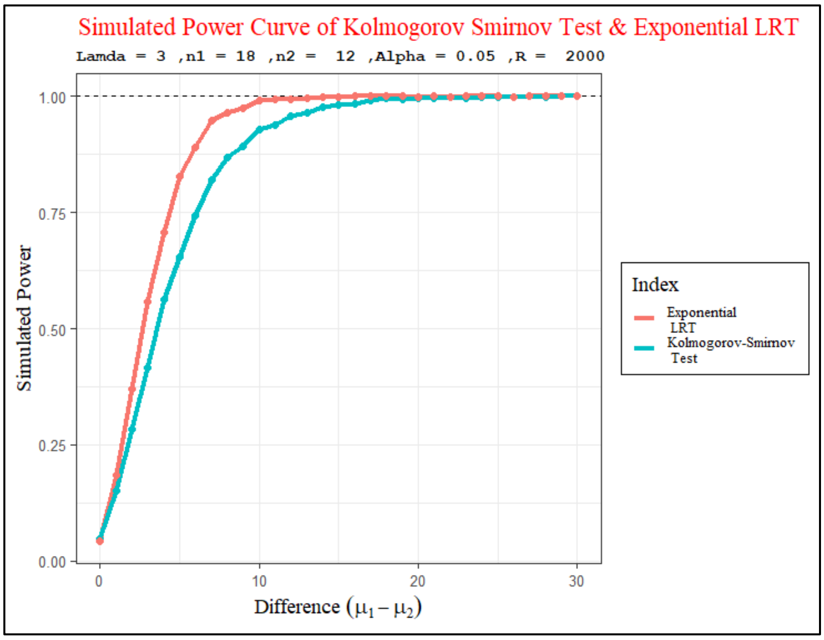


Diagram 19: Simulated Power Curve of Kolmogorov Smirnov test & Exponential LRT for $lamda = 3,\ n_1 = 18, n_2 = 12, R = 2000\ and\ \alpha = 0.05$

9.3.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is exponential, two sample Exponential LRT performs better than Kolmogorov-Smirnov test in terms of power. i.e., the two sample Exponential LRT is able to detect the same difference between the means of two independent exponential population more frequently than Kolmogorov-Smirnov test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are not very high. As sample sizes increase the difference between the power of two test decreases. It is clear from the graph also.

2. For small sample sizes (here, $n_1 = 5$ and $n_2 = 7$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance(alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.
3. The power of both the tests depends upon the value of $lamda$ (Y population mean). If $lamda$ increases or decreases, the value of power decreases or increases.
4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 50, n_2 = 50$ and $lamda = 3$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

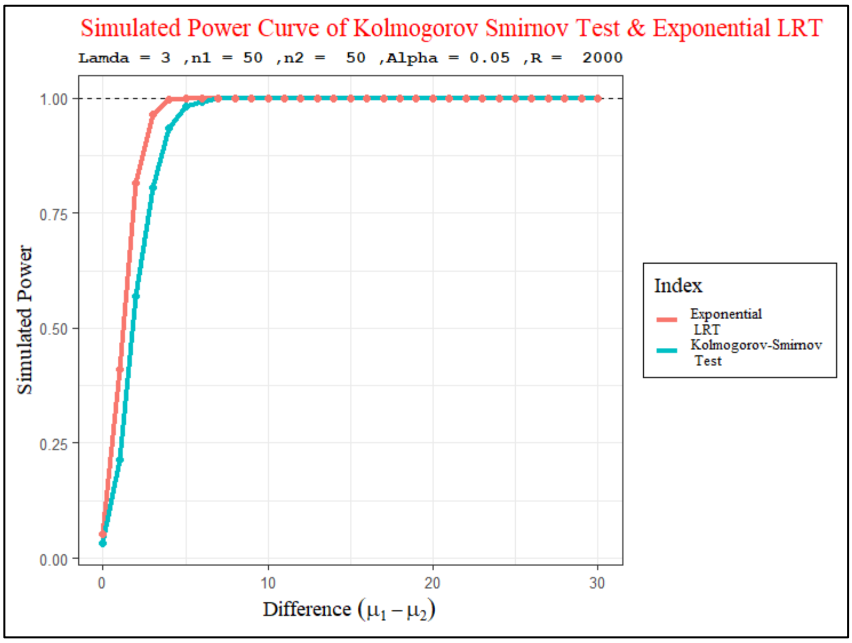


Diagram 20: Simulated Power Curve of Kolmogorov Smirnov test & Exponential LRT for $lamda = 3,\ n_1 = 50, n_2 = 50, R = 2000$ and $\alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except some differences. (Here, we have used large sample approximation for Kolmogorov-Smirnov test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.4. SIMULATION 03:

Comparison of Empirical Size and Power of Two sample T Test and Mann-Whitney U Test When Sampled Population is Normal

9.4.1. Motivation:

Mann Whitney U test is one of the popular tests for two sample problems. The applicability of the test is wide. It is mainly useful for testing location or shift alternatives. On the other hand, T test is another well-known test for testing equality of means of two independent normal populations. The main assumption of this test is that the two normal population differs with respect to mean only, that is variance is same. It is very interesting to study how the Mann-Whitney U test will perform for testing the equality of two means of two independent normal population.

9.4.2. Objective:

We wish to compare empirical size and power of T test and Mann-Whitney U Test when both the sampled population is Normal. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.4.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.T2$ The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are similar as $Power\_comparison.T1$.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.T2$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.4.4. Discussion:

For the discussion we will take variance of the two population as 1. We will consider $d = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$.

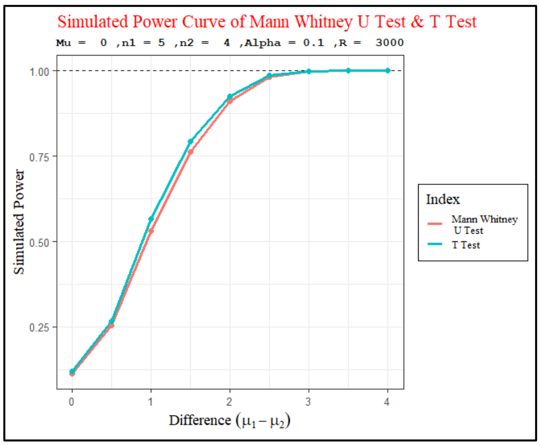
For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using table and graphs. We have used exact critical values from the table of Mann Whitney U Statistic. However, for large sample sizes we have used large sample approximation.

For $n_1 = 5, n_2 = 4$, we will construct the table of empirical size and power of Mann Whitney U test and T test side by side for $mu = 0$ and $alpha = 0.10, 0.05$.

Table 13: Empirical size and Power of Mann Whitney U Test and T test for $mu = 0$, $n_1 = 5, n_2 = 4$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	T Test	Mann Whitney U Test	T Test
0.0	0.11200	0.12033	0.04100	0.06500
0.5	0.25467	0.26667	0.10900	0.15700
1.0	0.53067	0.56633	0.28300	0.38933
1.5	0.76100	0.79367	0.52633	0.64133
2.0	0.91133	0.92600	0.74467	0.84067
2.5	0.98100	0.98733	0.90200	0.95767
3.0	0.99700	0.99867	0.96567	0.99067
3.5	0.99967	1.00000	0.98967	0.99933
4.0	1.00000	1.00000	0.99733	1.00000

As in case of Kolmogorov Smirnov Test (Sampled Population is Normal), here also for different values mu, we will get same table of empirical size and power. Which is very much intuitive. If we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)



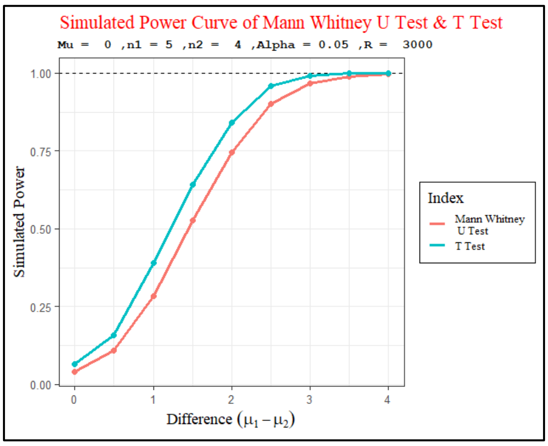


Diagram 21: Simulated Power Curve of Mann Whitney U test & T test for $mu = 0$, $n_1 = 5, n_2 = 4, R = 3000$ and $\alpha = 0.10, 0.05$

From the above table, we can say that the power of T-test more quickly reaches 1 than Mann-Whitney U Test and over all values of difference (d) T-test is uniformly powerful than Mann-Whitney U Test. It is evident from the graph also.

Now, if we take $n_1 = 4, n_2 = 5$ (that is, we interchange value of n_1 and n_2), then we get a different table. We have only presented the table for $mu = 2$.

Table 14: Empirical size and Power of Mann Whitney U Test and T test for $mu = 2$, $n_1 = 4, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	T Test	Mann Whitney U Test	T Test
0.0	0.10467	0.10767	0.03733	0.05767
0.5	0.25867	0.27467	0.10900	0.15500
1.0	0.51500	0.54533	0.28100	0.37600
1.5	0.75267	0.79467	0.51967	0.62833
2.0	0.91333	0.93333	0.74733	0.84967
2.5	0.97900	0.98833	0.90067	0.95400
3.0	0.99700	0.99833	0.96333	0.98933
3.5	0.99833	1.00000	0.99167	0.99933
4.0	1.00000	1.00000	0.99733	0.99967

Thus, the empirical power is not symmetric in sample sizes. (Note that, we are comparing the cases $mu = 0$, $mu = 2$, because we have seen they are comparable, as mu changes empirical power does not change). But it doesn't mean that power of two tests is not symmetric in sample sizes (n_1,n_2). It may happen due to sample fluctuations. Until we know the exact analytical form of power functions, we cannot say this explicitly.

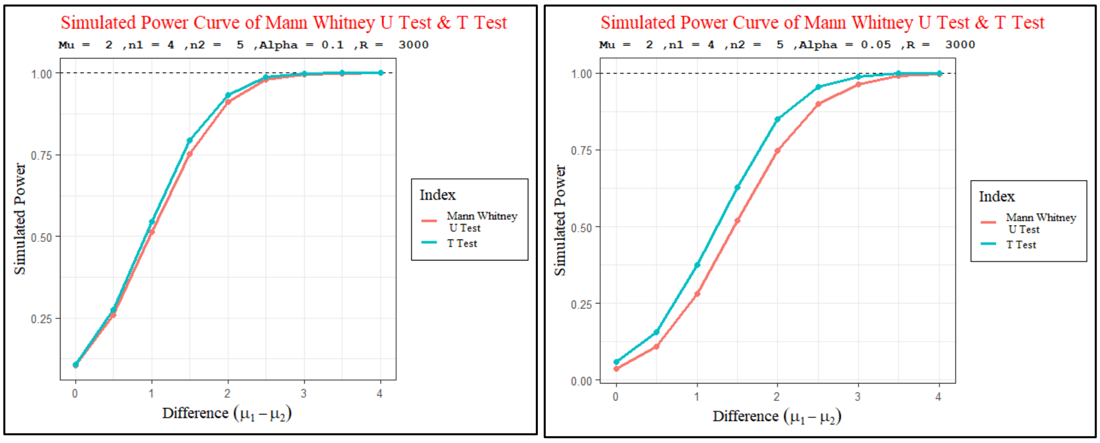


Diagram 22: Simulated Power Curve of Mann Whitney U test & T test for $mu = 2$, $n_1 = 4, n_2 = 5, R = 3000$ and $\alpha = 0.10,0.05$

Now, we will take $n_1 = 10, n_2 = 8$, we will construct the table of empirical size and power of Mann Whitney U test and T test side by side for $mu = 20$ and $alpha = 0.10,0.05$.

Table 15: Empirical size and Power of Mann Whitney U Test and T test for $mu = 20$, $n_1 = 10, n_2 = 8$ and $alpha = 0.10,0.05$, $R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	T Test	Mann Whitney U Test	T Test
0.0	0.08400	0.09550	0.04100	0.04700
0.5	0.34350	0.39200	0.22100	0.25950
1.0	0.74900	0.80050	0.60250	0.67250
1.5	0.94950	0.96700	0.88750	0.92150
2.0	0.99250	0.99550	0.98100	0.98800
2.5	1.00000	1.00000	0.99950	0.99950
3.0	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000
4.0	1.00000	1.00000	1.00000	1.00000

Here, the power reaches 1 more quickly than previous cases. Thus, as we increase the sample size power of both the test reaches 1 more quickly.

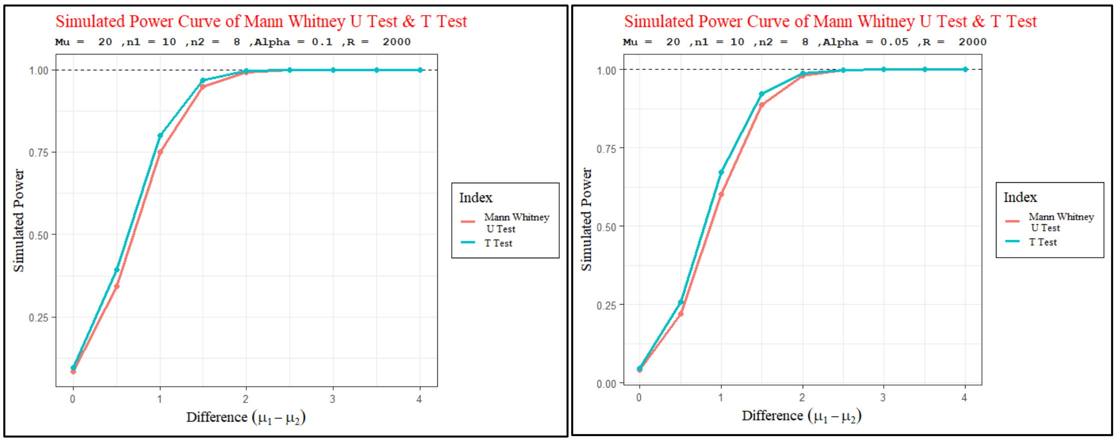


Diagram 23: Simulated Power Curve of Mann Whitney U test & T test for $mu = 20$, $n_1 = 10, n_2 = 8, R = 2000$ and $\alpha = 0.10, 0.05$

9.4.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is normal, T-test performs better than Mann Whitney U test in terms of power. i.e., the T-test is able to detect the same difference between the means of two independent normal population with same variance more frequently than Mann Whitney U test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are very small. As sample sizes increase the difference between the power of two test decreases. It is clear from the graphs also.
2. For small sample sizes (here, $n_1 = 5$ and $n_2 = 4$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance(alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.
3. The power of both the tests do not depend upon the value of mu (Y population mean). Other parameters remain fixed, if mu changes, it doesn't change the value of power. Instead of mu, if we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)
4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 40, n_2 = 40$ and $mu = 10$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

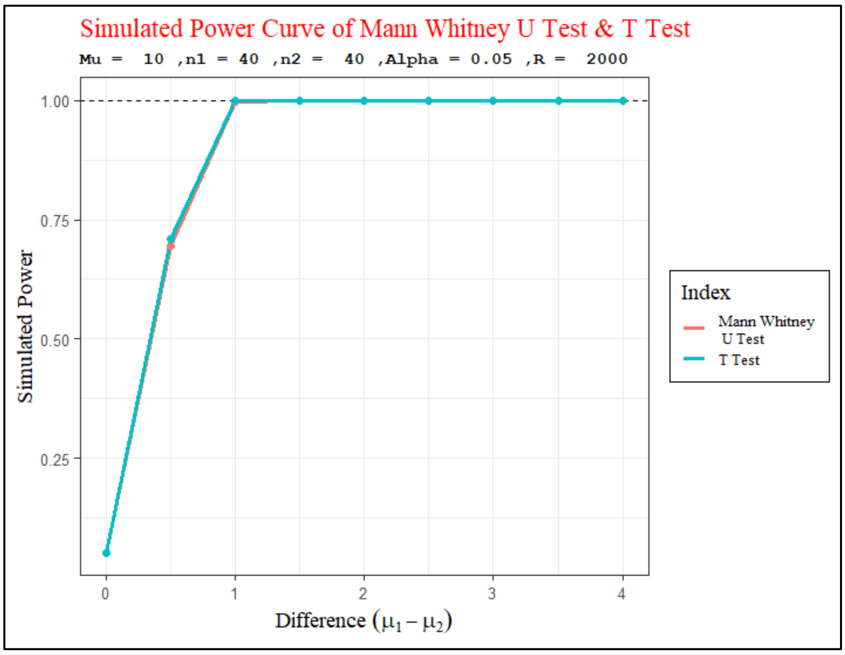


Diagram 24: Simulated Power Curve of Mann Whitney U test & T test for $mu = 10$, $n_1 = 40, n_2 = 40, R = 2000$ and $\alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except for $d = 0.5$. (Here, we have used large sample approximation for Mann-Whitney U test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.5. SIMULATION 04:

Comparison of Empirical Size and Power of Two sample Exponential LRT and Mann Whitney U Test When Sampled Population is Exponential

9.5.1. Motivation:

Mann-Whitney U test is one of the popular tests for two sample problems. The applicability of the test is wide. It is mainly useful for testing location or shift alternatives. On the other hand, two sample Exponential LRT is another well-known test for testing equality of means of independent exponential population. It is very interesting to study how the Mann-Whitney U test will perform for testing the equality of two means of two independent exponential population.

9.5.2. Objective:

We wish to compare empirical size and power of two sample Exponential LRT and Mann-Whitney U test when both the sampled population is Exponential. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.5.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.LRT2$. The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are similar as $Power\_comparison.LRT1$.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.LRT2$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.5.4. Discussion:

For the discussion we will consider $d = 0, 1, 2, 3, \ldots\ldots, 30$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using tables and graphs. We have used exact critical values from the table of Mann-Whitney U Statistic.

For $n_1 = 7, n_2 = 5$, we will construct the table of empirical size and power of Mann-Whitney U test and two sample Exponential LRT side by side for $lamda = 0.5, 1.5$ and $alpha = 0.10, 0.05$.

Table 16: Empirical size and Power of Mann Whitney U Test and Two sample Exponential LRT for $lamda = 0.5$, $n_1 = 7, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	Two sample Exponential LRT	Mann Whitney U Test	Two sample Exponential LRT
0	0.07167	0.09867	0.03300	0.05267
1	0.52200	0.68500	0.36300	0.53467
2	0.75267	0.91267	0.61200	0.82833

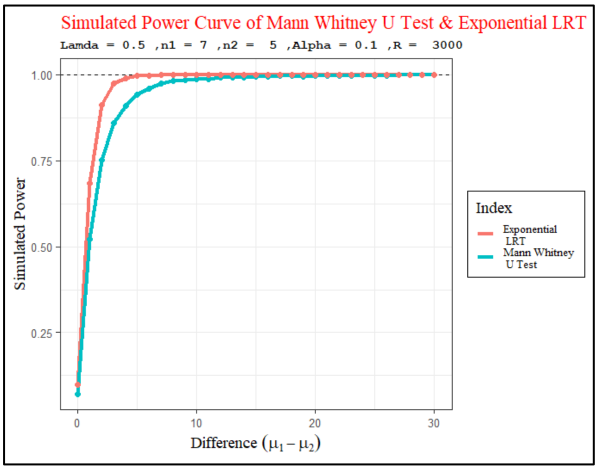
3	0.85900	0.97467	0.75800	0.94133
4	0.91000	0.99000	0.82667	0.97267
5	0.94200	0.99767	0.87800	0.99000
6	0.96033	0.99733	0.90733	0.99467
7	0.97567	0.99967	0.93667	0.99733
8	0.98133	0.99933	0.94233	0.99833
9	0.98400	0.99967	0.95067	0.99900
10	0.98833	1.00000	0.96733	1.00000
…..	……	…..	……	…..
20	0.99800	1.00000	0.99233	1.00000
21	0.99700	1.00000	0.99333	1.00000
22	0.99733	1.00000	0.99200	1.00000
…..	……	…..	……	…..
29	0.99900	1.00000	0.99433	1.00000
30	0.99900	1.00000	0.99533	1.00000

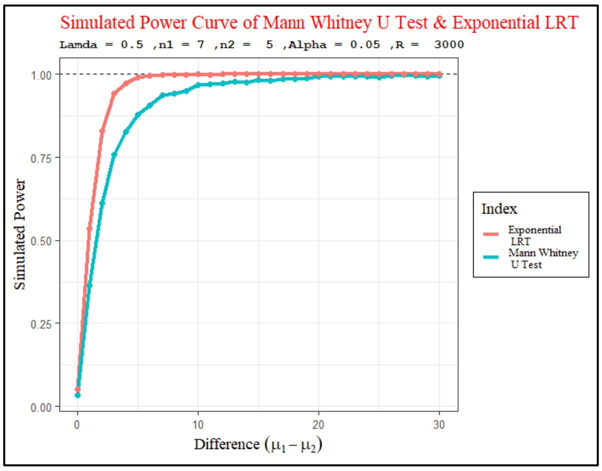
Table 17: Empirical size and Power of Mann Whitney U Test and Two sample Exponential LRT for $lamda = 1.5$, $n_1 = 7, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	Two sample Exponential LRT	Mann Whitney U Test	Two sample Exponential LRT
0	0.07167	0.09867	0.03300	0.05267
1	0.23367	0.31500	0.13200	0.20533
2	0.37933	0.52500	0.25367	0.37267
3	0.53100	0.70500	0.37533	0.55667
4	0.62200	0.80233	0.45100	0.66233
5	0.70533	0.87400	0.54333	0.77067
6	0.75233	0.91133	0.60633	0.84033
7	0.81300	0.94267	0.68967	0.88333
8	0.82933	0.96000	0.70900	0.91467
9	0.86033	0.97500	0.75233	0.94433
10	0.88467	0.98000	0.77633	0.95467

…..	……	…..	……	…..
15	0.94233	0.99767	0.87000	0.98600
16	0.94800	0.99500	0.88933	0.98700
17	0.96067	0.99733	0.90200	0.99233
18	0.95633	0.99833	0.90033	0.99400
19	0.96667	0.99933	0.91900	0.99500
20	0.97233	0.99967	0.93167	0.99700
…..	……	…..	……	…..
26	0.98433	1.00000	0.95267	0.99867
27	0.98600	1.00000	0.96167	0.99800
28	0.98567	1.00000	0.96433	0.99867
29	0.98667	1.00000	0.96467	0.99967
30	0.98867	1.00000	0.95900	0.99933

Notice, for different values of $lamda$ we get different table of empirical size and power. Which is not in case, when sampled population is normal. Actually, exponential distribution is characterized by mean of the distribution. The variance of the distribution also depends upon the value of mean of the distribution (Which is not in case of Normal Distribution.). That is why, as value of $lamda$ changes, power of both the test changes. It is empirically observed that as value of $lamda$ increases or decreases, the power of both the test decreases or increases.





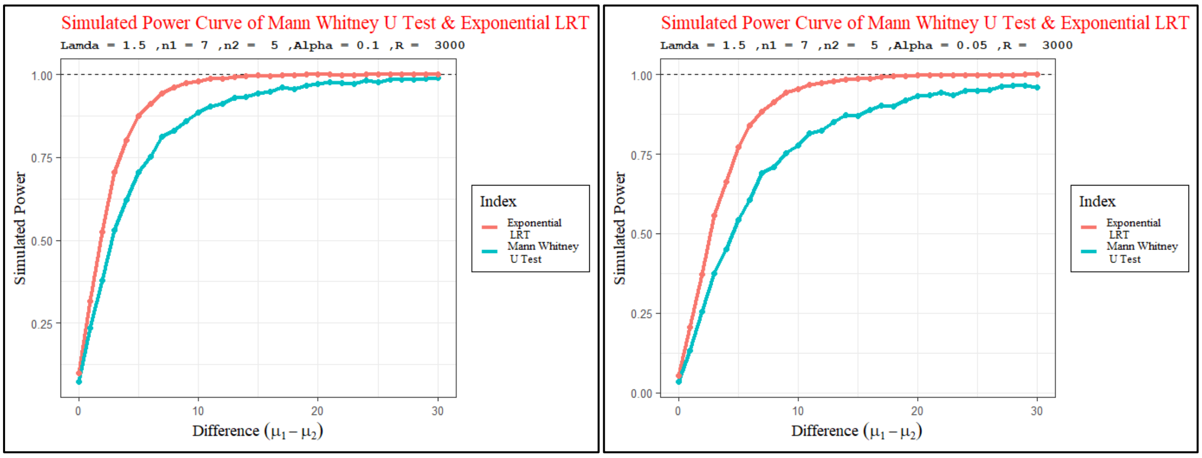


Diagram 25: Simulated Power Curve of Mann Whitney U test & Exponential LRT for $lamda = 0.5{,}1.5,\ n_1 = 7, n_2 = 5, R = 3000\ and\ \alpha = 0.10{,}0.05$

One thing to notice is that, the power of two sample Exponential LRT more quickly reaches 1 than Mann Whitney U Test and over all values of difference (d) two sample Exponential LRT is uniformly powerful than Mann Whitney U Test. If you closely observe the table of values of powers above, you will observe that the empirical power function is not monotone. It may happen due to sampling fluctuations.

In simulation-2, for $lamda = 1$, we have observed that the power of the two sample Exponential LRT reaches 1 very slowly. But here you observe that for $lamda = 0.5$, power reaches 1, at $d = 13$. But for $lamda = 1.5$, power reaches 1 at $d = 26$. It is mainly due to the fact that as $lamda$ increases (Keeping other test parameters constant) for same difference we will have less power.

Now, for $lamda = 1$ and $n_1 = 10, n_2 = 10$, we get these set of values for power.

Table 18: Empirical size and Power of Mann Whitney U Test and Two sample Exponential LRT for $lamda = 1, n_1 = 10, n_2 = 10$ and $alpha = 0.10{,}0.05, R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	Two sample Exponential LRT	Mann Whitney U Test	Two sample Exponential LRT
0	0.10250	0.10400	0.0530	0.0490
1	0.50300	0.59250	0.3445	0.4530
2	0.76750	0.88500	0.6175	0.7930

3	0.89150	0.96650	0.7845	0.9225
4	0.93950	0.98600	0.8700	0.9665
5	0.96550	0.99500	0.9105	0.9865
6	0.98250	0.99850	0.9525	0.9945
7	0.98500	0.99900	0.9605	0.9970
8	0.99050	0.99850	0.9740	0.9975
9	0.99450	1.00000	0.9820	0.9990
10	0.99800	1.00000	0.9865	1.0000
….	……	……	……	…….
20	0.99900	1.00000	0.9980	1.0000
21	1.00000	1.00000	0.9990	1.0000
….	……	……	……	…….
29	1.00000	1.00000	1.0000	1.0000
30	1.00000	1.00000	1.0000	1.0000

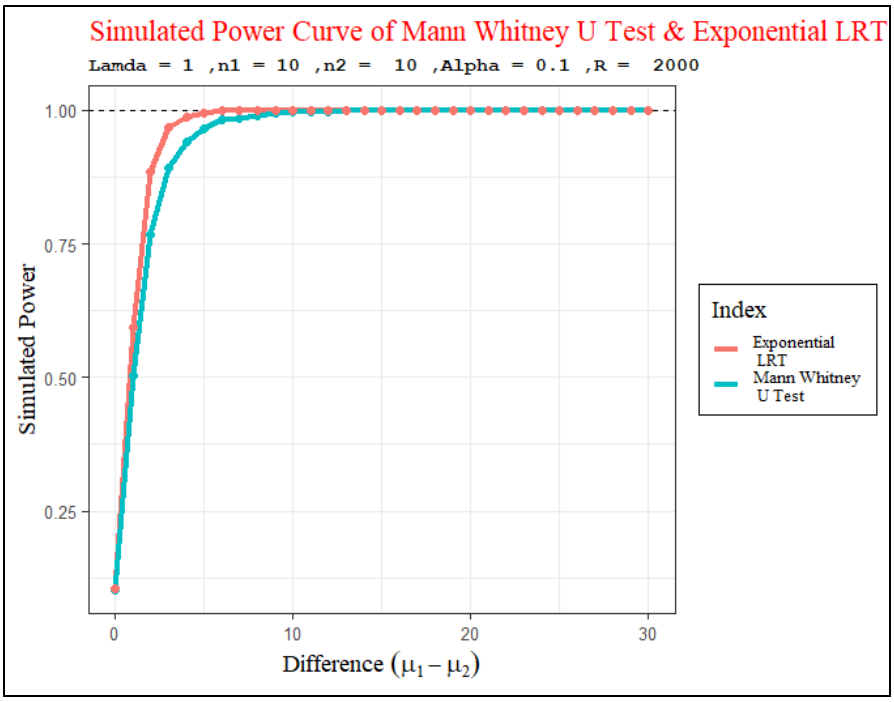


Diagram 26: Simulated Power Curve of Mann Whitney U test & Exponential LRT for $lamda = 1,\ n_1 = 10, n_2 = 10, R = 2000\ and\ \alpha = 0.10$

Now, for $lamda = 1$ and $n_1 = 18, n_2 = 12$, we get these set of values for power.

Table 19: Empirical size and Power of Mann Whitney U Test and Two sample Exponential LRT for $lamda = 1$, $n_1 = 18, n_2 = 12$ and $alpha = 0.10, 0.05$, $R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Mann Whitney U Test	Two sample Exponential LRT	Mann Whitney U Test	Two sample Exponential LRT
0	0.10200	0.09400	0.04350	0.04200
1	0.61800	0.70100	0.46250	0.57400
2	0.89750	0.95800	0.79000	0.90100
3	0.96500	0.99150	0.91850	0.98000
4	0.98950	0.99950	0.97250	0.99650
5	0.99450	0.99950	0.98400	0.99850
6	0.99750	1.00000	0.99400	0.99950
7	0.99950	1.00000	0.99750	1.00000
8	1.00000	1.00000	0.99900	1.00000
9	1.00000	1.00000	1.00000	1.00000
…..	…..	……	…….	……
29	1.00000	1.00000	1.00000	1.00000
30	1.00000	1.00000	1.00000	1.00000

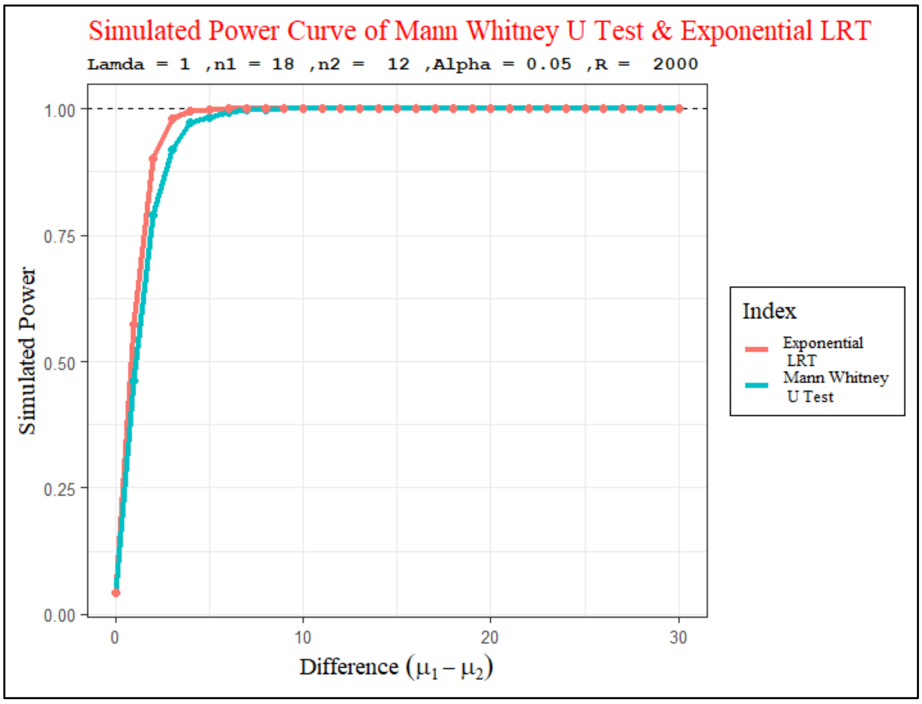


Diagram 27: Simulated Power Curve of Mann Whitney U test & Exponential LRT for $lamda = 1,\ n_1 = 18, n_2 = 12, R = 2000\ and\ \alpha = 0.05$

From the above graphs, we see that as sample sizes increases power increases more quickly and also the difference between powers of two tests decreases.

9.5.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is exponential, two sample Exponential LRT performs better than Mann-Whitney U test in terms of power. i.e., the two sample Exponential LRT is able to detect the same difference between the means of two independent normal population with same variance more frequently than Mann-Whitney U test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are not very high. As sample sizes increase the difference between the power of two test decreases. It is clear from the graph also.
2. For small sample sizes (here, $n_1 = 7\ and\ n_2 = 5$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance(alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.

3. The power of both the tests depends upon the value of $lamda$ (Y population mean). If $lamda$ increases or decreases, the value of power decreases or increases.
4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 45, n_2 = 45$ and $lamda = 1.5$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

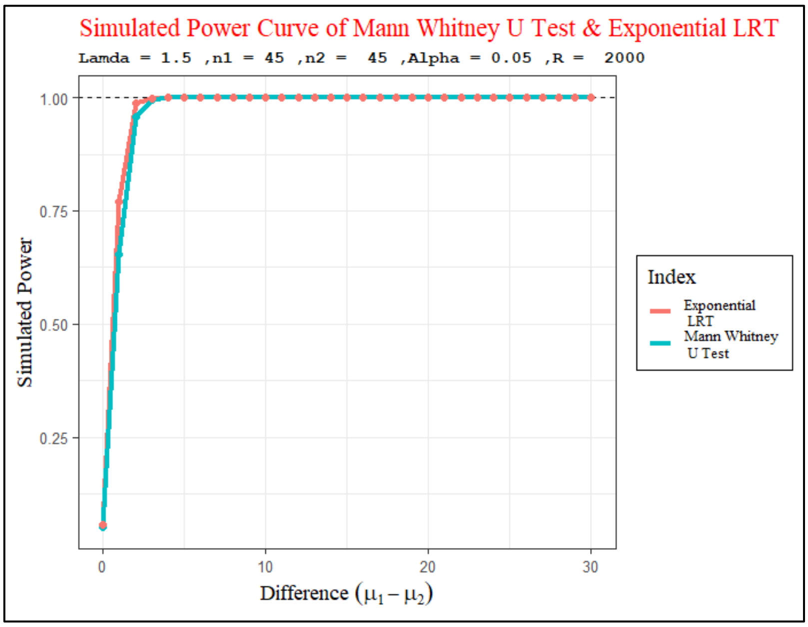


Diagram 28: Simulated Power Curve of Mann Whitney U test & Exponential LRT for $lamda = 1.5,\ n_1 = 45, n_2 = 45, R = 2000\ and\ \alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except some differences. (Here, we have used large sample approximation for Mann-Whitney U test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.6. SIMULATION 05:

Comparison of Empirical Size and Power of Two sample T Test and Wilcoxon Rank Sum Test When Sampled Population is Normal

9.6.1. Motivation:

Wilcoxon Rank Sum test is one of the popular tests for two sample problems. The applicability of the test is wide. It is similar as Mann Whitney U Test. Like MWU test, it is mainly useful for testing location or shift alternatives. On the other hand, Two Sample T test is another well-known test for testing equality of means of two independent normal populations. The main assumption of this test is that the two normal population differs with respect to mean only, that is variance is same. It is very interesting to study how the Wilcoxon Rank Sum test will perform for testing the equality of two means of two independent normal population.

9.6.2. Objective:

We wish to compare empirical size and power of T test and Wilcoxon Rank Sum Test when both the sampled population is Exponential. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.6.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.T3$ The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are similar as $Power\_comparison.T1$.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.T3$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.6.4. Discussion:

For the discussion we will take variance of the two population as 1. We will consider $d = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using table and graphs. We have used exact critical values from the table of Wilcoxon Rank Sum Statistic. However, for large sample sizes we have used large sample approximation.

For $n_1 = 5, n_2 = 4$, we will construct the table of empirical size and power of Wilcoxon Rank Sum Test and T test side by side for $mu = 0$ and $alpha = 0.10,0.05$.

Table 20: Empirical size and Power of Wilcoxon Rank Sum test and T test for $mu = 0$, $n_1 = 5, n_2 = 4$ and $alpha = 0.10,0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	T Test	Wilcoxon Rank Sum Test	T Test

0.0	0.11200	0.12033	0.04100	0.06500
0.5	0.25467	0.26667	0.10900	0.15700
1.0	0.53067	0.56633	0.28300	0.38933
1.5	0.76100	0.79367	0.52633	0.64133
2.0	0.91133	0.92600	0.74467	0.84067
2.5	0.98100	0.98733	0.90200	0.95767
3.0	0.99700	0.99867	0.96567	0.99067
3.5	0.99967	1.00000	0.98967	0.99933
4.0	1.00000	1.00000	0.99733	1.00000

As in case of Kolmogorov Smirnov Test and Mann Whitney U Test (Sampled Population is Normal), here also for different values mu, we will get same table of empirical size and power. Which is very much intuitive. If we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)

Notice that the values of empirical size and power is same, as in case of Mann Whitney U test (for same values of parameter). It is not a coincidence. We have already discussed theoretically that the Mann-Whitney U test and Wilcoxon Rank Sum test are exactly same in terms of power (for same size). They differ only by the value of test statistic and this difference in the value of test statistics depends upon only sample size drawn from X population. Here, we have empirically justified the theory.

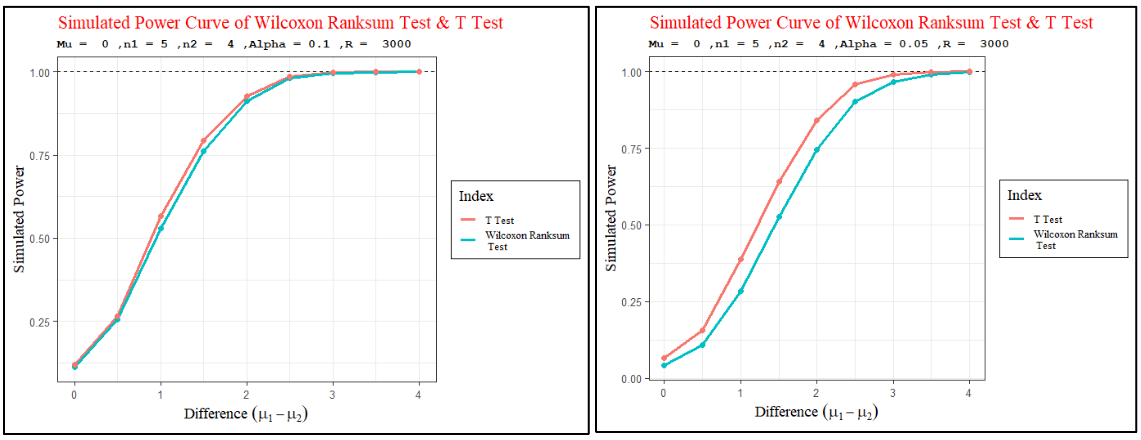


Diagram 29: Simulated Power Curve of Wilcoxon Rank Sum test & T test for $mu = 0$, $n_1 = 5, n_2 = 4, R = 3000$ and $\alpha = 0.10, 0.05$

From the above table, we can say that the power of T-test more quickly reaches 1 than Wilcoxon Rank Sum Test and over all values of difference (d) T-test is uniformly powerful than Wilcoxon Rank Sum Test. It is evident from the graph also.

Actually, the properties that we have empirically studied in case of Mann Whitney U test are same here also. Here, also the empirical power is not symmetric in sample size. But it doesn't mean that power of two tests is not symmetric in sample sizes (n_1,n_2). It may happen due to sample fluctuations. Until we know the exact analytical form of power functions, we cannot say this explicitly.

Now, if we take large sample sizes than previous, say, $n_1 = 10, n_2 = 10$. And $mu = 0, \alpha = 0.10, 0.05$. Then, we see that the power increases more rapidly than previous.

Table 21: Empirical size and Power of Wilcoxon Rank Sum test and T test for $mu = 0$, $n_1 = 10, n_2 = 10$ and $alpha = 0.10{,}0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	T Test	Wilcoxon Rank Sum Test	T Test
0.0	0.09667	0.10167	0.04533	0.05033
0.5	0.39833	0.41667	0.26133	0.27867
1.0	0.78300	0.81167	0.64867	0.68567
1.5	0.97033	0.97900	0.92200	0.93967
2.0	0.99900	0.99900	0.99500	0.99700
2.5	1.00000	1.00000	1.00000	1.00000
3.0	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000
4.0	1.00000	1.00000	1.00000	1.00000

From the above table we see that, even if at $d = 2.5$, the empirical power already reaches 1. The empirical power curves from the above simulation are given below.

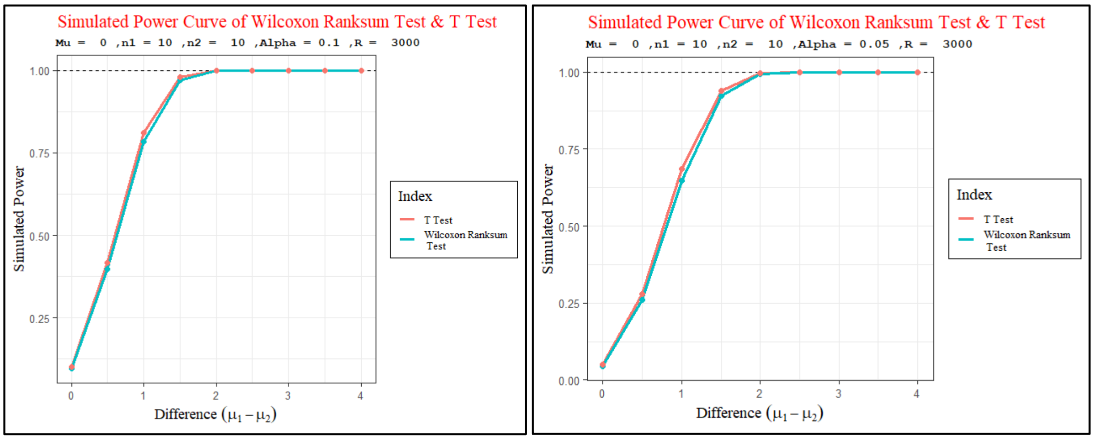


Diagram 30: Simulated Power Curve of Wilcoxon Rank Sum test & T test for $mu = 0$, $n_1 = 10, n_2 = 10, R = 3000$ *and* $\alpha = 0.10, 0.05$

9.6.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is normal, T-test performs better than Wilcoxon Rank Sum test in terms of power. i.e., the T-test is able to detect the same difference between the means of two independent normal population with same variance more frequently than Wilcoxon Rank Sum test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are very small. As sample sizes increase the difference between the power of two test decreases. It is clear from the graphs also.
2. For small sample sizes (here, $n_1 = 5$ *and* $n_2 = 4$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance (alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.
3. The power of both the tests do not depend upon the value of mu (Y population mean). Other parameters remain fixed, if mu changes, it doesn't change the value of power. Instead of mu, if we change the value of common variance, then we will see that the values of empirical size and power will change. (Actually, value of empirical size will remain near α, but as variance increases or decreases power will decrease or increase accordingly.)
4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 35, n_2 = 35$ and mu = 2, Alpha = 0.10, then the empirical power curves of the two tests are given by –

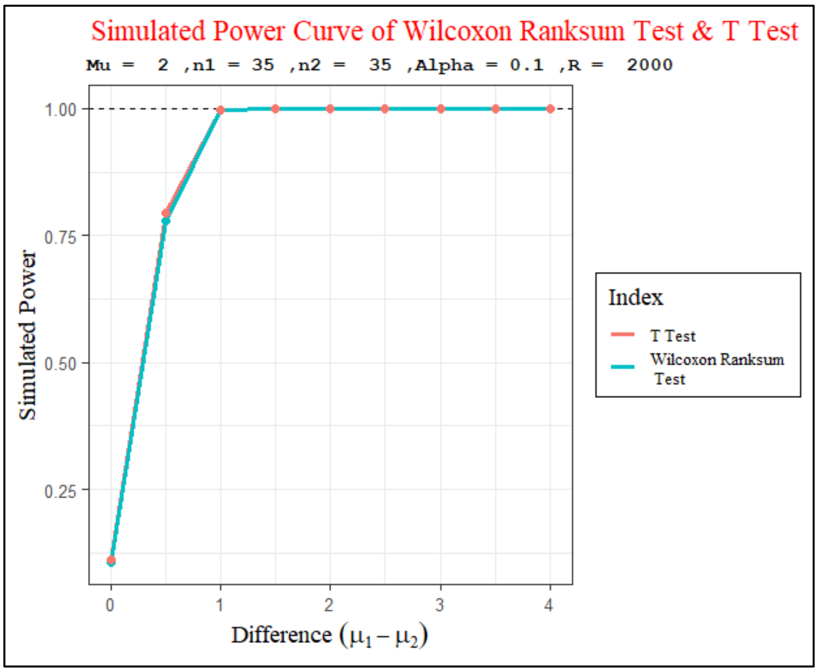


Diagram 31: Simulated Power Curve of Wilcoxon Rank Sum test & T test for $mu = 2$, $n_1 = 35, n_2 = 35, R = 2000$ and $\alpha = 0.10$

From the above graph we can see that, the simulated powers are 1, except for $d = 0.5$. (Here, we have used large sample approximation for Wilcoxon Rank Sum test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.7. SIMULATION 06:

Comparison of Empirical Size and Power of Two sample Exponential LRT and Wilcoxon Rank Sum Test When Sampled Population is Exponential

9.7.1. Motivation:

Wilcoxon Rank Sum test is one of the popular tests for two sample problems. The applicability of the test is wide. It is similar as Mann Whitney U Test. Like MNW U test, it is mainly useful for testing location or shift alternatives. On the other hand, two sample Exponential LRT is another well-known test for testing equality of means of independent exponential population. It is very interesting to study how the Wilcoxon Rank Sum test will perform for testing the equality of two means of two independent exponential population.

9.7.2. Objective:

We wish to compare empirical size and power of two sample Exponential LRT and Wilcoxon Rank Sum test when both the sampled population is Exponential. We will compare them for varying sample size, different level of significances, different values of $\mu_1 - \mu_2$.

9.7.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparison.LRT3$. The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are similar as $Power\_comparison.LRT1$.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparison.LRT3$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.7.4. Discussion:

For the discussion we will consider $d = 0, 1, 2, 3, \ldots\ldots, 30$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using tables and graphs. We have used exact critical values from the table of Wilcoxon Rank Sum Statistic.

For $n_1 = 7, n_2 = 5$, we will construct the table of empirical size and power of Wilcoxon Rank Sum test and two sample Exponential LRT side by side for $lamda = 0.5, 2.0$ and $alpha = 0.10, 0.05$.

Table 22: Empirical size and Power of Wilcoxon Rank Sum test and Two sample Exponential LRT for $lamda = 0.5$, $n_1 = 7, n_2 = 5$ and $alpha = 0.10, 0.05$, $R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	Two sample Exponential LRT	Wilcoxon Rank Sum Test	Two sample Exponential LRT
0	0.07167	0.09867	0.03300	0.05267
1	0.52200	0.68500	0.36300	0.53467
2	0.75267	0.91267	0.61200	0.82833

3	0.85900	0.97467	0.75800	0.94133
4	0.91000	0.99000	0.82667	0.97267
5	0.94200	0.99767	0.87800	0.99000
6	0.96033	0.99733	0.90733	0.99467
7	0.97567	0.99967	0.93667	0.99733
8	0.98133	0.99933	0.94233	0.99833
9	0.98400	0.99967	0.95067	0.99900
10	0.98833	1.00000	0.96733	1.00000
…..	……	…..	……	…..
20	0.99800	1.00000	0.99233	1.00000
21	0.99700	1.00000	0.99333	1.00000
22	0.99733	1.00000	0.99200	1.00000
…..	……	…..	……	…..
28	0.99967	1.00000	0.99633	1.00000
29	0.99900	1.00000	0.99433	1.00000
30	0.99900	1.00000	0.99533	1.00000

Notice that the values of empirical size and power is same, as in case of Mann Whitney U test (for same values of parameter). It is not a coincidence. We have already discussed theoretically that the Mann-Whitney U test and Wilcoxon Rank Sum test are exactly same in terms of power (for same size). They differ only by the value of test statistic and this difference in the value of test statistics depends upon only sample size drawn from X population. Here, we have empirically justified the theory.

Now, if we take a different value of $lamda$, say $lamda = 2$. We see that we get different values of empirical size and power as compared to $lamda = 0.5$.

Table 23: Empirical size and Power of Wilcoxon Rank Sum test and Two sample Exponential LRT for $lamda = 2.0$, $n_1 = 7, n_2 = 5$ and $alpha = 0.10, 0.05, R = 3000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	Two sample Exponential LRT	Wilcoxon Rank Sum Test	Two sample Exponential LRT
0	0.07167	0.09867	0.03300	0.05267
1	0.18667	0.26367	0.10433	0.16167
2	0.30433	0.42233	0.18933	0.27767

3	0.43367	0.59033	0.27900	0.43400
4	0.50833	0.68433	0.35000	0.53200
5	0.60133	0.77800	0.44233	0.63867
6	0.65067	0.84467	0.49900	0.71633
7	0.73300	0.88733	0.58633	0.78900
8	0.74700	0.91567	0.61567	0.83533
9	0.78767	0.94433	0.65700	0.86833
10	0.81767	0.95300	0.68900	0.89833
…..	……	…..	……	…..
20	0.95000	0.99667	0.88433	0.99133
21	0.95333	0.99733	0.89200	0.99133
22	0.95900	0.99867	0.90333	0.99467
…..	……	…..	……	…..
28	0.97467	0.99800	0.93833	0.99567
29	0.97300	0.99967	0.93700	0.99800
30	0.97500	0.99933	0.93867	0.99833

Notice, for different values of $lamda$ we get different table of empirical size and power. Which is not in case, when sampled population is normal. Actually, exponential distribution is characterized by mean of the distribution. The variance of the distribution also depends upon the value of mean of the distribution (Which is not in case of Normal Distribution.). Here, $lamda$ is actually the Y population mean and X population mean is $lamda + d$. That is why, as value of $lamda$ changes, power of both the test changes. It is empirically observed that as value of $lamda$ increases or decreases, the power of both the test decreases or increases.

One thing to notice is that, the power of two sample Exponential LRT more quickly reaches 1 than Wilcoxon Rank Sum Test and over all values of difference (d) two sample Exponential LRT is uniformly powerful than Wilcoxon Rank Sum Test. If you closely observe the table of values of powers above, you will observe that the empirical power function is not monotone. It may happen due to sampling fluctuations. It was observed in case of Mann Whitney U test also. Actually, the properties that we have empirically studied are in case of Mann Whitney U test is same here also.

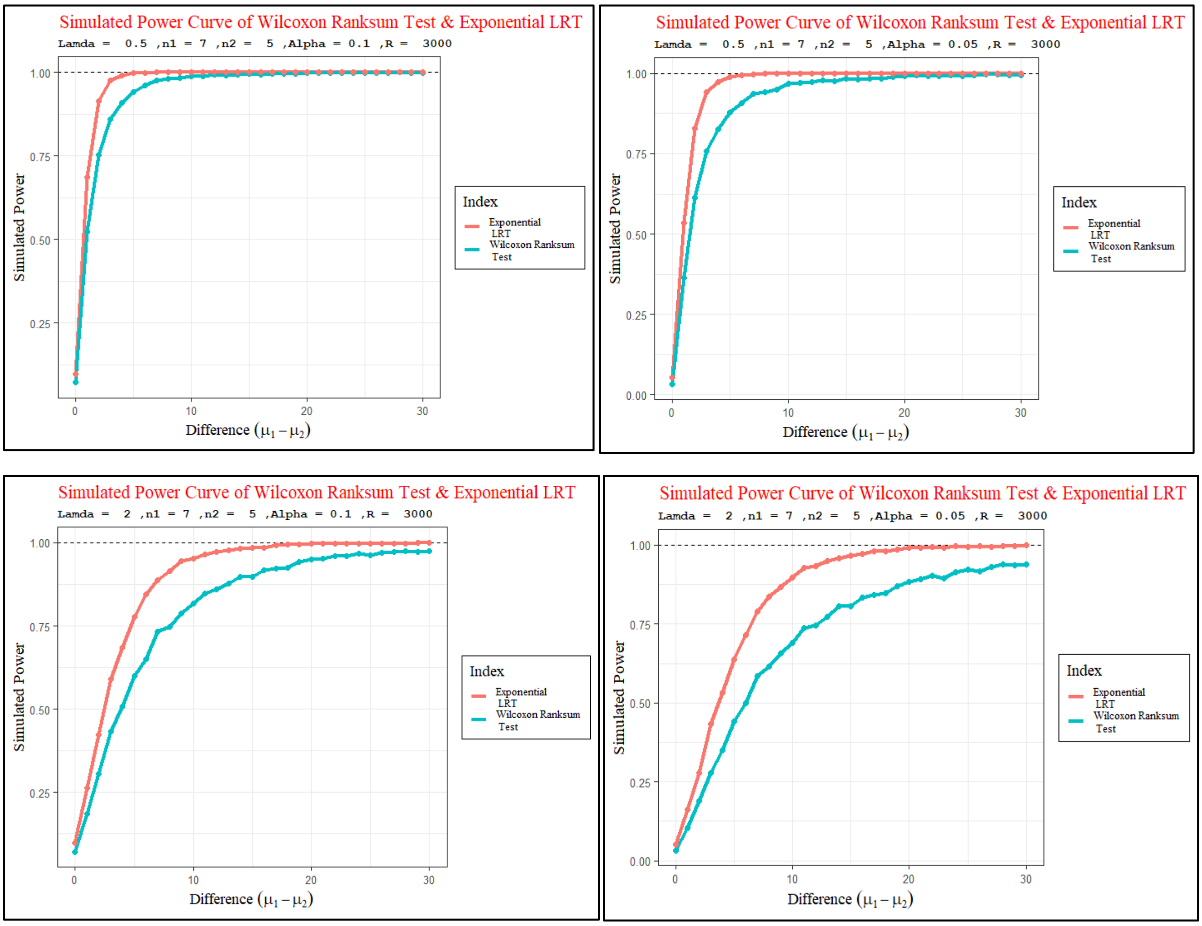


Diagram 32: Simulated Power Curve of Wilcoxon Rank Sum test & Exponential LRT for $lamda = 0.5,2,\ n_1 = 7, n_2 = 5, R = 3000\ and\ \alpha = 0.10,0.05$

If we take large sample sizes, we will see that the empirical power of both the tests reaches 1 more quickly as compared to small sample sizes.

Now, for $lamda = 2$ and $n_1 = 14, n_2 = 14$, we get these set of values for power.

Table 24: Empirical size and Power of Wilcoxon Rank Sum test and Two sample Exponential LRT for $lamda = 2.0$, $n_1 = 14, n_2 = 14$ and $alpha = 0.10,0.05, R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	Two sample Exponential LRT	Wilcoxon Rank Sum Test	Two sample Exponential LRT
0	0.08500	0.09650	0.0420	0.0525
1	0.34700	0.40750	0.2260	0.2650
2	0.60050	0.70750	0.4515	0.5645

3	0.76050	0.86100	0.6185	0.7575
4	0.88350	0.94750	0.7740	0.8845
5	0.92350	0.97150	0.8485	0.9450
6	0.95200	0.99250	0.8895	0.9805
7	0.97700	0.99750	0.9415	0.9915
8	0.98650	0.99900	0.9595	0.9945
9	0.98950	0.99750	0.9715	0.9950
10	0.99450	0.99850	0.9815	0.9965
…..	…….	…….	……..	……..
21	0.99950	1.00000	0.9995	1.0000
22	1.00000	1.00000	0.9995	1.0000
…..	…….	…….	……..	……..
29	1.00000	1.00000	1.0000	1.0000
30	1.00000	1.00000	1.0000	1.0000

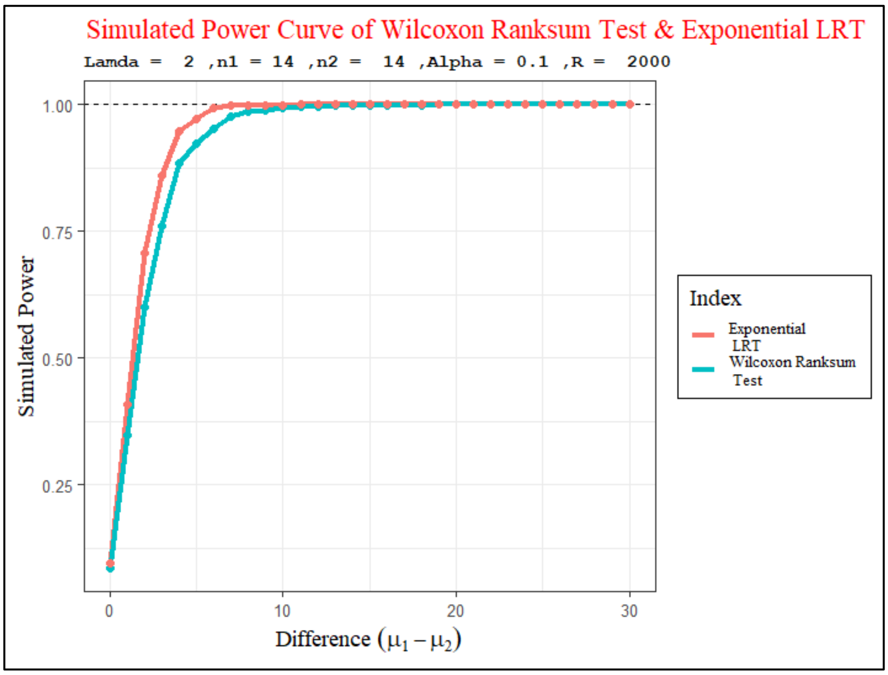


Diagram 33: Simulated Power Curve of Wilcoxon Rank Sum test & Exponential LRT for $lamda = 2,\ n_1 = 14, n_2 = 14, R = 2000\ and\ \alpha = 0.10$

Now, for $lamda = 1$ and $n_1 = 20, n_2 = 15$, we get these set of values for power.

Table 25: Empirical size and Power of Wilcoxon Rank Sum test and Two sample Exponential LRT for $lamda = 1.0$, $n_1 = 20, n_2 = 15$ and $alpha = 0.10, 0.05, R = 2000$

Difference (d)	Level of Significance (Alpha)			
	0.10		0.05	
	Wilcoxon Rank Sum Test	Two sample Exponential LRT	Wilcoxon Rank Sum Test	Two sample Exponential LRT
0	0.10150	0.10600	0.05350	0.06050
1	0.68100	0.76950	0.53700	0.64750
2	0.92200	0.97250	0.84750	0.94200
3	0.98350	0.99650	0.96050	0.99200
4	0.99350	0.99900	0.98100	0.99800
5	0.99800	1.00000	0.99650	0.99950
6	0.99950	1.00000	0.99900	1.00000
7	1.00000	1.00000	0.99900	1.00000
8	1.00000	1.00000	1.00000	1.00000
…..	…….	…….	……	……..
30	1.00000	1.00000	1.00000	1.00000

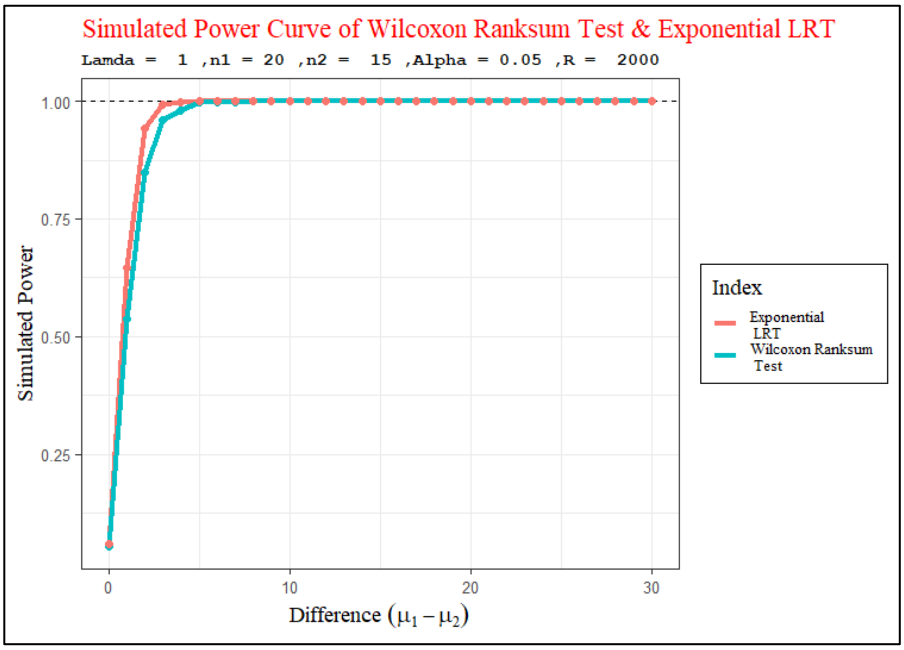


Diagram 34: Simulated Power Curve of Wilcoxon Rank Sum test & Exponential LRT for $lamda = 1, n_1 = 20, n_2 = 15, R = 2000$ and $\alpha = 0.05$

9.7.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is exponential, two sample Exponential LRT performs better than Wilcoxon Rank Sum test in terms of power. i.e., the two sample Exponential LRT is able to detect the same difference between the means of two independent normal population with same variance more frequently than Wilcoxon Rank Sum test. However, from the table of values of powers, we can notice that the difference between the empirical powers of the two tests are not very high. As sample sizes increase the difference between the power of two test decreases. It is clear from the graph also.
2. For small sample sizes (here, $n_1 = 7 \; and \; n_2 = 5$) the power of both the tests reaches 1 very slowly. However, as sample size increases the power of both the tests reaches 1 very quickly. For small level of significance(alpha) the power reaches 1 less quickly than large level of significance(alpha). In our discussion, for $\alpha = 0.10$ the power reaches 1 more quickly than $\alpha = 0.05$.
3. The power of both the tests depends upon the value of $lamda$ (Y population mean). If $lamda$ increases or decreases, the value of power decreases or increases.
4. Both the tests are consistent test. Because, as sample sizes increase, the power of both the test tends to 1. For example, if we take $n_1 = 38, n_2 = 38$ and $lamda = 2$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

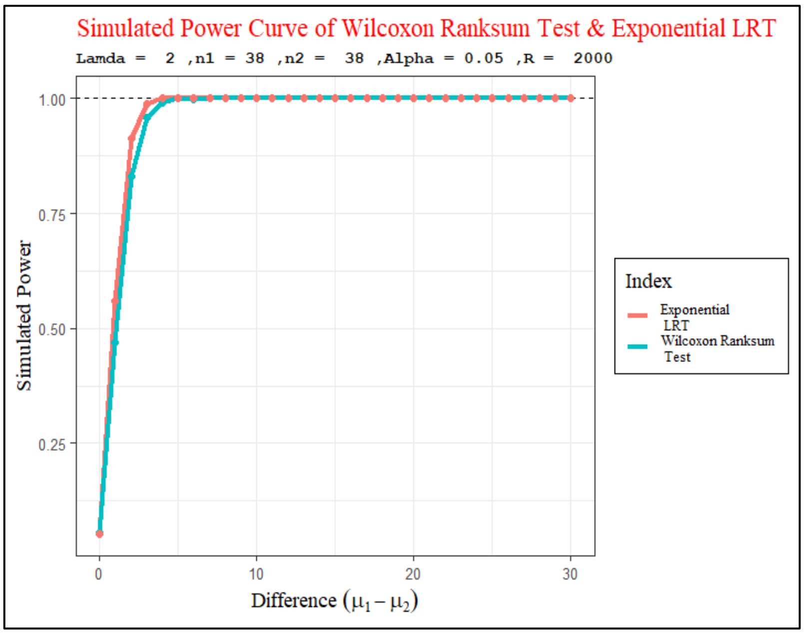


Diagram 35: Simulated Power Curve of Wilcoxon Rank Sum test & Exponential LRT for $lamda = 2, \; n_1 = 38, n_2 = 38, R = 2000 \; and \; \alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except some differences. (Here, we have used large sample approximation for Wilcoxon Rank Sum test, as tables of critical values are not available for those large values of n_1 and n_2.)

9.8. SIMULATION 07:

Simultaneous Comparison of Empirical Size and Power of Two sample T Test and Different Non-Parametric Tests When Sampled Population is Normal

9.8.1. Motivation:

Two sample T test is very well-known test for two sample problem, when sampled population is normal. This test assumes that the variances of the two independent normal populations are same. That is, if the two population differs, they will differ only by their means. On the other hand, non-parametric tests are also used for two sample problems. Here, we want to make comparative study between the two sample T test and different non-parametric tests for two sample problem of equality of means of two independent populations, when sampled population is normal. We have already made a comparative study of T-test with Kolmogorov-Smirnov Test, Mann Whitney U test and Wilcoxon Rank Sum test separately. So, here we will compare all of them simultaneously.

9.8.2. Objective:

We will study how Kolmogorov Smirnov test, Mann Whitney U test, Wilcoxon Rank Sum test will behave in respect to power, when sampled populations are normal. So, we will make a comparative study of empirical size, empirical power of all of these tests. We will compare them for varying sample sizes, different level of significances, different values of $\mu_1 - \mu_2$.

9.8.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparisonALL.1$. The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

mu: Mean of Y population.

d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $mu + d$. Where, mu is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.
$alpha$: It is level of significance of the test.
$exact.crit$: It is vector of critical values of different non-parametric tests. By default, it is set at $c(F, F, F)$. If we change any of the element of the vector to exact critical value (For example, $c(F, 21, F)$), then the function calculates empirical size and power using that exact critical value. We need to give exact critical points of different tests (For the given sample size and level of significance) as input. Otherwise, it will calculate empirical size and power using large sample approximation of the test-statistic. The first element of the vector is critical value for Mann Whitney U test, second element is for Wilcoxon Rank Sum test and last element is for Kolmogorov Smirnov test.
From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparisonALL.1$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.8.4. Discussion:

For the discussion we will take variance of the two population as 1. We will consider $d = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$.
For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using tables and graphs. In most of the cases, we have used exact critical values from the tables of Kolmogorov-Smirnov test, Mann Whitney U test, Wilcoxon Rank Sum test. But for large sample sizes we have used large sample approximations.
For $n_1 = 4, n_2 = 5$, we will construct the table of empirical size and power of T test, Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test side by side for $mu = 5$ and $alpha = 0.10, 0.05$. Considering only one value of mu will give us sufficient idea. Because, we have already empirically seen that (other parameters kept fixed) if we change mu, it doesn't change power of the test.

Table 26: Empirical size and Power of Two sample T test and Different Non-parametric Tests For $n_1 = 4, n_2 = 5, mu = 5, R = 3000$ and $\alpha = 0.10, 0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0.0	0.10767	0.08033	0.10467	0.10467	0.05767	0.04733	0.03733	0.03733
0.5	0.27467	0.19533	0.25867	0.25867	0.15500	0.12500	0.10900	0.10900
1.0	0.54533	0.41100	0.51500	0.51500	0.37600	0.28167	0.28100	0.28100
1.5	0.79467	0.65300	0.75267	0.75267	0.62833	0.49400	0.51967	0.51967
2.0	0.93333	0.85233	0.91333	0.91333	0.84967	0.70833	0.74733	0.74733
2.5	0.98833	0.94700	0.97900	0.97900	0.95400	0.85833	0.90067	0.90067
3.0	0.99833	0.98667	0.99700	0.99700	0.98933	0.94967	0.96333	0.96333
3.5	1.00000	0.99767	0.99833	0.99833	0.99933	0.98400	0.99167	0.99167
4.0	1.00000	1.00000	1.00000	1.00000	0.99967	0.99433	0.99733	0.99733

From the above table, we can see that as value of d increases empirical power increases for all the tests. For a fixed level of significance, we have already seen that T test is uniformly more powerful than Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test.

Here also we are observing the same thing. Additionally, we see that among the non-parametric tests (which we are discussing) on an average Kolmogorov-Smirnov test has less power than other two non-parametric tests. For $\alpha = 0.10$, Kolmogorov Smirnov test has uniformly less power among all the tests discussed. But, for $\alpha = 0.05$, Initially, Kolmogorov Smirnov test has more power than Mann Whitney U test and Wilcoxon Rank sum test. But after a certain value of d, we see that the power of Mann Whitney U test and Wilcoxon Rank sum test is greater than Kolmogorov-Smirnov test.

We now present the empirical power curve of the above simulation below.

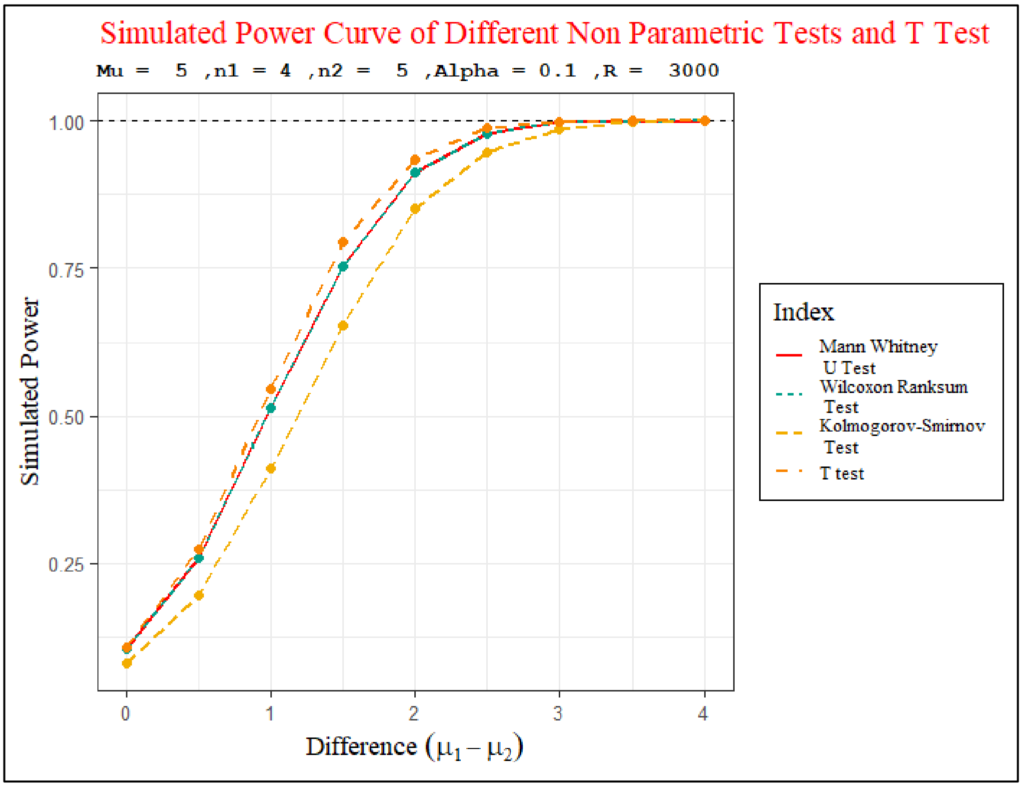


Diagram 36: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 5$, $n_1 = 4, n_2 = 5, R = 3000$ *and* $\alpha = 0.10$

From the above graph we can see that empirical power of Kolmogorov-Smirnov test is uniformly less than empirical power of Mann Whitney U test, Wilcoxon Rank Sum test and Two sample T test. Also, the empirical powers of Mann Whitney U test and Wilcoxon Rank Sum test are exactly same. We have already discussed about this equivalence between the two tests and also justified using simulation.

In the second graph we see that at first empirical power of Kolmogorov Smirnov test is greater than empirical power of Wilcoxon Rank Sum test and Mann Whitney u test. But later Kolmogorov-Smirnov test has again less empirical power. It may happen due to sampling fluctuation. Actually, in small sample sizes we can expect such behaviour.

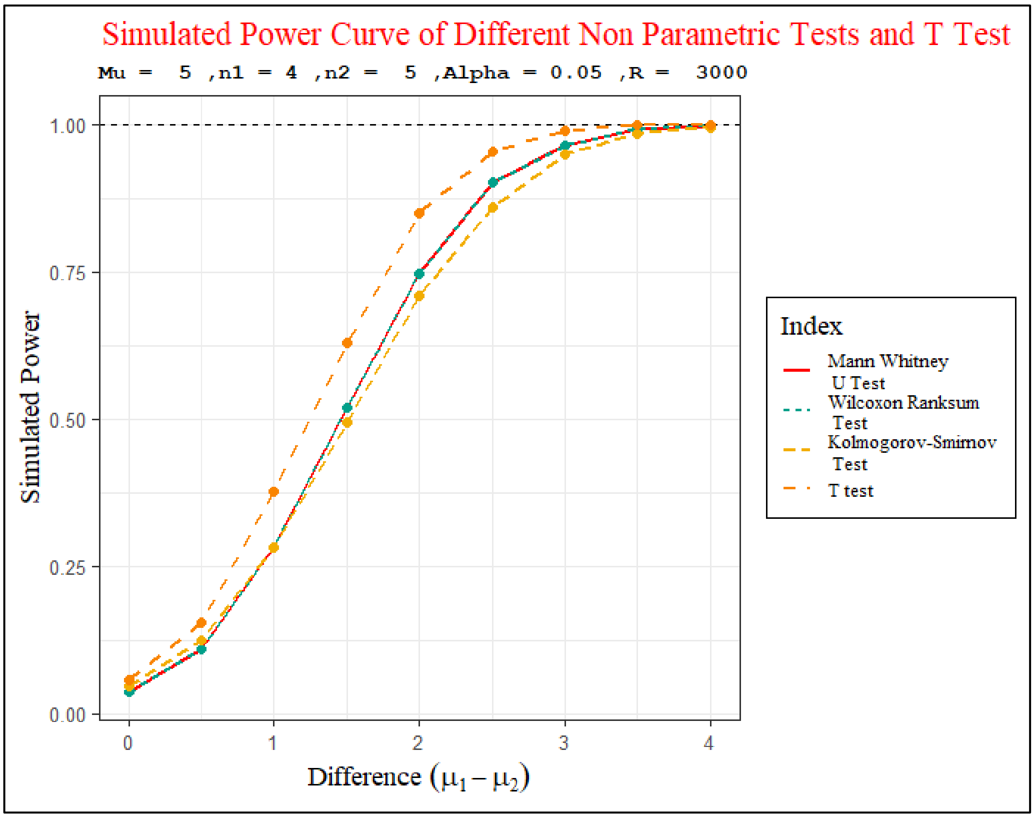


Diagram 37: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 5$, $n_1 = 4, n_2 = 5, R = 3000$ *and* $\alpha = 0.05$

As it is expected that, for $\alpha = 0.10$, the empirical power of all the tests reaches 1 more quickly than $\alpha = 0.05$.

Now, we will consider a different pair of sample size. We take $n_1 = 7, n_2 = 5$ and $mu = 20$. We present the results using table below.

From the table we see that for $\alpha = 0.10$, Kolmogorov-Smirnov Test has still uniformly less empirical power than other tests. But for, $\alpha = 0.05$, for some initial values of d, Kolmogorov-Smirnov Test has more empirical power than the other two non-parametric test. But after some values of d, Kolmogorov-Smirnov test has again less power than the other two non-parametric tests. So, again we see that, for $\alpha = 0.05$ we cannot say that Kolmogorov-Smirnov test has uniformly less power than other tests. These are just empirical observations, here it may happen that we are observing such things due to sampling fluctuations.

Table 27: Empirical size and Power of Two sample T test and Different Non-parametric Tests For $n_1 = 7, n_2 = 5, mu = 20, R = 3000\ and\ \alpha = 0.10, 0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0.0	0.10667	0.09100	0.08000	0.08000	0.05233	0.06100	0.03933	0.03933
0.5	0.33000	0.25967	0.26400	0.26400	0.19800	0.20100	0.15133	0.15133
1.0	0.65033	0.53167	0.55667	0.55667	0.47700	0.43767	0.39433	0.39433
1.5	0.88000	0.78433	0.81233	0.81233	0.76533	0.69733	0.67900	0.67900
2.0	0.97433	0.93567	0.94900	0.94900	0.93600	0.88767	0.89067	0.89067
2.5	0.99867	0.98200	0.99400	0.99400	0.99067	0.97300	0.97967	0.97967
3.0	1.00000	0.99867	0.99900	0.99900	0.99867	0.99433	0.99600	0.99600
3.5	1.00000	0.99967	1.00000	1.00000	1.00000	0.99967	0.99967	0.99967
4.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

We present the empirical power curves of the above simulations below.

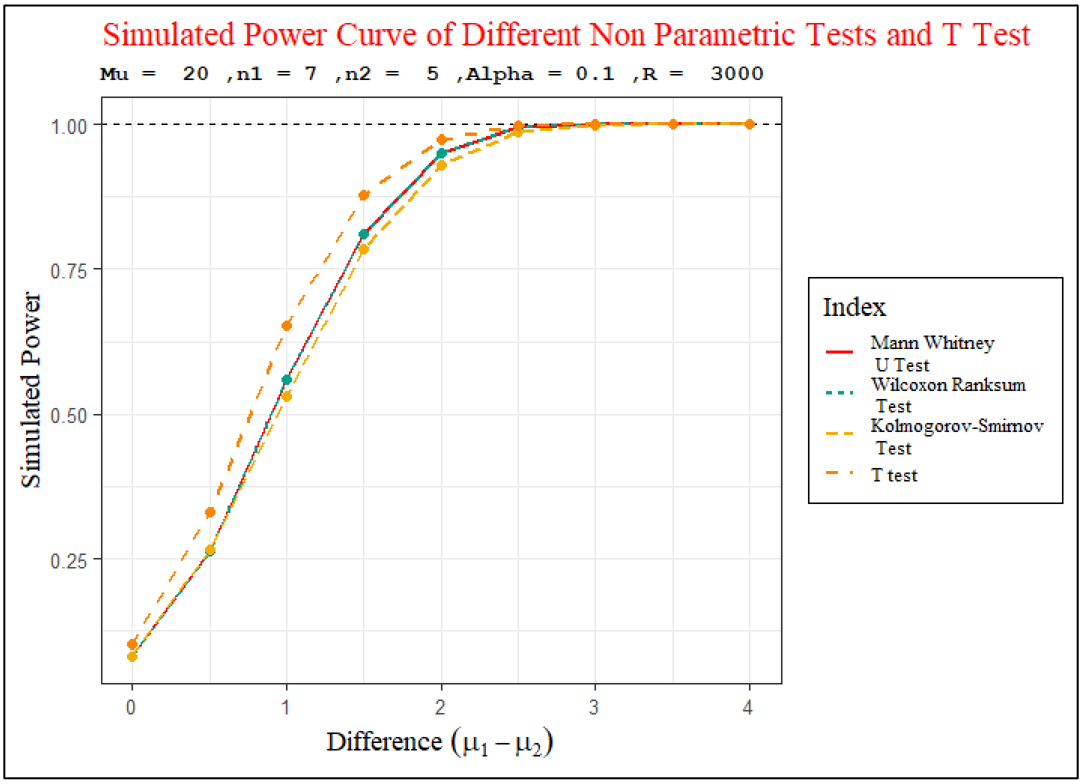


Diagram 38: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 20,\ n_1 = 7, n_2 = 5, R = 3000\ and\ \alpha = 0.10$

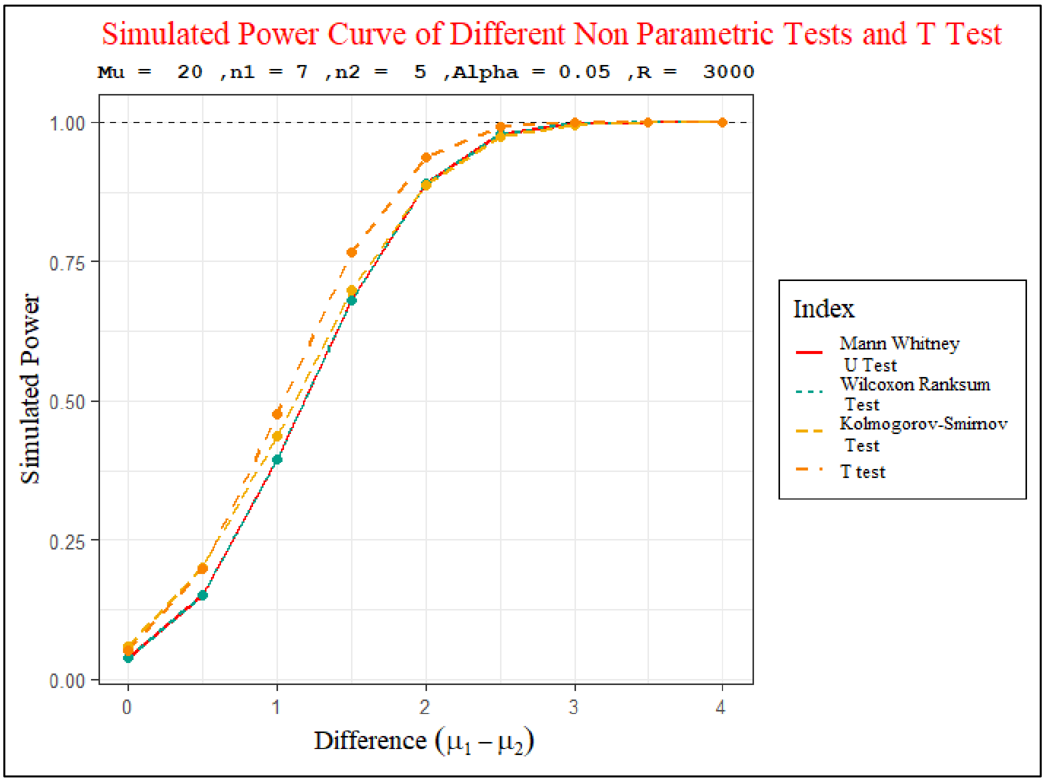


Diagram 39: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 20,\ n_1 = 7, n_2 = 5, R = 3000\ and\ \alpha = 0.05$

Now, we will consider a different pair of sample sizes. We take $n_1 = 12, n_2 = 18$. Here, we take $mu = 10$ and $\alpha = 0.10,0.05$. We have presented the results of the simulation using tables and graphs below.

Table 28: Empirical size and Power of Two sample T test and Different Non-parametric Tests For $n_1 = 12, n_2 = 18, mu = 10, R = 2000$ and $\alpha = 0.10,0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0.0	0.08950	0.09900	0.08600	0.08600	0.04100	0.03800	0.03450	0.03450
0.5	0.51500	0.45500	0.50100	0.50100	0.36250	0.29850	0.34700	0.34700
1.0	0.91150	0.84200	0.89500	0.89500	0.82800	0.70450	0.79750	0.79750
1.5	0.99600	0.98900	0.99400	0.99400	0.98950	0.96350	0.98550	0.98550
2.0	1.00000	0.99900	0.99950	0.99950	0.99950	0.99700	0.99900	0.99900
2.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
3.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
4.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

** For T test and Kolmogorov-Smirnov test only, we have used exact critical value.

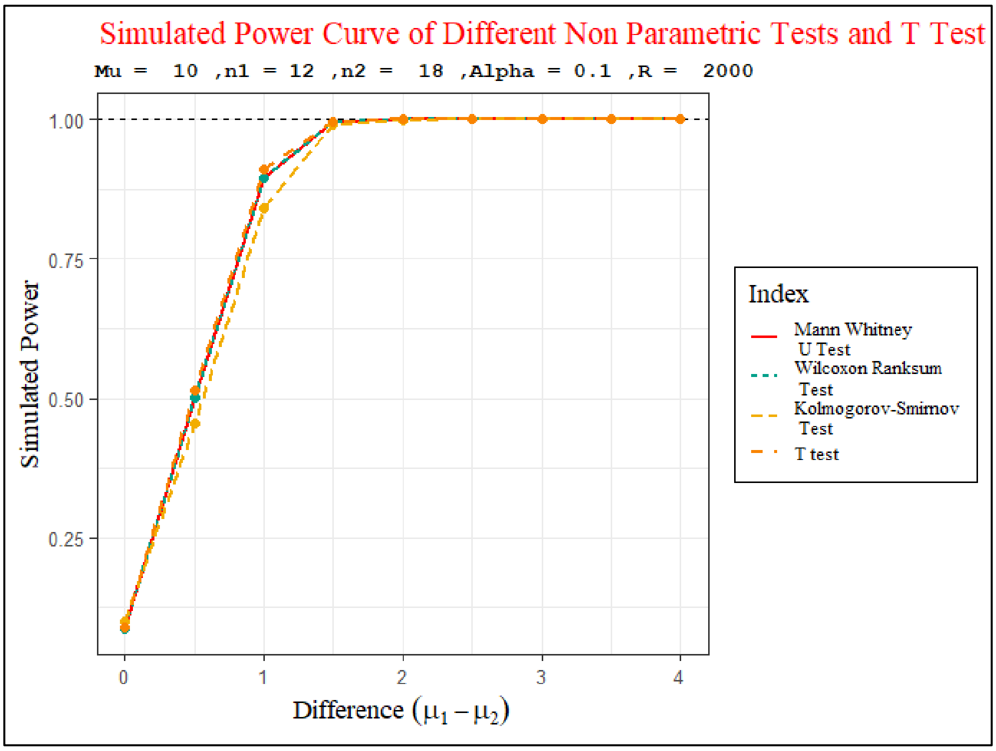


Diagram 40: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 10, n_1 = 12, n_2 = 18, R = 2000 \ and \ \alpha = 0.10$

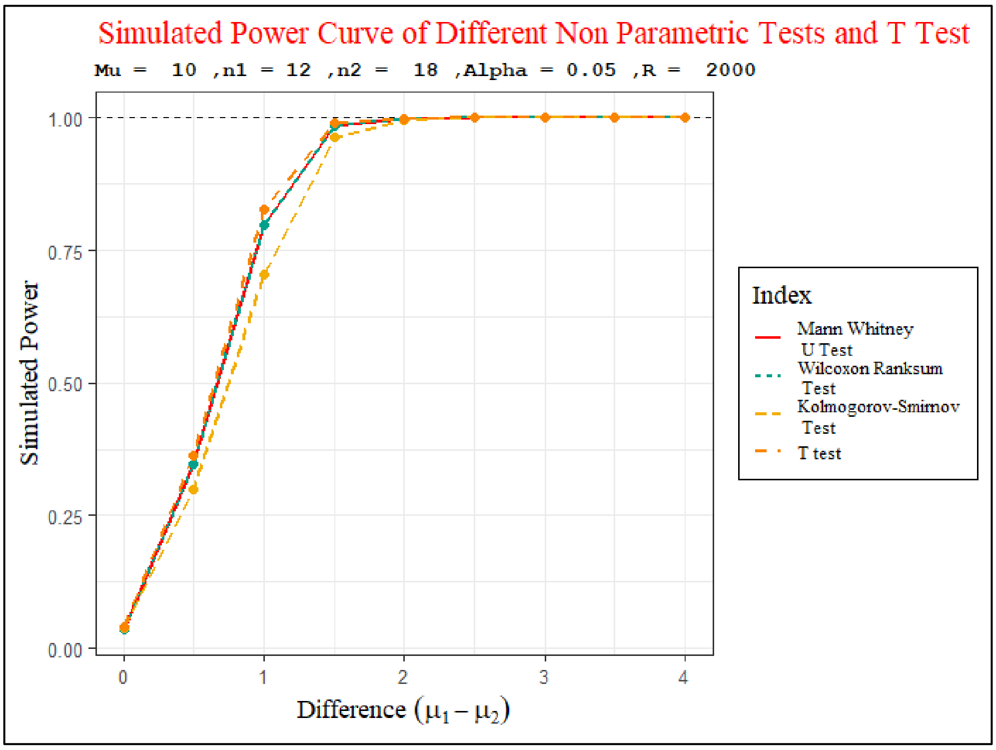


Diagram 41: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 10, n_1 = 12, n_2 = 18, R = 2000 \ and \ \alpha = 0.05$

Here, for both values of α, we see that Kolmogorov-Smirnov test has uniformly less power than all the other tests. Also, here, power of all tests reaches 1 more quickly.

Finally, we consider $n_1 = 22, n_2 = 22$ and $mu = 10, \alpha = 0.10,0.05$. We have presented the empirical size and power values using tables and graphs.

Table 29: Empirical size and Power of Two sample T test and Different Non-parametric Tests For $n_1 = 22, n_2 = 22, mu = 10, R = 2000\ and\ \alpha = 0.10,0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	T Test	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0.0	0.10200	0.06300	0.10250	0.10250	0.05500	0.03000	0.05700	0.05700
0.5	0.63100	0.42550	0.60250	0.60250	0.48550	0.29500	0.46350	0.46350
1.0	0.98000	0.89750	0.97050	0.97050	0.94900	0.79150	0.93800	0.93800
1.5	1.00000	0.99600	1.00000	1.00000	0.99950	0.98800	0.99950	0.99950
2.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
3.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
3.5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

4.0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

** We have used Large Sample Approximation for non-parametric tests.

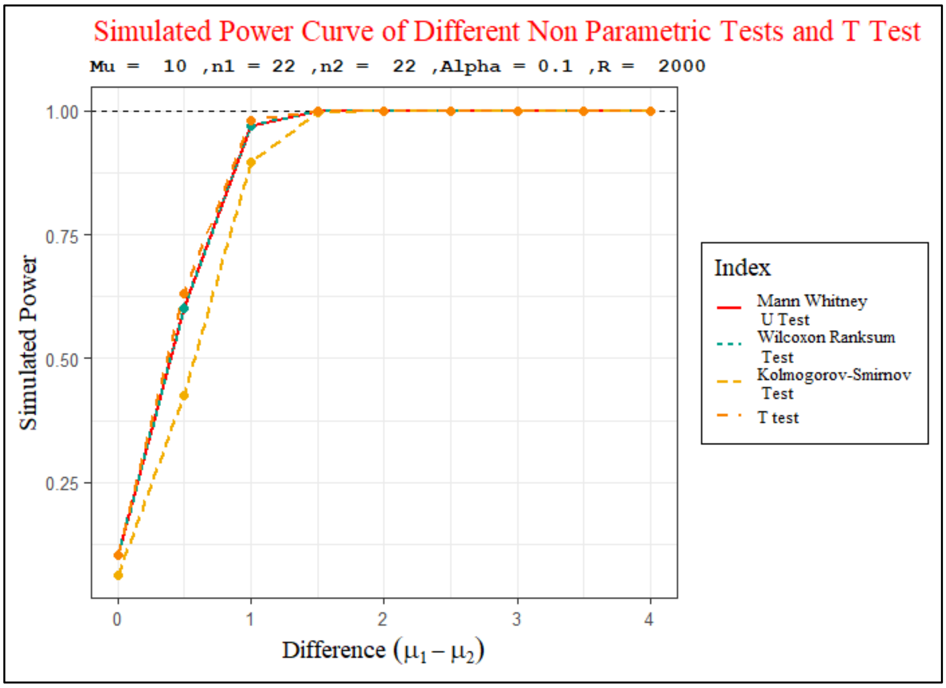


Diagram 42: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 10, n_1 = 22, n_2 = 22, R = 2000 \text{ and } \alpha = 0.10$

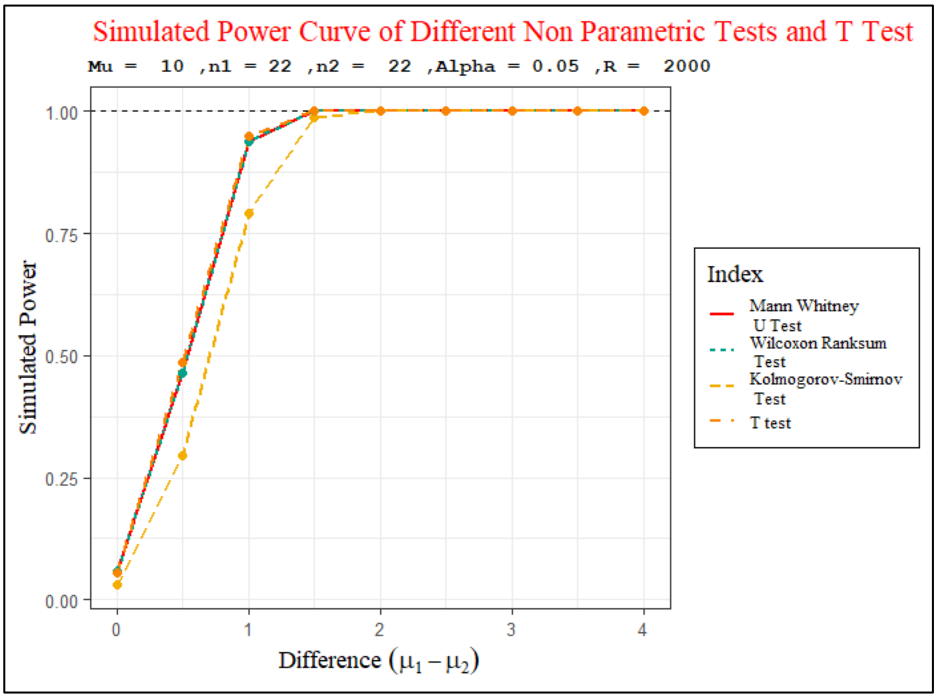


Diagram 43: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 10, n_1 = 22, n_2 = 22, R = 2000 \text{ and } \alpha = 0.05$

We see that, the empirical power of Kolmogorov-Smirnov test is uniformly less than that of other tests. Also, we see that the empirical power curves of T-test, Mann Whitney U test, Wilcoxon Rank Sum test overlap on each other. Which means that they become almost equivalent in terms of power.

9.8.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is normal, two sample T-test (Which is actually LRT) performs better than Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test in terms of power. i.e., the two sample T test is able to detect the same difference between the means of two independent normal population with same variance more frequently than other tests (non-parametric tests, which we have discussed). However, as sample sizes increase the power of Mann Whitney U test, Wilcoxon Rank Sum test increases and become almost same with the power of two sample T test. It is clear from the above graphs also.
2. For small sample sizes, sometimes Kolmogorov-Smirnov test performs better than Mann-Whitney U test and Wilcoxon Rank Sum tests in terms of power. But in most of the cases, MNW test and Wilcoxon Rank Sum test performs better than Kolmogorov-Smirnov test. The power of MNW test and Wilcoxon test reaches 1 more quickly as compared to Kolmogorov-Smirnov test. Actually, here we are testing Location type alternatives for normal population, so, it is quite natural to see Mann Whitney U test and Wilcoxon Rank Sum tests to perform well than Kolmogorov Smirnov test.
3. The power of all the tests doesn't depend upon the value of mu (Y population mean). Other parameters remain fixed, if mu increases or decreases, the value of power does not change.
4. All the tests are consistent test. Because, as sample sizes increase, the power of all the tests tends to 1. For example, if we take $n_1 = 30, n_2 = 30$ and $mu = 10$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

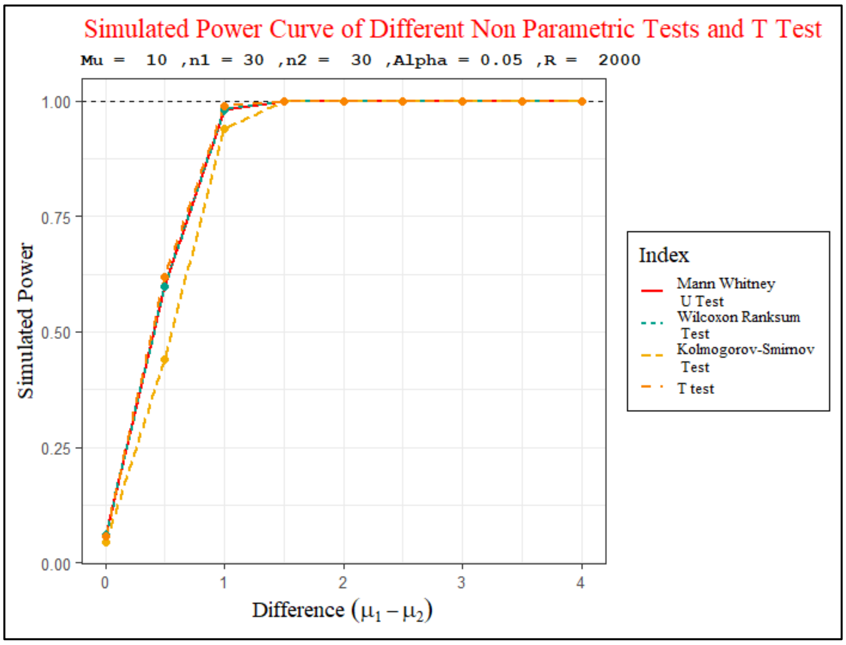


Diagram 44: Simulated Power Curve of Different Non-parametric tests & T test for $mu = 10,\ n_1 = 30, n_2 = 30, R = 2000\ and\ \alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except some differences. (Here, we have used large sample approximation for non-parametric tests, as tables of critical values are not available for those large values of n_1 and n_2.)

9.9. SIMULATION 08:

Simultaneous Comparison of Empirical Size and Power of Two sample Exponential LRT and Different Non-Parametric Tests When Sampled Population is Exponential

9.9.1. Motivation:

Two sample Exponential LRT is very well-known test for two sample problem, when sampled population is exponential. On the other hand, non-parametric tests are also used for two sample problems. Here, we want to make comparative study between the two sample Exponential LRT and different non-parametric tests for two sample problem related to equality of means of two independent populations, when sampled populations are exponential. We have already made a comparative study of Two sample Exponential LRT with Kolmogorov- Smirnov Test, Mann Whitney U test and Wilcoxon Rank Sum test separately. So, here we will compare all of them simultaneously.

9.9.2. Objective:

We will study how Kolmogorov Smirnov test, Mann Whitney U test, Wilcoxon Rank Sum test will behave in respect to power, when sampled population is exponential. So, we will make a comparative study of empirical size, empirical power of all of these tests. We will compare them for varying sample sizes, different level of significances, different values of $\mu_1 - \mu_2$.

9.9.3. Algorithm:

We have used a user-defined function. The name of the function is $Power\_comparisonALL.2$. The function takes different arguments as input and on the basis of the arguments taken as input, it calculates empirical size and power using the steps discussed in the previous section. The different arguments of the function are –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

$lamda$: Mean of Y population.

d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $lamda + d$. Where, $lamda$ is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.

$alpha$: It is level of significance of the test.

$exact.crit$: It is vector of critical values of different non-parametric tests. By default, it is set at $c(F, F, F)$. If we change any of the element of the vector to exact critical value (For example, $c(F, 21, F)$), then the function calculates empirical size and power using that exact critical value. We need to give exact critical points of different tests (For the given sample size and level of significance) as input. Otherwise, it will calculate empirical size and power using large sample approximation of the test-statistic. The first element of the vector is critical value for Mann Whitney U test, second element is for Wilcoxon Rank Sum test and last element is for Kolmogorov Smirnov test.

From the function we get empirical size and power as output. Using that output we can draw the empirical power curve. We can separately do this using the user-defined function - $Visualize\_Power\_comparisonALL.2$. Which takes similar arguments as input. We will provide the codes for these functions at the end of the article.

9.9.4. Discussion:

For the discussion we will consider $d = 0, 1, 2, 3, \ldots\ldots, 30$.

For different combinations of sample size and level of significance (In our study we have only considered 0.05 and 0.10) we will calculate the empirical size and power and will present them using tables and graphs. In most of the cases, we have used exact critical values from the tables of Kolmogorov-Smirnov test, Mann Whitney U test, Wilcoxon Rank Sum test. But for large sample sizes we have used large sample approximations.

For $n_1 = 5, n_2 = 4$, we will construct the table of empirical size and power of Two Sample Exponential LRT, Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test side by side for $lamda = 1$ and $alpha = 0.10, 0.05$.

Table 30: Empirical size and Power of Two sample Exponential LRT and Different Non-parametric Tests For $n_1 = 5, n_2 = 4, lamda = 1, R = 3000\ and\ \alpha = 0.10, 0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0	0.10200	0.08233	0.10367	0.10367	0.05633	0.04333	0.03733	0.03733
1	0.38567	0.23967	0.32667	0.32667	0.24900	0.16600	0.14867	0.14867
2	0.62500	0.40033	0.52200	0.52200	0.47000	0.32367	0.26833	0.26833
3	0.76700	0.50633	0.62233	0.62233	0.62100	0.42267	0.36700	0.36700
4	0.84500	0.57800	0.70267	0.70267	0.72067	0.48500	0.44000	0.44000
5	0.90933	0.66367	0.77733	0.77733	0.81167	0.57467	0.51867	0.51867
6	0.93700	0.71733	0.81367	0.81367	0.86967	0.64367	0.57533	0.57533

7	0.94733	0.76300	0.85267	0.85267	0.89733	0.68933	0.62200	0.62200
8	0.96867	0.78667	0.86767	0.86767	0.93067	0.73000	0.64467	0.64467
9	0.97733	0.81567	0.88867	0.88867	0.94633	0.74600	0.69067	0.69067
10	0.98533	0.86033	0.91433	0.91433	0.96467	0.80833	0.73400	0.73400
…..	……	…..	……	…..	…..	……	…..	……
20	0.99833	0.94100	0.96933	0.96933	0.99367	0.91833	0.84567	0.84567
21	0.99833	0.94933	0.97167	0.97167	0.99567	0.92833	0.85533	0.85533
22	0.99900	0.95233	0.97700	0.97700	0.99667	0.93200	0.86533	0.86533
…..	……	…..	……	…..	…..	……	…..	……
28	1.00000	0.96800	0.98633	0.98633	0.99833	0.95200	0.88833	0.88833
29	1.00000	0.97133	0.98733	0.98733	0.99967	0.95867	0.89867	0.89867
30	0.99933	0.97133	0.98733	0.98733	0.99867	0.95700	0.90067	0.90067

From the above table, we can see that as value of d increases empirical power increases for all the tests. For a fixed level of significance, we have already seen that Two Sample LRT is uniformly more powerful than Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test. Here also we are observing the same thing. Additionally, we see that among the non-parametric tests (which we are discussing) on an average Kolmogorov-Smirnov test has less power than other two non-parametric tests. For $\alpha = 0.10$, Kolmogorov Smirnov test has uniformly less power among all the tests discussed. But, for $\alpha = 0.05$, Kolmogorov Smirnov test has uniformly more empirical power than Mann Whitney U test and Wilcoxon Rank sum test. To check whether this happened due to sampling fluctuation, we have increased the number of replications (R), to 5000. We see that, still Kolmogorov-Smirnov test is uniformly more powerful than the other two discussed non-parametric tests. So, we expect that this may be true actually.

We now present the empirical power curve of the above simulation below.

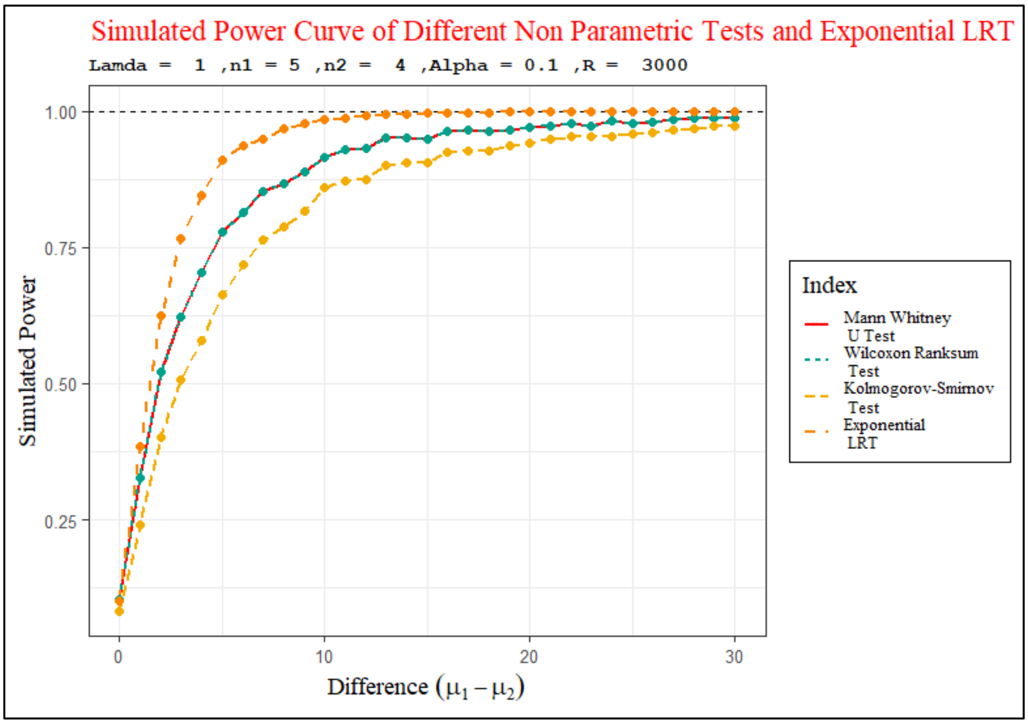


Diagram 45: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 1,\ n_1 = 5, n_2 = 4, R = 3000\ and\ \alpha = 0.10$

From the above graph we can see that empirical power of Kolmogorov-Smirnov test is uniformly less than empirical power of Mann Whitney U test, Wilcoxon Rank Sum test and Two sample Exponential LRT for $\alpha = 0.10$. Also, the empirical powers of Mann Whitney U test and Wilcoxon Rank Sum test are exactly same. We have already discussed about this equivalence between the two tests and also justified using simulation.

In the second graph we see that empirical power of Kolmogorov Smirnov test is uniformly greater than empirical power of Wilcoxon Rank Sum test and Mann Whitney u test.

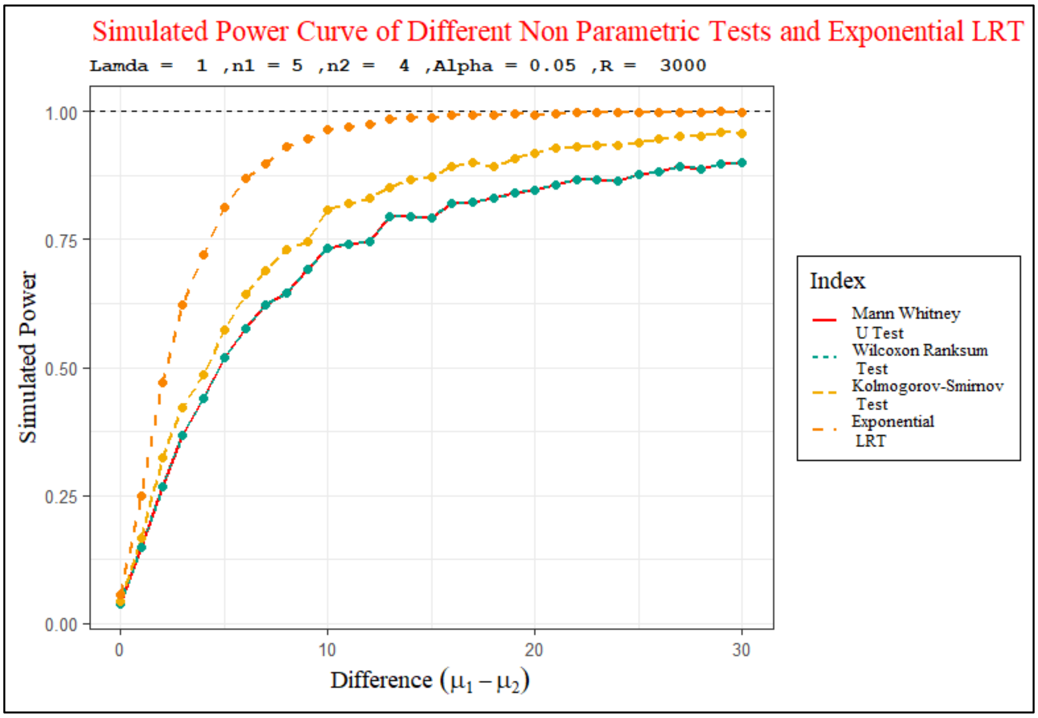


Diagram 46: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 1,\ n_1 = 5, n_2 = 4, R = 3000\ and\ \alpha = 0.05$

As it is expected that, for $\alpha = 0.10$, the empirical power of all the tests reaches 1 more quickly than $\alpha = 0.05$.

Now, we will consider a different pair of sample size. We take $n_1 = 6, n_2 = 10$ and $lamda = 1$. We present the results using table below.

From the table we see that for $\alpha = 0.10,0.05$. Kolmogorov-Smirnov Test has on an average uniformly less empirical power than other tests. But in the last simulation, we have seen different result for $\alpha = 0.05$. So, to check whether this happened due to sampling fluctuation, we increase the number of replications to $R = 5000$ and still we observe that Kolmogorov Smirnov test has uniformly less empirical power for this pair of sample sizes. So, we expect that actually for this choice of sample sizes, Kolmogorov-Smirnov test has less uniformly power than the other tests discussed.

Table 31: Empirical size and Power of Two sample Exponential LRT and Different Non-parametric Tests For $n_1 = 6, n_2 = 10, lamda = 1, R = 3000$ and $\alpha = 0.10, 0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0	0.10333	0.09167	0.09067	0.09067	0.05433	0.04700	0.04733	0.04733
1	0.52300	0.36833	0.42467	0.42467	0.37300	0.24233	0.28767	0.28767
2	0.78800	0.57600	0.63367	0.63367	0.68000	0.43167	0.49567	0.49567
3	0.91500	0.72767	0.79033	0.79033	0.84233	0.59533	0.66567	0.66567
4	0.95400	0.81067	0.85300	0.85300	0.91800	0.70300	0.74833	0.74833
5	0.97800	0.86933	0.90600	0.90600	0.95700	0.78433	0.83367	0.83367
6	0.98800	0.89533	0.93033	0.93033	0.97267	0.83133	0.87267	0.87267
7	0.99533	0.93000	0.95633	0.95633	0.98733	0.87100	0.89967	0.89967
8	0.99433	0.93900	0.95533	0.95533	0.99033	0.88767	0.90567	0.90567

9	0.99733	0.95167	0.96500	0.96500	0.99500	0.90833	0.92700	0.92700
10	0.99800	0.96767	0.97500	0.97500	0.99633	0.93333	0.95000	0.95000
…..	……	…..	……	…..	…..	……	…..	……
20	1.00000	0.99400	0.99700	0.99700	1.00000	0.98567	0.98800	0.98800
21	1.00000	0.99200	0.99500	0.99500	1.00000	0.98667	0.98700	0.98700
22	1.00000	0.99367	0.99633	0.99633	1.00000	0.98800	0.98967	0.98967
…..	……	…..	……	…..	…..	……	…..	……
28	1.00000	0.99567	0.99667	0.99667	1.00000	0.99300	0.99133	0.99133
29	1.00000	0.99567	0.99667	0.99667	1.00000	0.99200	0.99033	0.99033
30	1.00000	0.99800	0.99767	0.99767	1.00000	0.99367	0.99200	0.99200

We have presented the empirical power curves of the above simulations below.

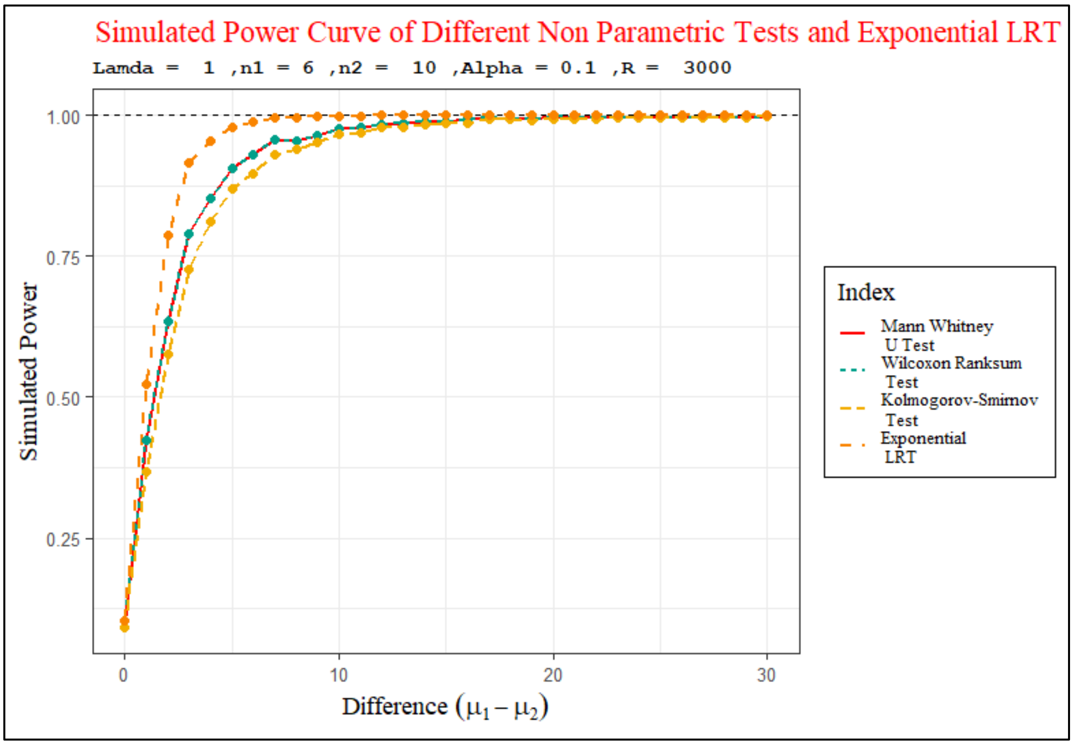


Diagram 47: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 1,\ n_1 = 6, n_2 = 10, R = 3000\ and\ \alpha = 0.10$

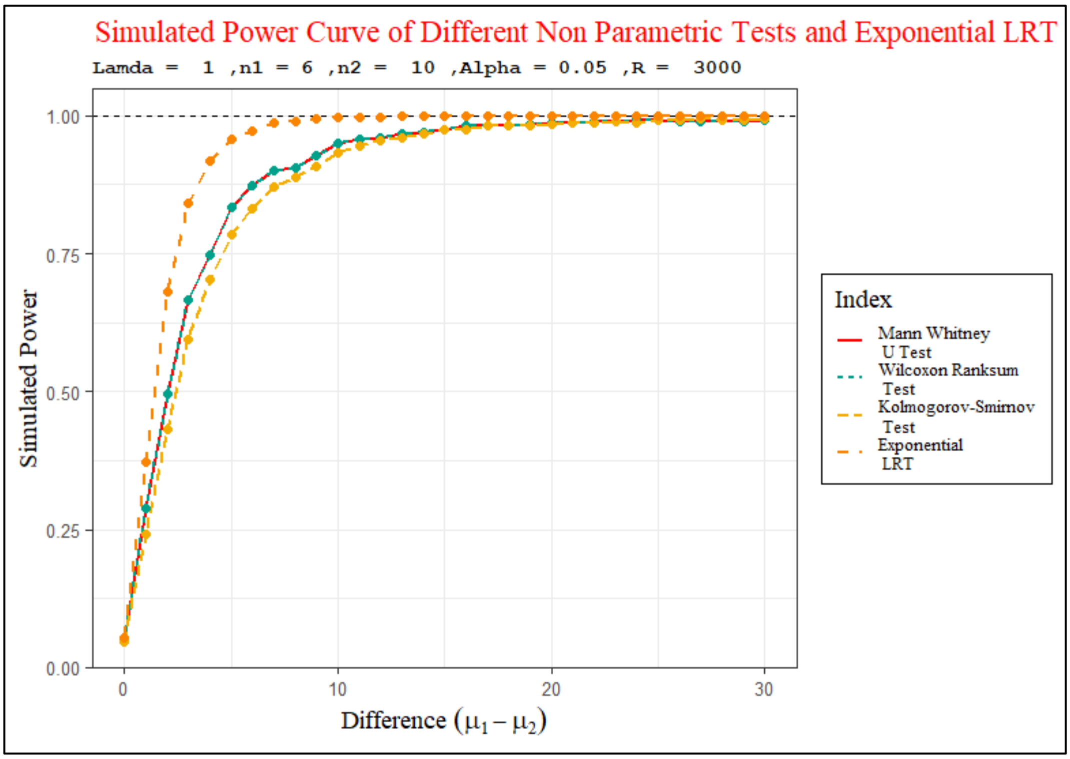


Diagram 48: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 1,\ n_1 = 6, n_2 = 10, R = 3000\ and\ \alpha = 0.05$

Now, we will consider a different pair of sample sizes. We take $n_1 = 16, n_2 = 20$. Here, we take $lamda = 2$ and $\alpha = 0.10, 0.05$. We have presented the results of the simulation using tables and graphs below.

Table 32: Empirical size and Power of Two sample T test and Different Non-parametric Tests For $n_1 = 16, n_2 = 20, lamda = 2, R = 2000\ and\ \alpha = 0.10, 0.05$

Difference (d)	Level of Significance(α)							
	0.10				0.05			
	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test	Exponential LRT	Kolmogorov Smirnov Test	Mann Whitney U Test	Wilcoxon Rank Sum Test
0	0.11450	0.08850	0.11200	0.11200	0.06350	0.04700	0.05300	0.05300
1	0.45850	0.34350	0.40050	0.40050	0.31500	0.20950	0.26450	0.26450
2	0.78500	0.59700	0.66700	0.66700	0.65450	0.44250	0.52600	0.52600
3	0.92650	0.78950	0.84000	0.84000	0.86400	0.66000	0.72300	0.72300
4	0.97000	0.88800	0.92750	0.92750	0.93800	0.79300	0.85550	0.85550
5	0.99350	0.93700	0.95850	0.95850	0.98400	0.87550	0.91200	0.91200
6	0.99500	0.96150	0.97550	0.97550	0.98650	0.92050	0.94500	0.94500
7	0.99850	0.98200	0.99250	0.99250	0.99700	0.96300	0.97250	0.97250

8	0.99950	0.98850	0.99350	0.99350	0.99700	0.97050	0.98400	0.98400
9	1.00000	0.99100	0.99250	0.99250	0.99850	0.97900	0.98350	0.98350
10	1.00000	0.99700	0.99900	0.99900	1.00000	0.99100	0.99600	0.99600
…..	……	…..	……	…..	…..	……	…..	……
20	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
21	1.00000	0.99950	1.00000	1.00000	1.00000	0.99950	1.00000	1.00000
…..	……	…..	……	…..	…..	……	…..	……
29	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
30	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

** For T test and Kolmogorov-Smirnov test only, we have used exact critical value.

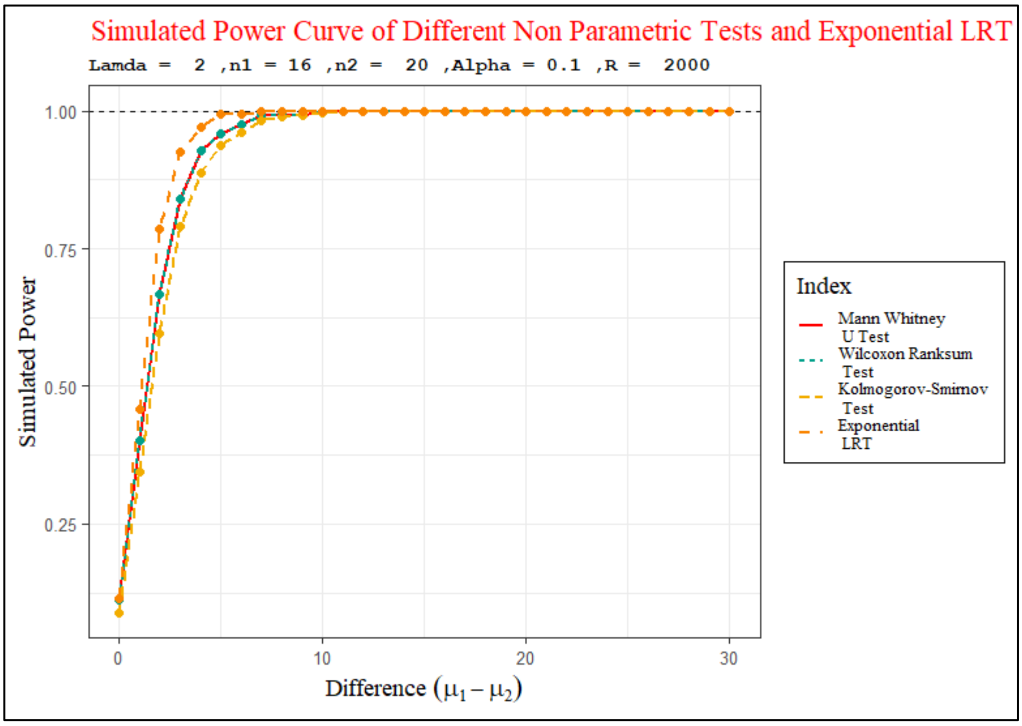


Diagram 49: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 2, n_1 = 16, n_2 = 20, R = 2000 \; and \; \alpha = 0.10$

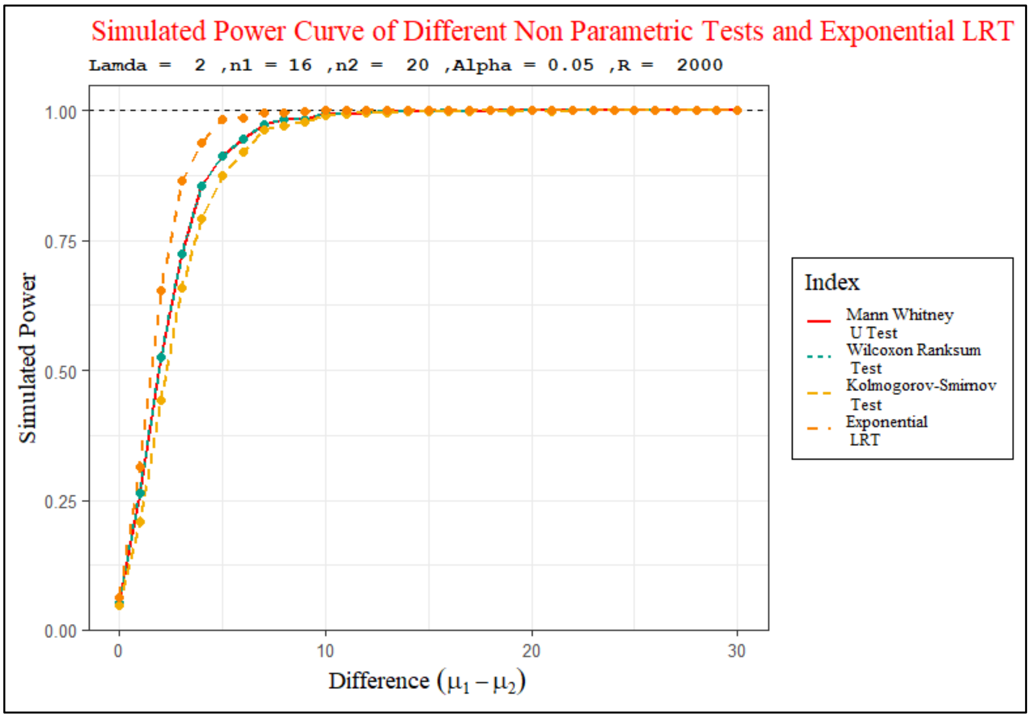


Diagram 50: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 2, n_1 = 16, n_2 = 20, R = 2000 \; and \; \alpha = 0.05$

Here, for both values of α, we see that Kolmogorov-Smirnov test has uniformly less power than all the other tests. Also, here, power of all tests reaches 1 more quickly.

Finally, we consider $n_1 = 24, n_2 = 24$ and $lamda = 2, \alpha = 0.10, 0.05$. We have presented the empirical size and power values using graphs (We have used Large Sample Approximation for non-parametric tests.).

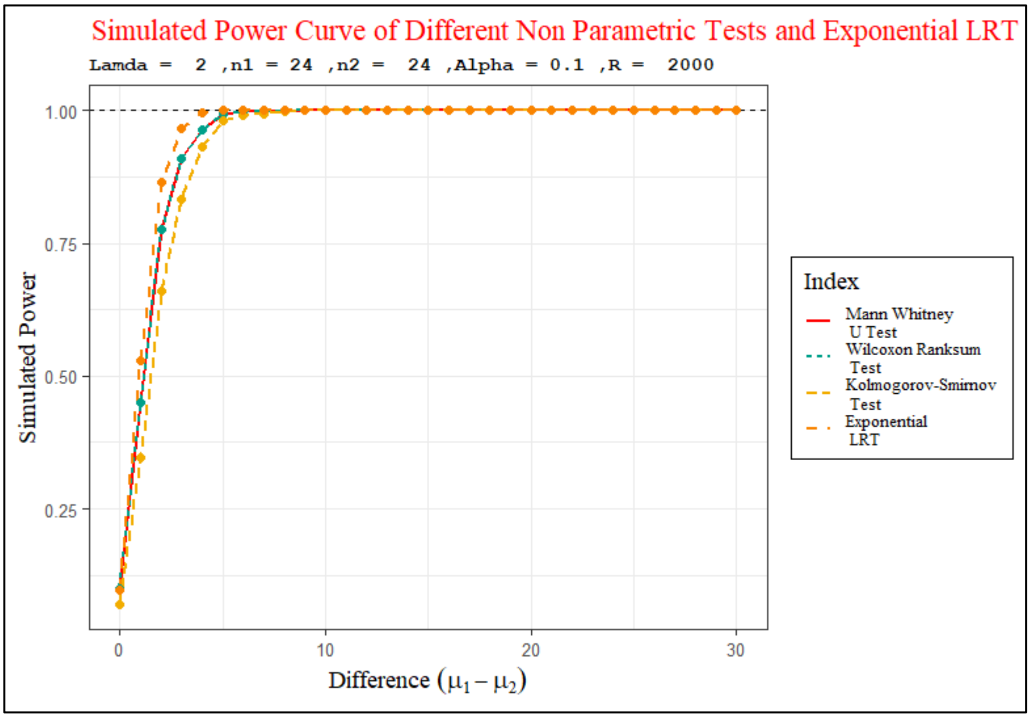


Diagram 51: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 2,\ n_1 = 24, n_2 = 24, R = 2000\ and\ \alpha = 0.10$

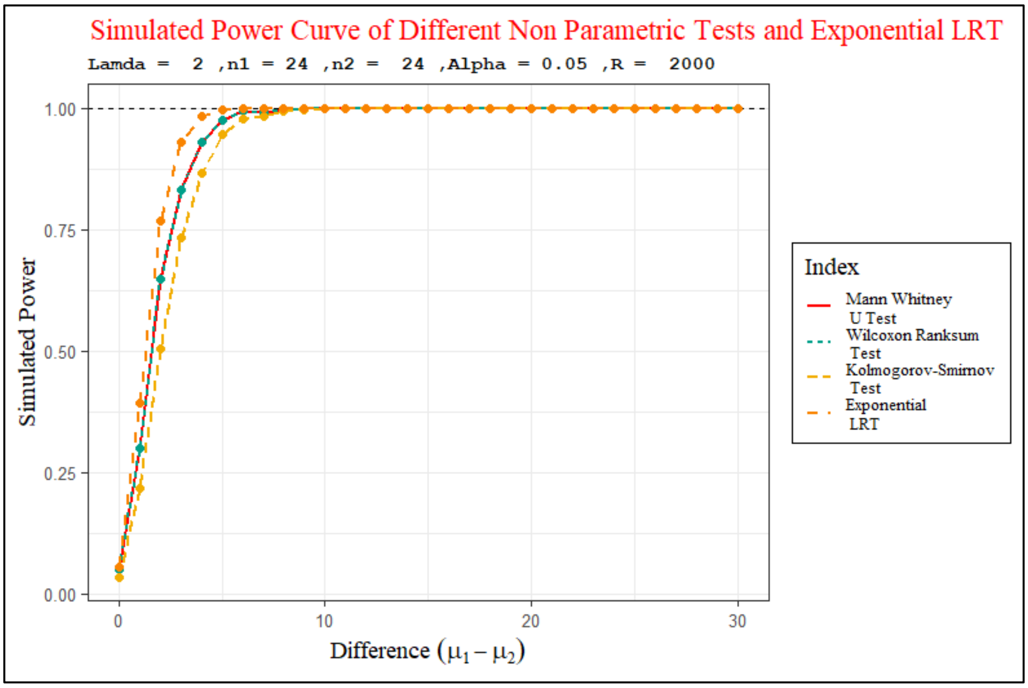


Diagram 52: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 2,\ n_1 = 24, n_2 = 24, R = 2000\ and\ \alpha = 0.05$

We see that, the empirical power of Kolmogorov-Smirnov test is uniformly less than that of other tests. Also, we see that the empirical power curves of Mann Whitney U test, Wilcoxon Rank Sum test does not overlap on empirical power curve of Two sample exponential LRT, which was in case of T test (LRT for normal). Which means that they become almost equivalent in terms of power.

9.9.5. Observations:

Now, we will summarize our observations from the above discussion below –

1. From the above discussion we have observed that when the sampled population is exponential, two-sample exponential LRT performs better than Kolmogorov-Smirnov test, Mann Whitney U test and Wilcoxon Rank Sum test in terms of power. i.e., the two-sample exponential LRT is able to detect the same difference between the means of two independent normal population with same variance more frequently than other tests (non-parametric tests, which we have discussed). However, as sample sizes increase the power of Mann Whitney U test, Wilcoxon Rank Sum test increases and become almost same with the power of two-sample exponential LRT. It is clear from the above graphs also.
2. For small sample sizes, sometimes Kolmogorov-Smirnov test performs better than Mann-Whitney U test and Wilcoxon Rank Sum tests in terms of power. But in most of the cases, MNW test and Wilcoxon Rank Sum test performs better than Kolmogorov-Smirnov test. The power of MNW test and Wilcoxon test reaches 1 more quickly as compared to Kolmogorov-Smirnov test.
3. The power of all the tests doesn't depend upon the value of $lamda$ (Y population mean). Other parameters remain fixed, if $lamda$ increases or decreases, the value of power decreases or increases.
4. All the tests are consistent test. Because, as sample sizes increase, the power of all the tests tends to 1. For example, if we take $n_1 = 40, n_2 = 40$ and $lamda = 2$, Alpha = 0.05, then the empirical power curves of the two tests are given by –

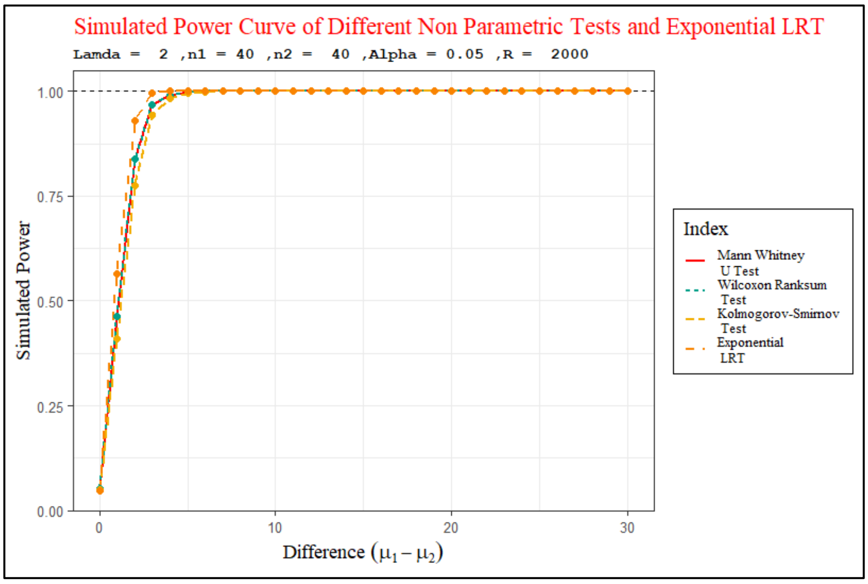


Diagram 53: Simulated Power Curve of Different Non-parametric tests & Two sample Exponential LRT for $lamda = 2,\ n_1 = 40, n_2 = 40, R = 2000\ and\ \alpha = 0.05$

From the above graph we can see that, the simulated powers are 1, except some differences. (Here, we have used large sample approximation for non-parametric tests, as tables of critical values are not available for those large values of n_1 and n_2.)

9.10. SIMULATION 09:

Study of Probability Distributions of Test Statistics of Different Non-parametric Tests using Simulation

9.10.1. Motivation:

In the previous simulations, we have made a comparative study between different parametric tests (LRT) with different non-parametric tests, when sampled population is normal and exponential separately. We have observed that, for same difference d between the two independent population means, same non-parametric tests perform well, when sampled population is normal as compared to when sampled population is exponential. This provokes us to find out the reason. One may say that, how can we compare two different situations certainly. Actually, we are not comparing two different situations. we just want to know why for the same test power reaches 1 when $d = 30$ (Sampled population is exponential) and $d = 4$ (Sample population is normal).

9.10.2. Objective:

Here, we will try to find out a reason of the above query. To understand the reason, we need basic understanding of power of a test. Which we have already discussed. Basically, power of a test is not just calculating a probability. But it is very deeply related to probability distribution of the test-statistic. We will study simulated distributions of test-statistics of different non-parametric tests when sampled population is normal and exponential (separately) for different values of $d = \mu_1 - \mu_2$.

9.10.3. Algorithm:

For this purpose, we have used three user defined functions, all having same arguments. The names of the functions are - $Simulated\_KolmogorovSmirnov$, $Simulated\_MannWhiteny$, $Simulated\_WilcoxonRanksum$. The arguments of given below –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

mu: Mean of Y population. (For normal as well as exponential)

d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $lamda + d$. Where, $lamda$ is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.

These functions take different arguments as input and gives us graph as output. Where, we have $length(d)$ number of panels and in each panel, for an element of the vector $d,$ we plot the histogram of simulated distributions of test-statistic when sampled population is normal as well as exponential. (How to find simulate distribution of test statistic is already discussed under the section of introduction to simulation studies). Note that, to make things comparable, we have fixed the mean of Y population same for both normal and exponential distribution. Codes of these functions are given later.

9.10.4. Discussion:

For this discussion we will take $d = 0{,}0.5{,}1{,}..7.5$.

We will plot the histogram of simulated distribution of test statistic of Kolmogorov-Smirnov test for different sample sizes.

For $n_1 = 5, n_2 = 4$ and $mu = 1, R = 3000$. We get the simulated distributions as below-

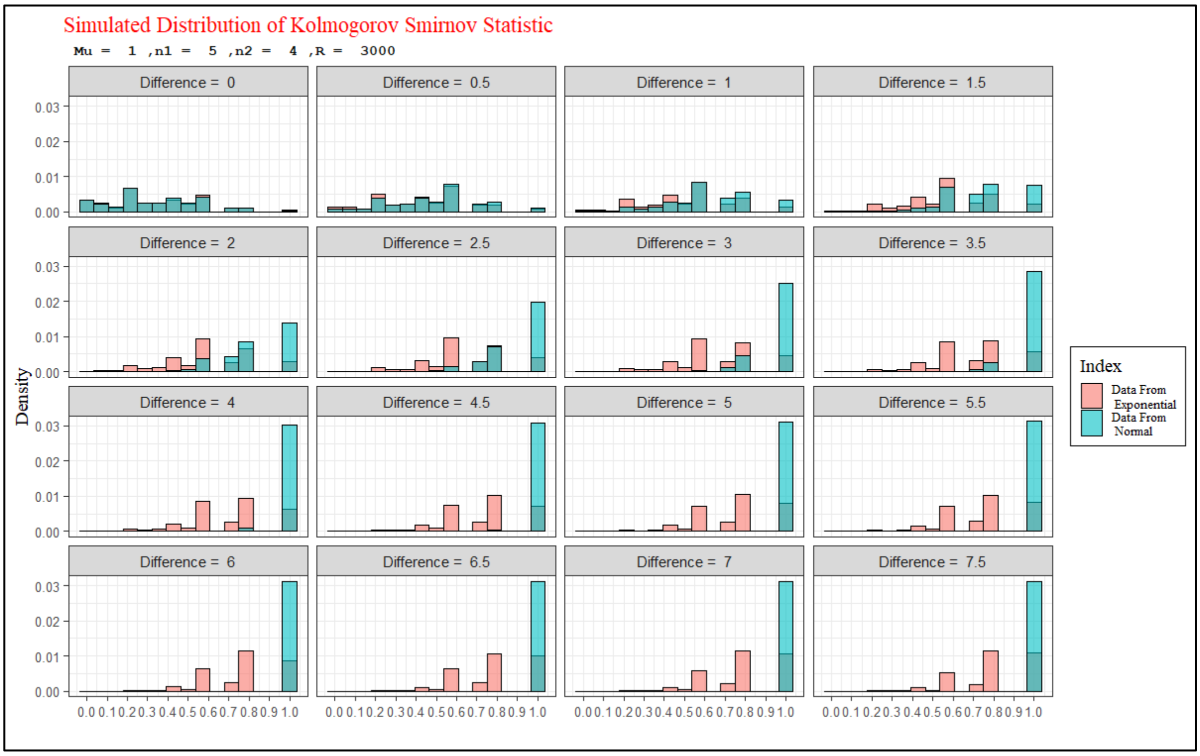


Diagram 54: Simulated Distribution of Kolmogorov-Smirnov Statistic for $mu = 1, n_1 = 5, n_2 = 4, R = 3000$

From the above graph we see that, for $d = 0$, the distribution of the test-statistic overlaps for both cases. That is when data is from normal and data is from exponential. But as value of d increases, we see that the distribution of the test-statistic shifts towards extreme right for normal distribution very quickly. That is, graphically, the power of Kolmogorov-Smirnov test becomes 1 very quickly for normal distribution. While, for exponential distribution, we have still some middle range values of test-statistic. That is why, power of Kolmogorov-Smirnov test for normal distribution reaches very quickly to 1.

Now, we consider for $n_1 = 5, n_2 = 4$ and $mu = 2, R = 3000$. We get the simulated distributions as below-

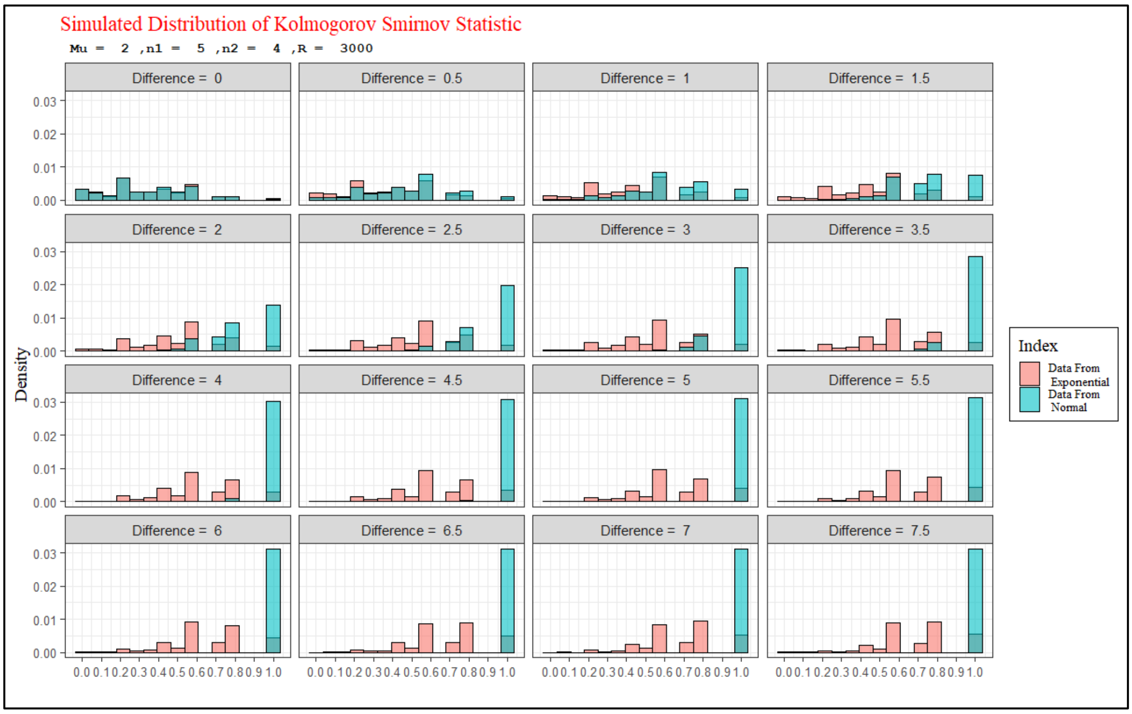


Diagram 55: Simulated Distribution of Kolmogorov-Smirnov Statistic for $mu = 2,\ n_1 = 5, n_2 = 4, R = 3000$

Here, we have increased the value of mu. We have empirically verified that this has no affect on the power of Kolmogorov-Smirnov test, when sampled population is normal but it has indeed effect on power of Kolmogorov-Smirnov test, when sampled population is exponential. By previous simulations studies, we can say that power of test must decreases here for sample population exponential. Now, see in the diagram. As it is expected that, the increases in the value of mu is not affecting the distribution of the test-statistic as well (for sampled population normal). But, the distribution of test statistic is affected, when sampled population is exponential. We see a greater number of observations are in the middle range than previous.

Now, similarly we can draw the simulated distribution for $n_1 = 10, n_2 = 10$ and $mu = 2, R = 3000$.

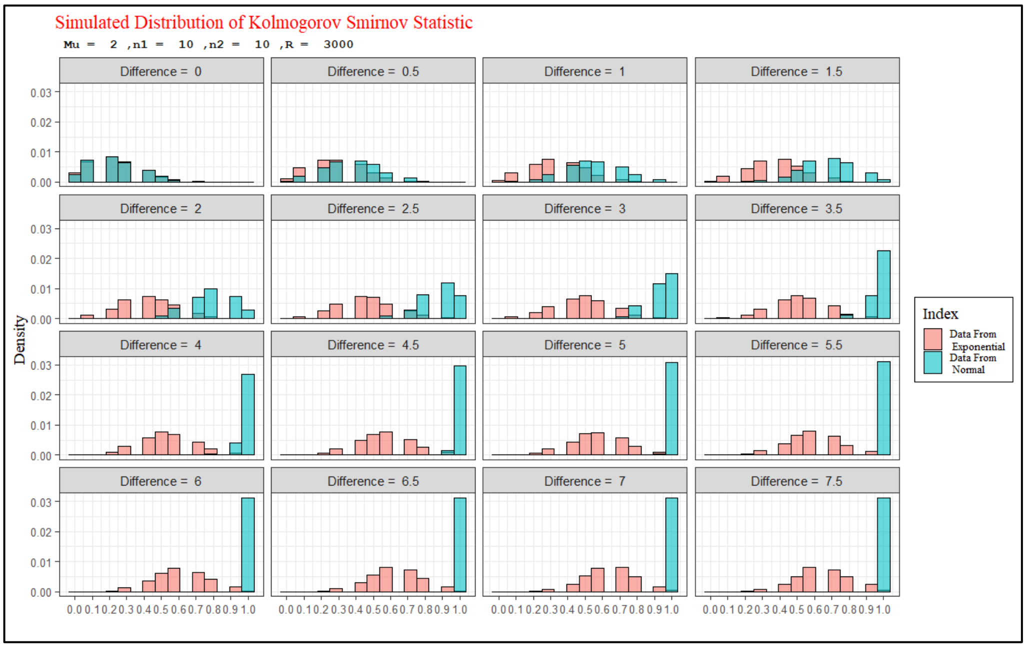


Diagram 56: Simulated Distribution of Kolmogorov-Smirnov Statistic for $mu = 2, n_1 = 10, n_2 = 10, R = 3000$

From the above graph, we observe that distribution of the test-statistic shifts toward extreme right more quickly than previous. Which in terms of power means that power of the reaches more quickly to 1.

Now, we will draw histogram of the simulated distributions for Mann-Whitney U statistic, for various choices of n_1, n_2, mu. We have presented them below –

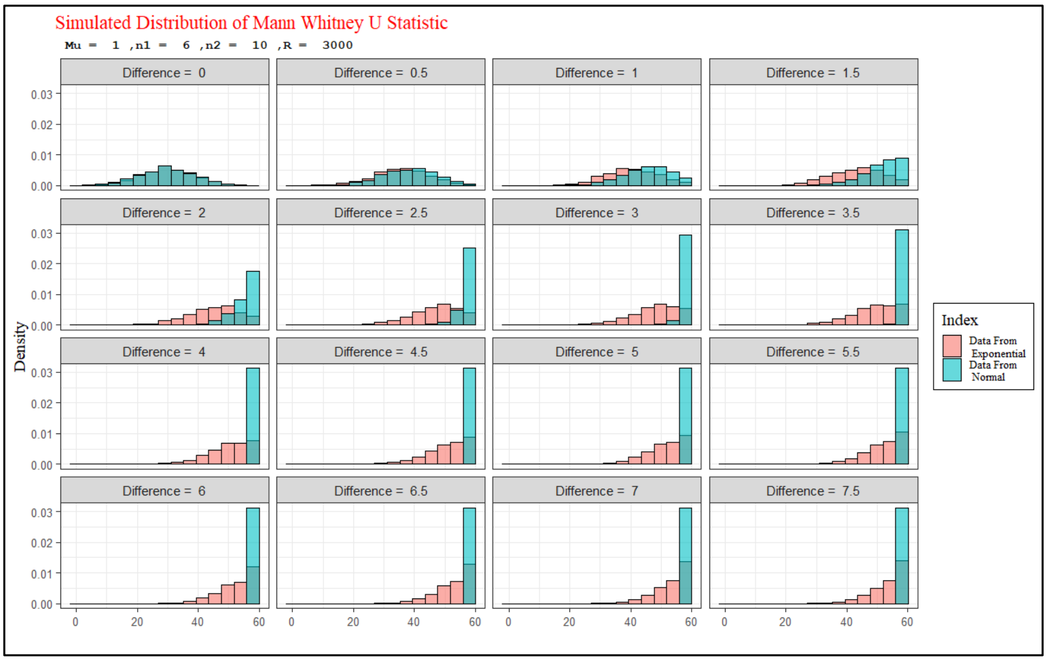


Diagram 57: Simulated Distribution of Mann Whitney U Statistic for $mu = 1, n_1 = 6, n_2 = 10, R = 3000$

From the above diagram, we see that the NULL distribution of Mann Whitney U statistic is symmetric about its mean. But as value of d increases, it deviates from the symmetric nature of the distribution. As value of d increases, we see that the distribution of the test-statistic shifts towards extreme right for normal distribution very quickly. That is, graphically, the power of Mann Whitney U test becomes 1 very quickly for normal distribution. While, for exponential distribution, we have still some middle range values of test-statistic. That is why, power of Mann Whitney U test for normal distribution reaches very quickly to 1.

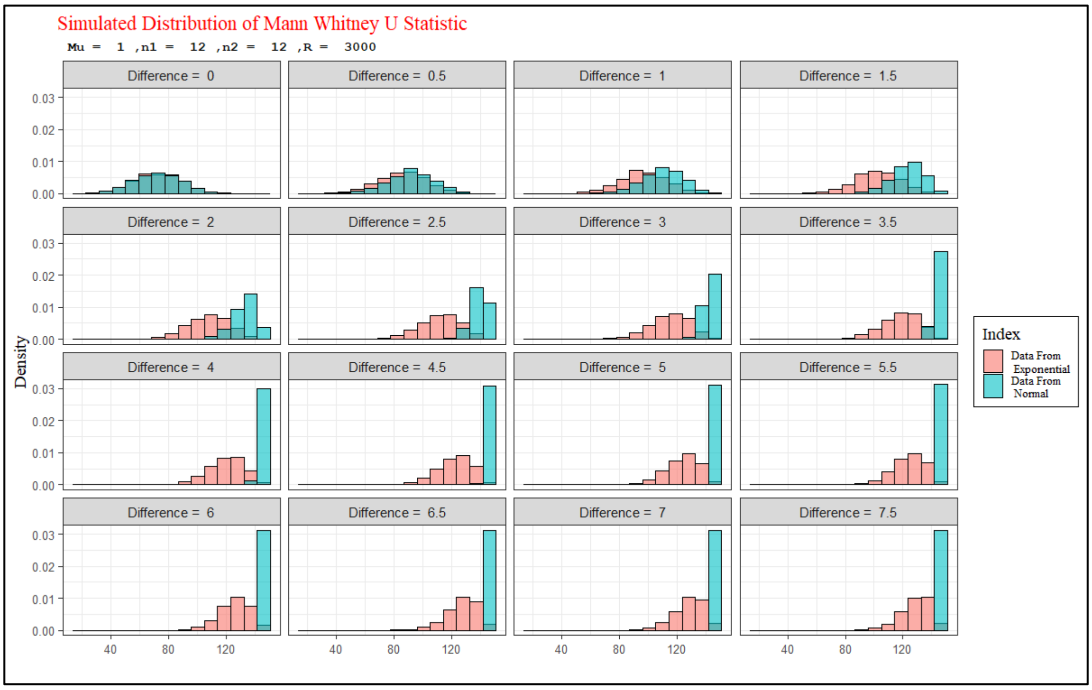


Diagram 58: Simulated Distribution of Mann Whitney U Statistic for $mu = 1,\ n_1 = 12, n_2 = 12, R = 3000$

From the above graph, we observe that distribution of the test-statistic shifts toward extreme right more quickly than previous. Which in terms of power means that power of the reaches more quickly to 1.

Similarly for Wilcoxon Rank Sum we get same type of observations.

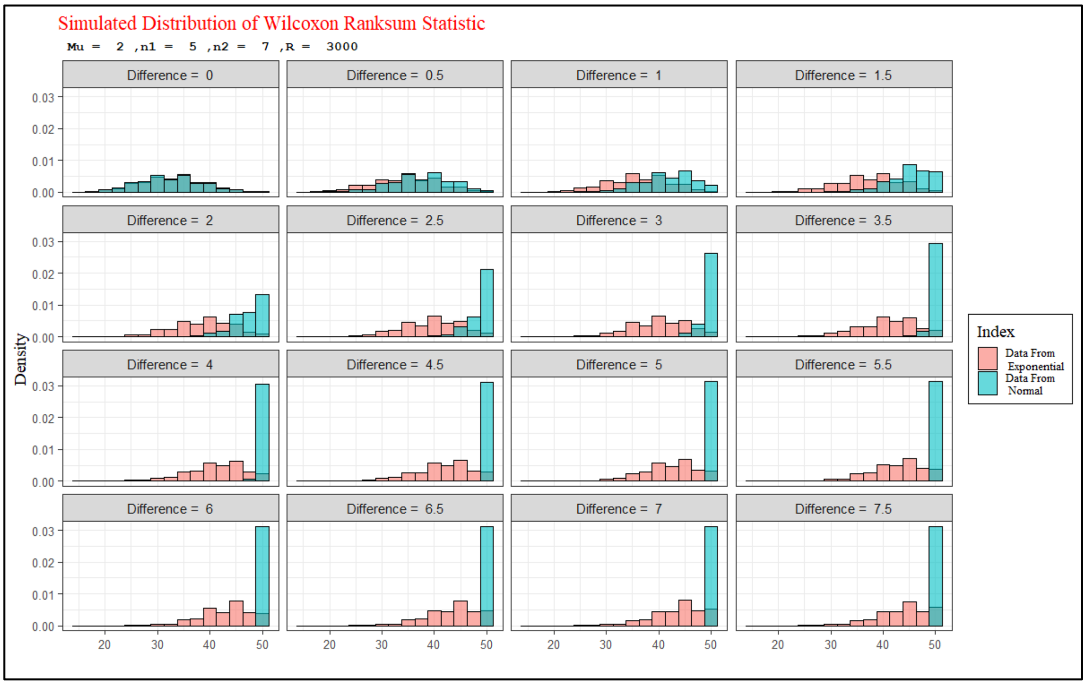


Diagram 59: Simulated Distribution of Wilcoxon Rank Sum Statistic for $mu = 2$ $n_1 = 5, n_2 = 7, R = 3000$

We know that the test statistic for Mann-Whitney U test and Wilcoxon Rank Sum test differs only by a constant (Which depend upon sample size from X population). That is why the simulated distributions of the Mann Whitney U statistic and Wilcoxon Rank Sum statistic is exactly same. There, is only a shift of origin in the distribution.

9.10.5. Observations:

So, we are seeing that the observations we get from previous simulations are very much consistent with the observations of this simulation. The tests may be non-parametric, but we have seen that the distribution of test-statistic under alternative hypothesis depends upon from which population samples are drawn. That is, power of the tests depends to some extent on the type of population sampled.

Also, it is important to note that the Mann-Whitney U test, Wilcoxon Rank Sum test is particularly useful for location alternatives. That is, when the two populations, from which samples are drawn differs only with respect to mean. In other words, when the two independent populations belong to the same location family. For, normal distrbution under H_1, the two independent populations belong to the same location family. But, for exponential distribution, under H_1, the two independent populations belong to the same scale family, not location family. So, that may also be reason for which Mann Whitney U test and Wilcoxon Rank Sum performs well when sampled population is normal.

10. Conclusion:

From the above analysis of empirical size and power of different parametric and non-parametric tests for two sample problem of equality of means of two independent populations, when sampled populations are normal or exponential, we have understood that, whatever be the situations, if the assumptions of parametric tests (here LRT) hold good, then parametric test is always better in terms of power than the non-parametric tests, which we have discussed. We have seen that Mann Whitney U test and Wilcoxon Rank Sum test is better than Kolmogorov-Smirnov test in almost every case. Although for small sample sizes, we generally see some exceptions.

In two sample problems, if the samples are known to have come from two independent normal populations having common variance and if someone still wishes to use non-parametric test for testing equality of means of these two independent normal populations instead of using the popular two sample T test, then Mann Whitney U test and Wilcoxon Rank Sum test are good choices than other non-parametric tests. Because, we have empirically seen that for sample sizes greater then 10, the Mann Whitney U test and Wilcoxon Rank Sum test perform almost similarly as Two sample T test in terms of power.

Also, if the samples are known to have come from two independent exponential populations and if someone still wishes to use non-parametric test for testing equality of means of these two independent exponential populations instead of using the parametric Exponential LRT, then Mann Whitney U test and Wilcoxon Rank Sum test are good choices than other non-parametric tests. Because, we have empirically seen that for sample sizes greater then 14, the Mann Whitney U test and Wilcoxon Rank Sum test perform almost similarly as Two sample exponential LRT in terms of power. (Here we could have use randomization to obtain exact size α test (For non-parametric tests), but in that case also, we will have same observations.)

Also, using simulated distributions of the test-statistic of different non-parametric tests, we have seen that the same non-parametric test is able to detect the same difference less frequently (i.e., power is less) when sampled population is exponential than when sampled population is normal. This may happen, because the Mann-Whitney U test and Wilcoxon Rank Sum test are particularly useful for location alternatives. That is, when the two populations, from which samples are drawn differs only with respect to location. In other words, when the two independent populations belong to the same location family. For, normal distrbution under H_1, the two independent populations belong to the same location family. But, for exponential distribution, under H_1, the two independent populations belong to the same scale family, not location family.

Thus, we can say that, if we have knowledge about type of population (here, we mean normal or exponential), from which samples are drawn, then we should go for the corresponding parametric tests. But still if we have doubt, then Mann Whitney U test or Wilcoxon Rank Sum test is a good choice for two sample problem related to equality of means of two independent populations.

11. Related Works and Prospective Works:

11.1 Related Works:

While doing this project, I tried to visualize the concept of power of a test in different ways. We have already understood that, when it is difficult to obtain any analytical expression of power function of test, we can go for empirical powers. Which are basically obtained using Monte-Carlo simulation. Again, the idea of power of a test can be visualized using the simulated distribution of the test-statistic under H_1. Here, we will use this approach, but in a sophisticated way. We will use animations to visualize the concept.

For this purpose, we have used three user defined functions, all having same arguments. The names of the functions are - $visualize\_Power.T1$, $visualize\_Power.LRT1$, $visualize\_Power.T2$, $visualize\_Power.LRT2$, $visualize\_Power.T3$, $visualize\_Power.LRT3$. The arguments of these functions are given below –

n_1, n_2: These are the size of the samples to be drawn from X and Y population respectively.

mu: Mean of Y population. (For normal) or, $lamda$: Mean of Y population (For Exponential).

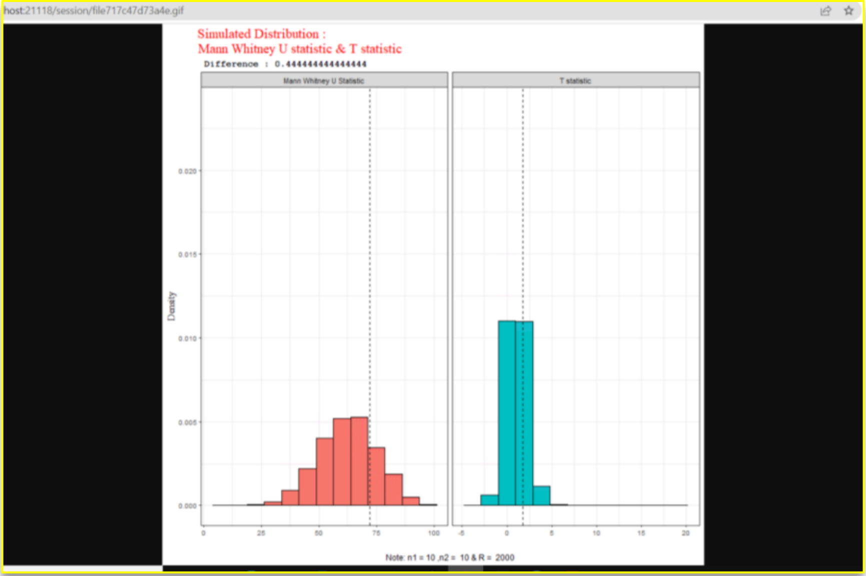
d: A vector of differences between population means of X and Y. We have already mentioned that mean of X population is $mu + d$ ($lamda + d$ for exponential). Where, mu ($lamda, for\ exponential$) is the argument of the function which is taken as input.

R: It is the replication number, that is number of times to repeat the whole simulation process.

$alpha$: It is level of significance of the test.

$exact.crit$: exact critical value of corresponding non-parametric tests.

From the function we get an animation as output. Which depicts, how as d (the difference between means of two independent population) increases, power changes, which is basically the area to the right of vertical line drawn in the animation. Since, animation cannot move in pdf. We will give a snapshot of it here.



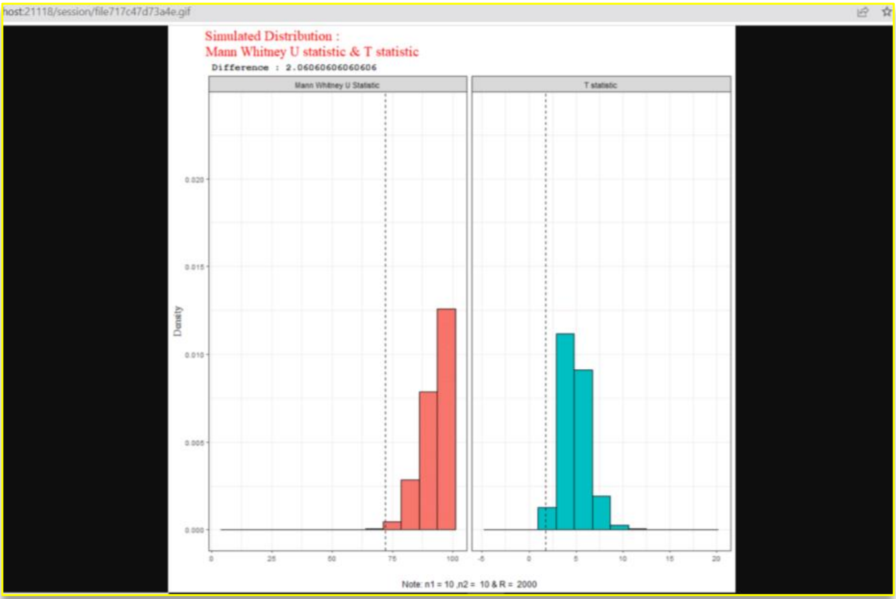


Diagram 60: Animated illustration of meaning of power for Mann Whitney U test and T test for $mu = 0,\ n_1 = 10, n_2 = 10, R = 2000$

Here, we have presented it for Mann Whitney U test only. $n_1 = 10, n_2 = 10, mu = 0, R = 2000$.

It is a great visualization of the underlying concept of power of a test.

11.2 Prospective Work:

We have already made a comparative study of size and power of parametric tests (LRT) and non-parametric tests in two sample problems (when sampled population is either normal or exponential). Now, we can do further study on -

1. How the two sample T test would behave when the assumption of homoscedasticity of two independent normal population is violated. In such a case, whether the popular non-parametric tests perform better than T test.
2. Whether the T test is robust to the assumption of normality of two independent populations. That is, how T test will behave if the sampled population is non normal. For example – Exponential, Double exponential, Cauchy, Gamma etc.
3. How in one sample problem or multi sample problem parametric tests (for example LRT) will behave as compared to the corresponding non-parametric tests.
4. Whether Two sample exponential LRT is robust to the assumption of population distribution (exponential)
5. Whether the large sample approximation of different non-parametric (which we have discussed) tests is valid for small sample sizes also. That is, how large sample sizes are needed for applicability of these approximations.

12. BIBLIOGRAPHY:

1. '*An Introduction to Probability and Statistics*' Book by *A. K. Md. Ehsanes Salah* and *V. K. Rohatgi*.
2. '*Introduction to the Theory of Statistics*' Book by *Alexander M. Mood, Duane C. Boes*, and *Franklin A. Graybill*
3. '*Nonparametric Statistical Inference*' Book by *Jean D. Gibbons* and *Subhabrata Chakraborti*
4. '*Statistical Inference*' Book by *George Casella and Roger Lee Berge*
5. http://www.ru.ac.bd/wp-content/uploads/sites/25/2019/03/308_03_Rubinstein_Simulation-and-the-Monte-Carlo-Method-Wiley-2017.pdf
6. http://onlinelibrary.wiley.com/doi/10.1002/9781118799635.oth1/pdf (Table of Critical Values)

13. APPENDIX

R CODES:

Required Packages

For the simulation studies, we will be required two packages $ggplot2, wesanderson$. They are basically needed for graphs and colors.

```
rm(list = ls(all = T))  #Removes all data

#Loading Required Library
library(ggplot2)              #for graphics
library(wesanderson)         #for colours
```

The codes of different simulations are added below.

SIMULATION 01:
Comparison of Empirical Size and Power of Two Sample T test and Kolmogorov-Smirnov Test When Sampled Population is Normal

```
# Here We will do a comparative study of Kolmogorov Smirnov test &
# T test in two sample problems, where data is from Normal Distribution

# A function to calculate D statistic
Kolmogorv_smirnov.stat <- function(x,y){

  Fx <- ecdf(x)     #empirical CDF of X
  Fy <- ecdf(y)     #empirical CDF of Y

  D <- max(Fy(c(x,y)) - Fx(c(x,y)))  #computing D statistic

  return(D)   #returns the value of u
}

#Equal Variance

Power_comparison.T1 <- function(n1,n2,d,mu,R,alpha,exact.crit = F){

  # The function takes sample sizes,a vector of differences &
  # Level of Significance as argument

  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)    #A matrix to store
power values
  colnames(power.matrix) = c('Kolmogorov Smirnov Test','T Test')
```

```
  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rnorm(n1,d[i] + mu,1); y <- rnorm(n2,mu,1)  #our sample

      #computation of test statistics

      #Kolmogorov U test statistic
      Kolmogorov_smirnov.statistic <- Kolmogorv_smirnov.stat(x,y)

      #T test statistic
      est.var <- ((n1-1)*var(x) + (n2-1)*var(y))/(n1+n2-2)   #pooled
variance
      T.statistic <- (mean(x) - mean(y))/sqrt(est.var*((1/n1) + (1/n2)))

      #Storing the values
      c(Kolmogorov_smirnov.statistic,T.statistic)
    })

    #Simulated power for Kolmogorov Smirnov test
    if(exact.crit){
      Kolmogorov_smirnov.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      cut.point.1 <- sqrt(-log(alpha)/2)
      Kolmogorov_smirnov.power <-
mean(test.statistic[1,]*sqrt((n1*n2)/(n1+n2)) > cut.point.1)
    }

    #Simulated power for t test
    cutpoint.2 <- qt(alpha,n1+n2-2,lower.tail = F)
    T.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(Kolmogorov_smirnov.power,T.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}


#Function to draw power curve
Visualize_Power_comparison.T1 <- function(n1,n2,d,mu,R,alpha,exact.crit =
F){

  #Storing Simulated Powers
  M <- Power_comparison.T1(n1,n2,d,mu,R,alpha,exact.crit)

  #Graph
  index.1 <- rep(c('Kolmogorov-Smirnov \n Test','T Test '),each = nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
```

```
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.5)
+
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Mu = ',mu,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Kolmogorov Smirnov Test & T Test') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 02:
Comparison of Empirical Size and Power of Two sample Exponential LRT and Kolmogorov-Smirnov Test When Sampled Population is Exponential

```
# Here We will do a comparative study of Kolmogorov Smirnov test &
# Exponential LRT in two sample problems, where data is from Exponential
Distribution

Power_comparison.LRT1 <- function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  # The function takes sample sizes,a vector of differences &
  # Level of Significance as argument

  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)   #A matrix to store
power values
  colnames(power.matrix) = c('Kolmogorov Smirnov Test','Exponential LRT')

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{
```

```
      x <- rexp(n1,rate = 1/(lamda + d[i])); y <- rexp(n2,rate = 1/lamda)
#our sample

      #computation of test statistics

      #Kolmogorov Smirnov test statistic
      Kolmogorov_smirnov.statistic <- Kolmogorv_smirnov.stat(x,y)

      #Exponential LRT statistic
      Exponential_LRT.statistic <- mean(x)/mean(y)

      #Storing the values
      c(Kolmogorov_smirnov.statistic,Exponential_LRT.statistic)
    })

    #Simulated power for Kolmogorov Smirnov test
    if(exact.crit){
      Kolmogorov_smirnov.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      cut.point.1 <- sqrt(-log(alpha)/2)
      Kolmogorov_smirnov.power <-
mean(test.statistic[1,]*sqrt((n1*n2)/(n1+n2)) > cut.point.1)
    }

    #Simulated power for Exponential LRT
    cutpoint.2 <- qf(alpha,2*n1,2*n2,lower.tail = F)
    Exponential_LRT.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(Kolmogorov_smirnov.power,Exponential_LRT.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}

#Function to Draw Power Curve
Visualize_Power_comparison.LRT1 <-
function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  #Storing Simulated Powers
  M <- Power_comparison.LRT1(n1,n2,d,lamda,R,alpha,exact.crit)

  #Graph
  index.1 <- rep(c('Kolmogorov-Smirnov \n Test','Exponential \n LRT'),each
= nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.5)
+
```

```r
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Lamda =',lamda,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Kolmogorov Smirnov Test &
Exponential LRT') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)

}
```

SIMULATION 03:
Comparison of Empirical Size and Power of Two sample T Test and Mann-Whitney U Test When Sampled Population is Normal

```r
# Here We will do a comparative study of Mann Whitney U test &
# T test in two sample problems, where data is from Normal Distribution

# A function to calculate U statistic
Ustat <- function(x,y){
  u <- 0          #to store the value
  for(i in 1:length(x)){
    u <-  u + sum(y < x[i])
  }
  return(u)   #returns the value of u
}

#Equal Variance

Power_comparison.T2 <- function(n1,n2,d,mu,R,alpha,exact.crit = F){

  # The function takes sample sizes,a vector of differences &
  # Replication Number,Level of Significance as argument
```

```
  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)   #A matrix to store
power values
  colnames(power.matrix) = c('MWU test','T Test')

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rnorm(n1,d[i] + mu,1); y <- rnorm(n2,mu,1)  #our sample

      #computation of test statistics

      #Mann Whitney U test statistic
      MWU.statistic <- Ustat(x,y)

      #T test statistic
      est.var <- ((n1-1)*var(x) + (n2-1)*var(y))/(n1+n2-2)   #pooled
variance
      T.statistic <- (mean(x) - mean(y))/sqrt(est.var*((1/n1) + (1/n2)))

      #Storing the values
      c(MWU.statistic,T.statistic)
    })

    #Simulated power for Mann Whitney U test
    if(exact.crit){
      MNW.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      a <- (n1*n2)/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.1 <- qnorm(alpha,lower.tail = F)
      MNW.power <- mean((test.statistic[1,] - a- 0.5)/sqrt(b) >
cutpoint.1) #after continuity correction
    }

    #Simulated power for t test
    cutpoint.2 <- qt(alpha,n1+n2-2,lower.tail = F)
    T.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(MNW.power,T.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}

#Function to draw Power Curve
Visualize_Power_comparison.T2 <- function(n1,n2,d,mu,R,alpha,exact.crit =
F){
```

```r
  #Storing Simulated Powers
  M <- Power_comparison.T2(n1,n2,d,mu,R,alpha,exact.crit)

  #Graph
  index.1 <- rep(c('Mann Whitney \n U Test','T Test'),each = nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.1)
+
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Mu = ',mu,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Mann Whitney U Test & T Test') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 04:
Comparison of Empirical Size and Power of Two sample Exponential LRT and Mann Whitney U Test When Sampled Population is Exponential

```r
# Here We will do a comparative study of Mann Whitney U test &
# Exponential LRT in two sample problems, where data is from Exponential
Distribution

Power_comparison.LRT2 <- function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  # The function takes sample sizes,a vector of diffreneces &
  # Level of Significance as argument

  set.seed(seed = 987654321)    #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)    #A matrix to store
```

```
power values
  colnames(power.matrix) = c('Mann Whitney U test','Exponential LRT')

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rexp(n1,rate = 1/(lamda + d[i])); y <- rexp(n2,rate = 1/lamda)
#our sample

      #computation of test statistics

      #Mann Whitney U test statistic
      MWU.statistic <- Ustat(x,y)

      #Exponential LRT statistic
      Exponential_LRT.statistic <- mean(x)/mean(y)

      #Storing the values
      c(MWU.statistic,Exponential_LRT.statistic)
    })

    #Simulated power for Mann Whitney U test
    if(exact.crit){
      MNW.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      a <- (n1*n2)/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.1 <- qnorm(alpha,lower.tail = F)
      MNW.power <- mean((test.statistic[1,] - a - 0.5)/sqrt(b) >
cutpoint.1)  #after continuity correction
    }

    #Simulated power for Exponential LRT of equality of mean
    cutpoint.2 <- qf(alpha,2*n1,2*n2,lower.tail = F)
    Exponential_LRT.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(MNW.power,Exponential_LRT.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}

#Function to draw power curve
Visualize_Power_comparison.LRT2 <-
function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  #Storing Simulated Powers
  M <- Power_comparison.LRT2(n1,n2,d,lamda,R,alpha,exact.crit)

  #Graph
```

```
  index.1 <- rep(c('Mann Whitney \n U Test','Exponential \n LRT'),each =
nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.5)
+
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Lamda =',lamda,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Mann Whitney U Test & Exponential
LRT') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 05:
Comparison of Empirical Size and Power of Two sample T Test and Wilcoxon Rank Sum Test When Sampled Population is Normal

```
# Here We will do a comparative study of Wilcoxon Rank Sum test &
# T test in two sample problems, where data is from Normal Distribution

# A function to calculate U statistic
Wilcoxon_Ranksum_stat <- function(x,y){
  z <- c(x,y)
  u <- sum(rank(z)[c(1:length(x))])         #to store the value

  return(u)   #returns the value of u
}

#Equal Variance

Power_comparison.T3 <- function(n1,n2,d,mu,R,alpha,exact.crit = F){
```

```
  # The function takes sample sizes,a vector of differences &
  # Level of Significance as argument

  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)   #A matrix to store
power values
  colnames(power.matrix) = c('Wilcoxon Ranksum test','T Test')

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rnorm(n1,mu + d[i],1); y <- rnorm(n2,mu,1)  #our sample

      #computation of test statistics

      #Wilcoxon Rank Sum test statistic
      WilcoxonRankSum.statistic <- Wilcoxon_Ranksum_stat(x,y)

      #T test statistic
      est.var <- ((n1-1)*var(x) + (n2-1)*var(y))/(n1+n2-2)   #pooled
variance
      T.statistic <- (mean(x) - mean(y))/sqrt(est.var*((1/n1) + (1/n2)))

      #Storing the values
      c(WilcoxonRankSum.statistic,T.statistic)
    })

    #Simulated power for Wilcoxon RankSum test
    if(exact.crit){
      WilcoxonRankSum.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      a <- (n1*(n1+n2+1))/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.1 <- qnorm(alpha,lower.tail = F)
      WilcoxonRankSum.power <- mean((test.statistic[1,] - a - 0.5)/sqrt(b)
> cutpoint.1) #after continuity correction
    }

    #Simulated power for t test
    cutpoint.2 <- qt(alpha,n1+n2-2,lower.tail = F)
    T.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(WilcoxonRankSum.power,T.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}
```

```r
#Function to draw the power curve
Visualize_Power_comparison.T3 <- function(n1,n2,d,mu,R,alpha,exact.crit =
F){

  #Storing Simulated Powers
  M <- Power_comparison.T3(n1,n2,d,mu,R,alpha,exact.crit)

  #Graph
  index.1 <- rep(c('Wilcoxon Ranksum \n Test','T Test'),each = nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.1)
+
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Mu = ',mu,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Wilcoxon Ranksum Test & T Test') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                 face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 06:
Comparison of Empirical Size and Power of Two sample Exponential LRT and Wilcoxon Rank Sum Test When Sampled Population is Exponential

```r
# Here We will do a comparative study of Wilcoxon Rank Sum U test &
# Exponential LRT in two sample problems, where data is from Exponential
Distribution

# A function to calculate U statistic
Wilcoxon_Ranksum_stat <- function(x,y){
  z <- c(x,y)
  u <- sum(rank(z)[c(1:length(x))])         #to store the value
```

```
  return(u)   #returns the value of u
}

Power_comparison.LRT3 <- function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  # The function takes sample sizes,a vector of differences &
  # Level of Significance as argument

  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 2)   #A matrix to store
power values
  colnames(power.matrix) = c('Wilcoxon Ranksum test','Exponential LRT')


  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rexp(n1,rate = 1/(lamda + d[i])); y <- rexp(n2,rate = 1/lamda)
#our sample

      #computation of test statistics

      #Wilcoxon Rank Sum test statistic
      WilcoxonRankSum.statistic <- Wilcoxon_Ranksum_stat(x,y)

      #Exponential LRT statistic
      Exponential_LRT.statistic <- mean(x)/mean(y)

      #Storing the values
      c(WilcoxonRankSum.statistic,Exponential_LRT.statistic)
    })

    #Simulated power for Wilcoxon Rank Sum test
    if(exact.crit){
      WilcoxonRankSum.power <- mean(test.statistic[1,] > exact.crit)
    }else{
      a <- (n1*(n1+n2+1))/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.1 <- qnorm(alpha,lower.tail = F)
      WilcoxonRankSum.power <- mean((test.statistic[1,] - a - 0.5)/sqrt(b)
> cutpoint.1) #after continuity correction
    }

    #Simulated power for Exponential LRT of equality of mean
    cutpoint.2 <- qf(alpha,2*n1,2*n2,lower.tail = F)
    Exponential_LRT.power <- mean(test.statistic[2,] > cutpoint.2)

    #Storing the simulated powers in matrix
    power.matrix[i,] <- c(WilcoxonRankSum.power,Exponential_LRT.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
```

```
difference column

  return(power.matrix)     #returns a matrix
}


#Function to draw power Curve
Visualize_Power_comparison.LRT3 <-
function(n1,n2,d,lamda,R,alpha,exact.crit = F){

  #Storing Simulated Powers
  M <- Power_comparison.LRT3(n1,n2,d,lamda,R,alpha,exact.crit)

  #Graph
  index.1 <- rep(c('Wilcoxon Ranksum \n Test','Exponential \n LRT'),each =
nrow(M))
  graph.data <- data.frame(x = c(M[,1],M[,1]),y = c(M[,2],M[,3]),Index =
index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') + geom_line(size = 1.5)
+
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Lamda = ',lamda,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Wilcoxon Ranksum Test & Exponential
LRT') +
    theme_bw(14)

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 07:
Simultaneous Comparison of Empirical Size and Power of Two sample T Test and Different Non-Parametric Tests When Sampled Population is Normal

```
# Here We will do a comparative study of Different Non-Parametric test &
# T test in two sample problems, where data is from Normal Distribution
#Equal Variance

Power_comparisonALL.1 <- function(n1,n2,d,mu,R,alpha,exact.crit =
c(F,F,F)){

  # The function takes sample sizes,a vector of differences &
  # Level of Significance as argument

  set.seed(seed = 987654321)   #for uniformity of result

  power.matrix <- matrix(0,nrow = length(d),ncol = 4)   #A matrix to store
power values
  colnames(power.matrix) = c('MNW Test','Wilcoxon Rank-Sum
Test','Kolmogorov Smirnov Test','T Test')
  sigma <- 1

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      x <- rnorm(n1,d[i] + mu,sigma); y <- rnorm(n2,mu,sigma)  #our sample

      #computation of test statistics

      #Wilcoxon Rank Sum test statistic
      WilcoxonRankSum.statistic <- Wilcoxon_Ranksum_stat(x,y)

      #Mann Whitney U test statistic
      MWU.statistic <- Ustat(x,y)

      #Kolmogorov Smirnov test statistic
      Kolmogorov_smirnov.statistic <- Kolmogorv_smirnov.stat(x,y)

      #T test statistic
      est.var <- ((n1-1)*var(x) + (n2-1)*var(y))/(n1+n2-2)   #pooled
variance
      T.statistic <- (mean(x) - mean(y))/sqrt(est.var*((1/n1) + (1/n2)))

      #Storing the values

c(MWU.statistic,WilcoxonRankSum.statistic,Kolmogorov_smirnov.statistic,T.s
tatistic)
    })

    #Simulated power for Mann Whitney U test
    if(exact.crit[1]){
      MNW.power <- mean(test.statistic[1,] > exact.crit[1])
    }else{
      a <- (n1*n2)/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.1 <- qnorm(alpha,lower.tail = F)
```

```
      MNW.power <- mean((test.statistic[1,] - a- 0.5)/sqrt(b) >
cutpoint.1) #after continuity correction
    }

    #Simulated power for Wilcoxon Rank Sum test
    if(exact.crit[2]){
      WilcoxonRankSum.power <- mean(test.statistic[2,] > exact.crit[2])
    }else{
      a <- (n1*(n1+n2+1))/2 ; b <- n1*n2*(n1+n2+1)/12
      cutpoint.2 <- qnorm(alpha,lower.tail = F)
      WilcoxonRankSum.power <- mean((test.statistic[2,] - a - 0.5)/sqrt(b)
> cutpoint.2) #after continuity correction
    }

    #Simulated power for Kolmogorov Smirnov test
    if(exact.crit[3]){
      Kolmogorov_smirnov.power <- mean(test.statistic[3,] > exact.crit[3])
    }else{
      cut.point.3 <- sqrt(-log(alpha)/2)
      Kolmogorov_smirnov.power <-
mean(test.statistic[3,]*sqrt((n1*n2)/(n1+n2)) > cut.point.3)
    }

    #Simulated power for t test
    cutpoint.4 <- qt(alpha,n1+n2-2,lower.tail = F)
    T.power <- mean(test.statistic[4,] > cutpoint.4)

    #Storing the simulated powers in matrix
    power.matrix[i,] <-
c(MNW.power,WilcoxonRankSum.power,Kolmogorov_smirnov.power,T.power)
  }

  power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

  return(power.matrix)     #returns a matrix
}

#FUnction to Draw Power Curve
Visualize_Power_comparisonALL.1 <- function(n1,n2,d,mu,R,alpha,exact.crit
= c(F,F,F)){

  #To store power
  M <- Power_comparisonALL.1(n1,n2,d,mu,R,alpha,exact.crit)

  #Graph
  index.1 <- factor(rep(c('Mann Whitney \n U Test','Wilcoxon Ranksum \n
Test','Kolmogorov-Smirnov \n Test','T test'),each = nrow(M)),
                    levels = c('Mann Whitney \n U Test','Wilcoxon Ranksum
\n Test','Kolmogorov-Smirnov \n Test','T test'))
  graph.data <- data.frame(x = c(M[,1],M[,1],M[,1],M[,1]),y =
c(M[,2],M[,3],M[,4],M[,5]),Index = index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') +
```

```
geom_line(aes(linetype = Index),size = 1) +
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Mu = ',mu,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Different Non Parametric Tests and T
Test') +
    theme_bw(14) + scale_color_manual(values = wes_palette(n = 4,name =
"Darjeeling1"))

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 08:
Simultaneous Comparison of Empirical Size and Power of Two sample Exponential LRT and Different Non-Parametric Tests When Sampled Population is Exponential

```
# Here We will do a comparative study of Different Non-Parametric &
# Exponential LRT in two sample problems, where data is from Exponential
Distribution

Power_comparisonALL.2 <- function(n1,n2,d,lamda,R,alpha,exact.crit =
c(F,F,F)){

    # The function takes sample sizes,a vector of differences &
    # Level of Significance as argument

    set.seed(seed = 987654321)    #for uniformity of result

    power.matrix <- matrix(0,nrow = length(d),ncol = 4)    #A matrix to
store power values
    colnames(power.matrix) = c('Mann Whitney U Test','Wilcoxon Rank-Sum
Test','Kolmogorov Smirnov Test','Exponential LRT')

    for(i in 1:length(d)){
```

```
      test.statistic <- replicate(R,{

        x <- rexp(n1,rate = 1/(lamda + d[i])); y <- rexp(n2,rate =
1/lamda)  #our sample

        #computation of test statistics

        #Wilcoxon Rank Sum test statistic
        WilcoxonRankSum.statistic <- Wilcoxon_Ranksum_stat(x,y)

        #Mann Whitney U test statistic
        MWU.statistic <- Ustat(x,y)

        #Kolmogorov Smirnov test statistic
        Kolmogorov_smirnov.statistic <- Kolmogorv_smirnov.stat(x,y)

        #Exponential LRT statistic
        Exponential_LRT.statistic <- mean(x)/mean(y)

        #Storing the values

c(MWU.statistic,WilcoxonRankSum.statistic,Kolmogorov_smirnov.statistic,Exp
onential_LRT.statistic)
      })

      #Simulated power for Mann Whitney U test
      if(exact.crit[1]){
        MNW.power <- mean(test.statistic[1,] > exact.crit[1])
      }else{
        a <- (n1*n2)/2 ; b <- n1*n2*(n1+n2+1)/12
        cutpoint.1 <- qnorm(alpha,lower.tail = F)
        MNW.power <- mean((test.statistic[1,] - a- 0.5)/sqrt(b) >
cutpoint.1) #after continuity correction
      }

      #Simulated power for Wilcoxon Rank Sum test
      if(exact.crit[2]){
        WilcoxonRankSum.power <- mean(test.statistic[2,] > exact.crit[2])
      }else{
        a <- (n1*(n1+n2+1))/2 ; b <- n1*n2*(n1+n2+1)/12
        cutpoint.2 <- qnorm(alpha,lower.tail = F)
        WilcoxonRankSum.power <- mean((test.statistic[2,] - a -
0.5)/sqrt(b) > cutpoint.2) #after continuity correction
      }

      #Simulated power for Kolmogorov Smirnov test
      if(exact.crit[3]){
        Kolmogorov_smirnov.power <- mean(test.statistic[3,] >
exact.crit[3])
      }else{
        cut.point.3 <- sqrt(-log(alpha)/2)
        Kolmogorov_smirnov.power <-
mean(test.statistic[3,]*sqrt((n1*n2)/(n1+n2)) > cut.point.3)
      }
```

```
      #Simulated power for Exponential LRT
      cutpoint.4 <- qf(alpha,2*n1,2*n2,lower.tail = F)
      Exponential_LRT.power <- mean(test.statistic[4,] > cutpoint.4)

      #Storing the simulated powers in matrix
      power.matrix[i,] <-
c(MNW.power,WilcoxonRankSum.power,Kolmogorov_smirnov.power,Exponential_LRT
.power)
    }

    power.matrix <- cbind(Diiference = d,power.matrix)      #adding the
difference column

    return(power.matrix)     #returns a matrix
  }


#Function to draw power curve
Visualize_Power_comparisonALL.2 <-
function(n1,n2,d,lamda,R,alpha,exact.crit = c(F,F,F)){

  #Storing power in matrix
  M <- Power_comparisonALL.2(n1,n2,d,lamda,R,alpha,exact.crit)

  #Graph
  index.1 <- factor(rep(c('Mann Whitney \n U Test','Wilcoxon Ranksum \n
Test','Kolmogorov-Smirnov \n Test','Exponential \n LRT'),each = nrow(M)),
                    levels = c('Mann Whitney \n U Test','Wilcoxon Ranksum
\n Test','Kolmogorov-Smirnov \n Test','Exponential \n LRT'))
  graph.data <- data.frame(x = c(M[,1],M[,1],M[,1],M[,1]),y =
c(M[,2],M[,3],M[,4],M[,5]),Index = index.1)
  graph.1 <- ggplot(graph.data,aes(x,y,col = Index)) +
    geom_hline(yintercept = 1,linetype = 'dashed') +
geom_line(aes(linetype = Index),size = 1) +
    geom_point(show.legend = F,size = 2) +
    labs(x = expression(paste('Difference ',(mu[1] - mu[2]))),y =
'Simulated Power',
         subtitle = paste('Lamda = ',lamda,',n1 =',n1,',n2 = ',n2,',Alpha
=',alpha,',R = ',R)) +
    ggtitle('Simulated Power Curve of Different Non Parametric Tests and
Exponential LRT') +
    theme_bw(14) + scale_color_manual(values = wes_palette(n = 4,name =
"Darjeeling1"))

  #Self_defined theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
```

```
                    legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                    legend.box.background = element_rect(colour = "black"))
  #Final Output
  return(graph.1 + mytheme)
}
```

SIMULATION 09:
Study of Probability Distributions of Test Statistics of Different Non-parametric Tests using Simulation

```
#Function to draw histogram Simulated distribution of Kolmogorov Smirnov
test statistic
Simulated_KolmogorovSmirnov <- function(n1,n2,mu,d,R){

  set.seed(seed = 987654321)           #for uniformity

  Simulated.Distribution <- NULL      #to store statistic value

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      #sample from normal distribution
      x.normal <- rnorm(n1,mean = mu + d[i],sd = 1)
      y.normal <- rnorm(n2,mean = mu,sd = 1)

      #sample from exponential distribution
      x.exp <- rexp(n1,rate = 1/(mu + d[i]))
      y.exp <- rexp(n2,rate = 1/mu)

      #calculating value of statistic
      KS.stat_Normal <- Kolmogorv_smirnov.stat(x.normal,y.normal)
      KS.stat_Exp <- Kolmogorv_smirnov.stat(x.exp,y.exp)

      c(KS.stat_Normal,KS.stat_Exp)

    })

    #Simulated Distribution
    Index.1 <- rep(paste('Difference = ',d[i]),each = R)
    Index.2 <- rep(c('Data From \n Normal','Data From \n
Exponential'),each = R)
    M <- data.frame(Value = c(test.statistic[1,],test.statistic[2,]),
                    Index1 = c(Index.1,Index.1),
                    Index2 = Index.2)
    Simulated.Distribution <- rbind(Simulated.Distribution,M)
  }
```

```
  #Graph data
  Simulated.Distribution[,2] <- factor(Simulated.Distribution[,2],levels =
paste('Difference = ',d))
  graphdata <- data.frame(Value = Simulated.Distribution[,1],Index =
Simulated.Distribution[,3],
                          diff_index = Simulated.Distribution[,2])

  #Graph
  graph.1 <- ggplot(graphdata,aes(Value,fill = Index)) +
    geom_histogram(aes(y = stat(count) / sum(count)),alpha = 0.6,bins =
15,col = 'black',
                   position = 'identity') +
    labs(x = '',y = 'Density',title = 'Simulated Distribution of
Kolmogorov Smirnov Statistic',
         subtitle = paste('Mu = ',mu,',n1 = ',n1,',n2 = ',n2,',R = ',R)) +
    facet_wrap(.~diff_index) + scale_x_continuous(breaks=seq(0,1,by=0.1))
+
    theme_bw(14)

  #Self Defined Theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))

  return(graph.1 + mytheme)
}

#Function to draw histogram Simulated distribution of Mann Whitney U
statistic
Simulated_MannWhiteny <- function(n1,n2,mu,d,R){

  set.seed(seed = 987654321)        #for uniformity

  Simulated.Distribution <- NULL   #to store the data

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      #sample from normal distribution
      x.normal <- rnorm(n1,mean = mu + d[i],sd = 1)
      y.normal <- rnorm(n2,mean = mu,sd = 1)

      #sample from exponential distribution
      x.exp <- rexp(n1,rate = 1/(mu + d[i]))
      y.exp <- rexp(n2,rate = 1/mu)
```

```
      #calculating value of statistic
      MNW.stat_Normal <- Ustat(x.normal,y.normal)
      MNW.stat_Exp <- Ustat(x.exp,y.exp)

      c(MNW.stat_Normal,MNW.stat_Exp)

    })

    #Simulated Distribution
    Index.1 <- rep(paste('Difference = ',d[i]),each = R)
    Index.2 <- rep(c('Data From \n Normal','Data From \n
Exponential'),each = R)
    M <- data.frame(Value = c(test.statistic[1,],test.statistic[2,]),
                    Index1 = c(Index.1,Index.1),
                    Index2 = Index.2)
    Simulated.Distribution <- rbind(Simulated.Distribution,M)
  }

  #Graph data
  Simulated.Distribution[,2] <- factor(Simulated.Distribution[,2],levels =
paste('Difference = ',d))
  graphdata <- data.frame(Value = Simulated.Distribution[,1],Index =
Simulated.Distribution[,3],
                          diff_index = Simulated.Distribution[,2])

  #Graph
  graph.1 <- ggplot(graphdata,aes(Value,fill = Index)) +
    geom_histogram(aes(y = stat(count) / sum(count)),alpha = 0.6,bins =
15,col = 'black',
                   position = 'identity') +
    labs(x = '',y = 'Density',title = 'Simulated Distribution of Mann
Whitney U Statistic',
         subtitle = paste('Mu = ',mu,',n1 = ',n1,',n2 = ',n2,',R = ',R)) +
    facet_wrap(.~diff_index) + theme_bw(14)

  #Self Defined Theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))

  return(graph.1 + mytheme)
}
```

```r
#Function to draw histogram Simulated distribution of Wilcoxon Rank Sum
test statistic
Simulated_WilcoxonRanksum <- function(n1,n2,mu,d,R){

  set.seed(seed = 987654321)           #for uniformity

  Simulated.Distribution <- NULL      #to store statistic value

  for(i in 1:length(d)){

    test.statistic <- replicate(R,{

      #sample from normal distribution
      x.normal <- rnorm(n1,mean = mu + d[i],sd = 1)
      y.normal <- rnorm(n2,mean = mu,sd = 1)

      #sample from exponential distribution
      x.exp <- rexp(n1,rate = 1/(mu + d[i]))
      y.exp <- rexp(n2,rate = 1/mu)

      #calculating value of statistic
      Wilcoxon.stat_Normal <- Wilcoxon_Ranksum_stat(x.normal,y.normal)
      Wilcoxon.stat_Exp <- Wilcoxon_Ranksum_stat(x.exp,y.exp)

      c(Wilcoxon.stat_Normal,Wilcoxon.stat_Exp)

    })

    #Simulated Distribution
    Index.1 <- rep(paste('Difference = ',d[i]),each = R)
    Index.2 <- rep(c('Data From \n Normal','Data From \n
Exponential'),each = R)
    M <- data.frame(Value = c(test.statistic[1,],test.statistic[2,]),
                    Index1 = c(Index.1,Index.1),
                    Index2 = Index.2)
    Simulated.Distribution <- rbind(Simulated.Distribution,M)
  }

  #Graph data
  Simulated.Distribution[,2] <- factor(Simulated.Distribution[,2],levels =
paste('Difference = ',d))
  graphdata <- data.frame(Value = Simulated.Distribution[,1],Index =
Simulated.Distribution[,3],
                          diff_index = Simulated.Distribution[,2])

  #Graph
  graph.1 <- ggplot(graphdata,aes(Value,fill = Index)) +
    geom_histogram(aes(y = stat(count) / sum(count)),alpha = 0.6,bins =
15,col = 'black',
                   position = 'identity') +
    labs(x = '',y = 'Density',title = 'Simulated Distribution of Wilcoxon
Ranksum Statistic',
         subtitle = paste('Mu = ',mu,',n1 = ',n1,',n2 = ',n2,',R = ',R)) +
    facet_wrap(.~diff_index) + theme_bw(14)
```

```
  #Self Defined Theme
  mytheme <- theme(plot.subtitle = element_text(family = "mono",size = 11,
                                                face = "bold",hjust =
0.01),axis.title = element_text(family = "serif"),
                   axis.text = element_text(size = 10),plot.title =
element_text(family = "serif",

colour = "red", hjust = -0.01),legend.text = element_text(size = 10,family
= "serif"),
                   legend.title = element_text(family =
"serif"),legend.background = element_blank(),
                   legend.box.background = element_rect(colour = "black"))

  return(graph.1 + mytheme)
}
```