Kolmogorov-Smirnov Test

A non-parametric Test for Goodness of fit

Uttaran Chatterjee(MD2227), Adrija Saha(MD2203), Shrayan Roy(MD2220)

Indian Statistical Institute (Delhi Centre)

28/03/2023

Kolmogorov-Smirnov Test as a Test of Goodness-of-Fit :

- Suppose we have a random sample $X_1, X_2, \ldots . X_n$ from some population. We want to fit a distribution to the unknown population by that what we mean is that we want to check that whether the sample can be considered as random sample from a population with a continuous distribution function F_o which is completely specified (for now) to us.

- Hence we set our null hypothesis as

$$\mathcal{H}_o : F(x) = F_o(x) \; for \; all \; x \in \mathbb{R}$$

- We can consider several alternate hypothesis from the as,

$$\mathcal{H}_1 : F(x) \neq F_o(x) \; for \; some \; x \in \mathbb{R} \; , \; \mathcal{H}_2 : F(x) \geq F_o(x) \; or \; \mathcal{H}_3 : F(x) \leq F_o(x) \; for \; all \; x \in \mathbb{R}$$

.

- In our testing problem we basically want to estimate F and check whether it agrees or disagrees with our above hypothesis.
- Since as we know for a fixed value of x the quantity $F(x)$ is nothing but a probability value which we generally not aware of, and our natural intuition of proportion and our inclination towards averages insists us to define an estimate of $F(x)$ as,

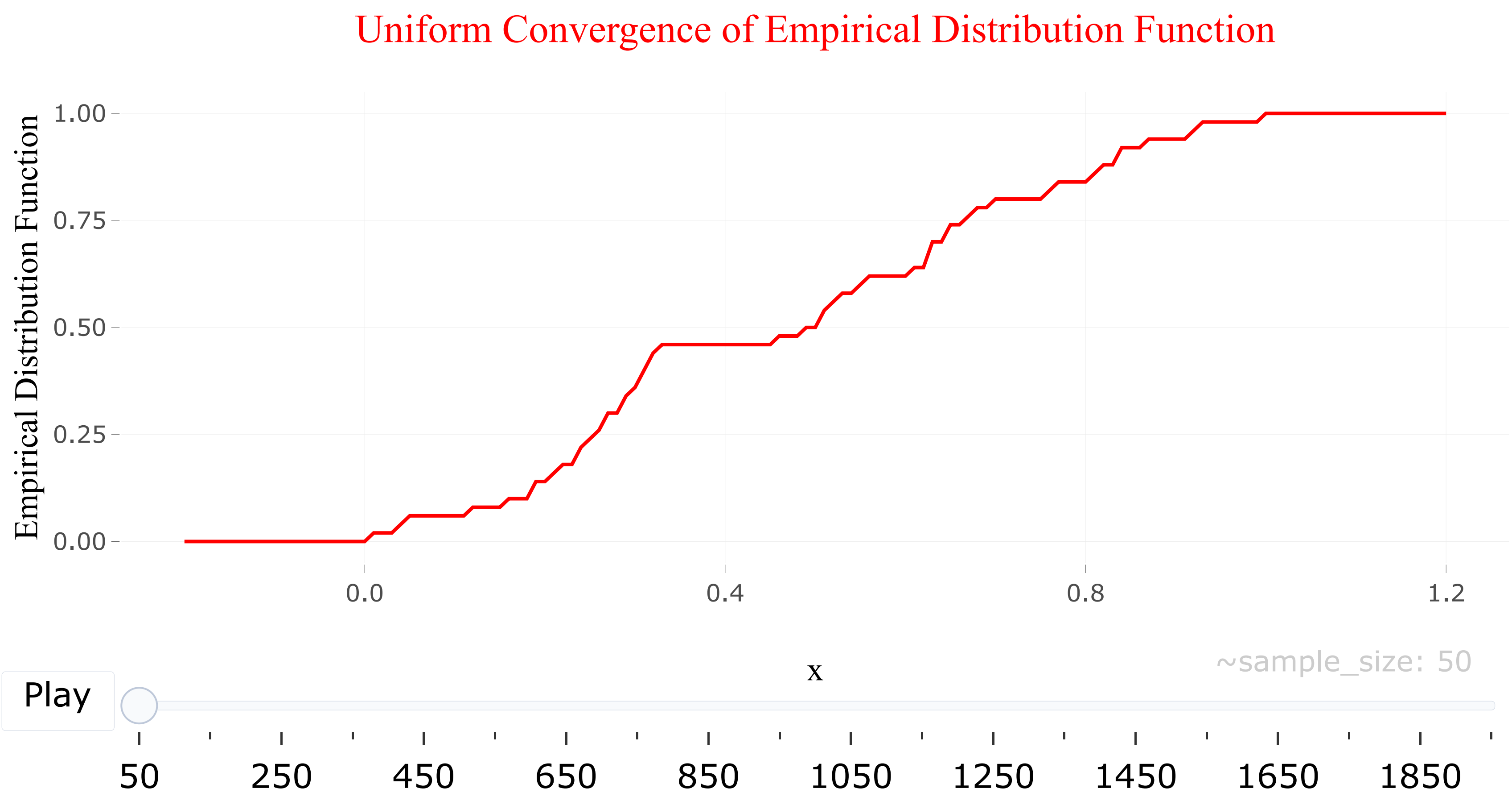
$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}$$

- This function defined above is called the empirical distribution function.

Empirical Distribution Function(ECDF) as an Estimator of the Distribution Function :

- Some quick observations that we can make immediately is that,
- $$n\mathbb{F}_n(x) \sim Binomial(n, F(x)).$$
- Also, $\textbf{Weak Law of Large Numbers}$ tells us, $\mathbb{F}_n(x) \xrightarrow{\mathbb{P}} F(x)$ as $n \to \infty$ for every $x \in \mathbb{R}$
- Hence we see that the empirical distribution function is weakly(infact strongly) consistent for the true distribution function.

Animatic View :



Kolmogorov-Smirnov Statistic :

- The Kolmogorov-Smirnov Statistic is defined as,

$$D_n = Sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - F_o(x)|.$$

- Further we can also define,

$$D_n{}^+ = Sup_{x \in \mathbb{R}} (\mathbb{F}_n(x) - F_o(x)) \ and \ D_n{}^- = Sup_{x \in \mathbb{R}} (F_o(x) - \mathbb{F}_n(x))$$

- What the quantity D_n is actually quantifying is the distance two functions $\mathbb{F}_n$ and F under the supremum metric.
- Hence, if our sample $X_1, \ldots., X_n$ is really a sample drawn from F_o then we expect D_n (even D_n^+ and D_n^-) to give negligible value. -Hence we reject null for large values of D_n (or D_n^+ or D_n^-).

Glivenko-Cantelli Theorem - The Fundamental Statistical Theorem :

- One of the theoretical motivation behind the use of this Kolmogorov statistic is the following theorem -
- $\mathbf{Glivenko\text{-}Cantelli\ Theorem}$

  For $\{X_n\}_{n\geq 1}$ be a sequence of random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If we define the empirical distribution function (edf) as defined earier, then we have,

$$\mathbb{P}\left(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)| = 0\right) = 1$$

- This theorem was proved by Glivenko for continuous distributions and the Cantelli proved the theorem for any distribution function.
- What the theorem says is remarkable and strong as it says that the true distribution function F can be $rediscovered\ from\ the\ data$ after making sufficiently large numbers of observations.

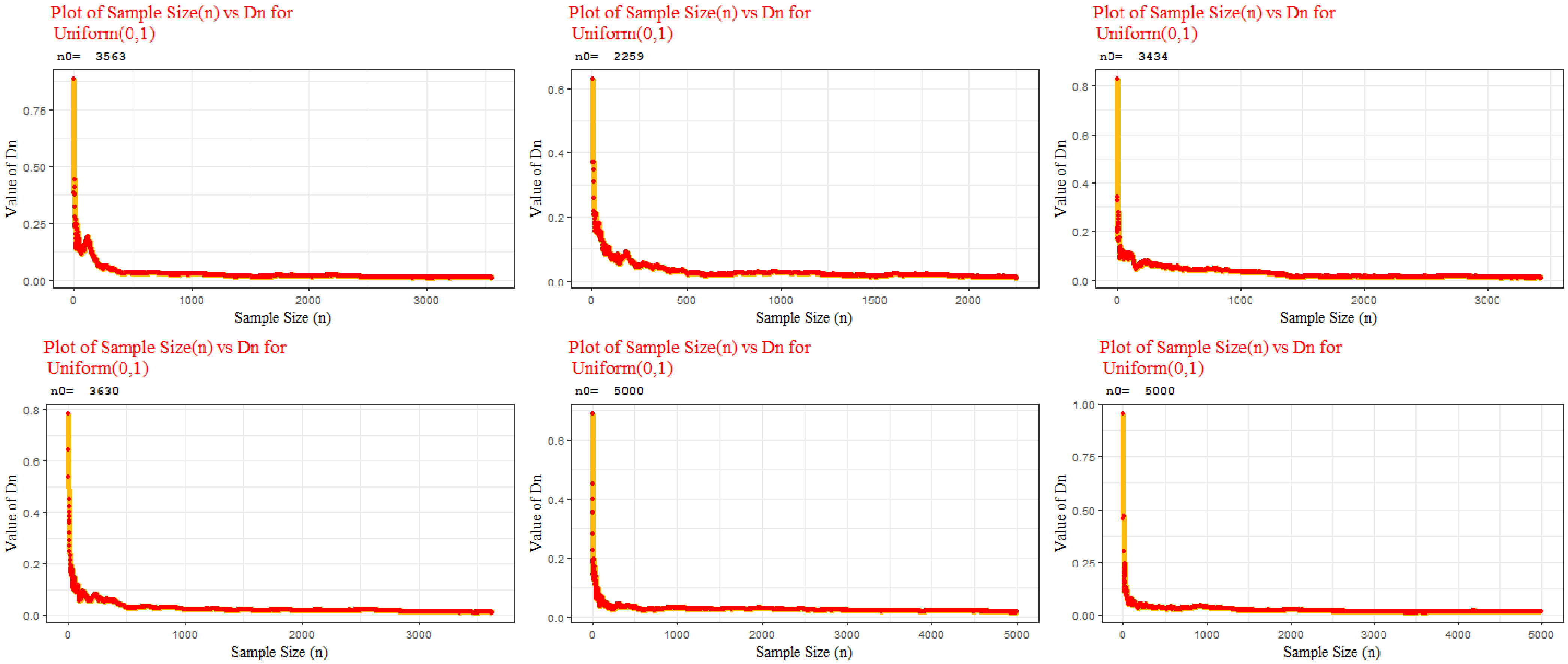
Glivenko-Cantelli Theorem - The Fundamental Statistical Theorem (Contd.) :

- The theorem says that suppose there are two experimenters A and B keep on taking sequence of observations from the same population, then for the A lets say her sample is $X_1(\omega), X_2(\omega), \ldots\ldots\ldots$ and B has drawn her sample to be $X_1(\omega'), X_2(\omega'), \ldots\ldots\ldots$ for $\omega, \omega' \in \Omega$ respectively, the theorem above ensures that the empirical distribution function $\mathbb{F}_n(x)$ converges to $F(x)$ uniformly in $x \in \mathbb{R}$ such that $\omega, \omega' \in N \subseteq \Omega$ and $\mathbb{P}(N) = 1$.

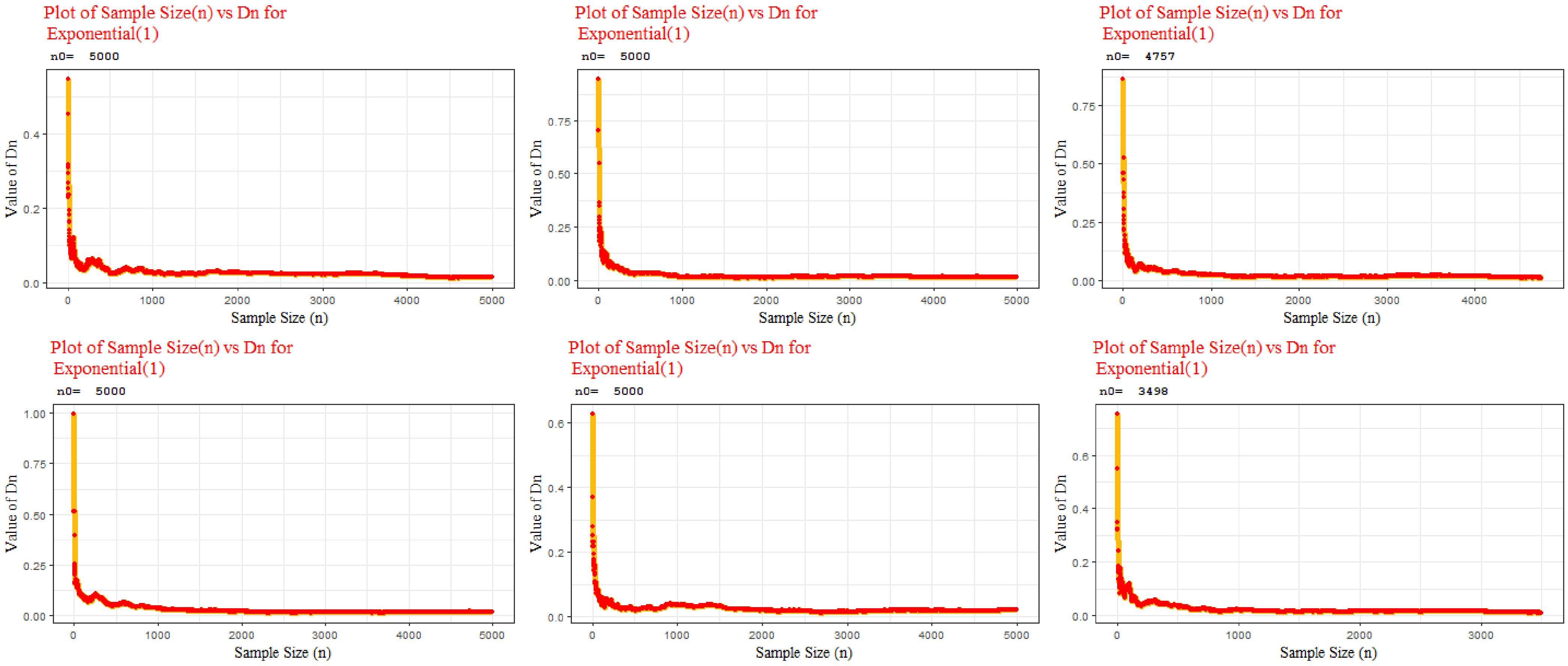
- Hence, it was quite rightfully referred as the **Fundamental Statistical Theorem** by Renyi and as **Central Statistical Theorem** by Loeve.

- Clearly, one of the immediate consequence of this theorem is **Kolmogorov-Smirnov Test** which we will be studying in detail through simulations.

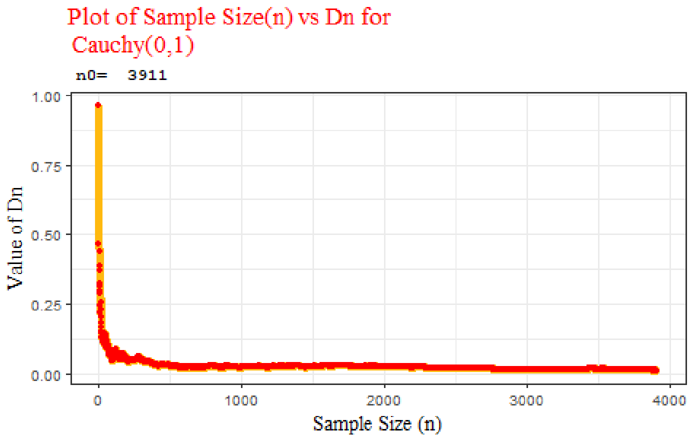
Convergence of the Kolmogorov Statistics D_n :

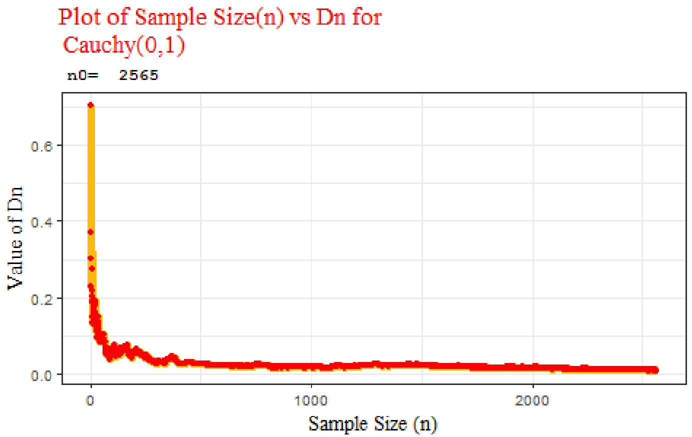


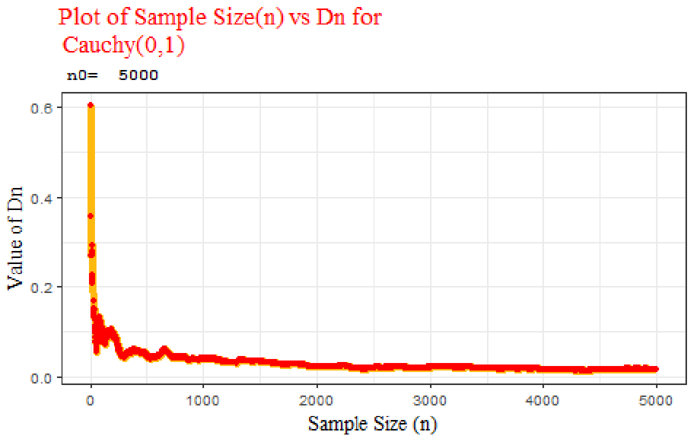
Convergence of D_n for Exponential :

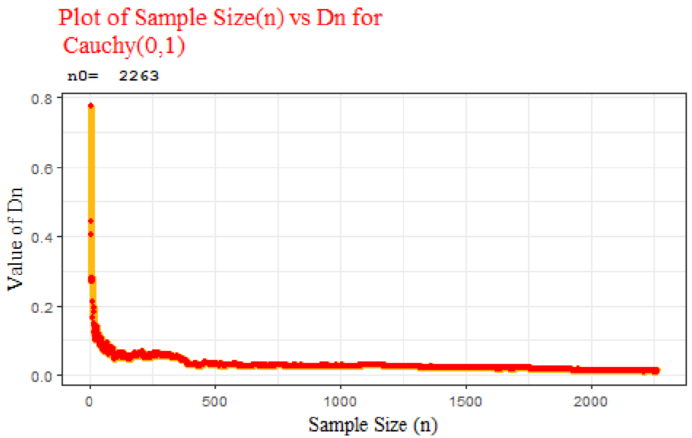


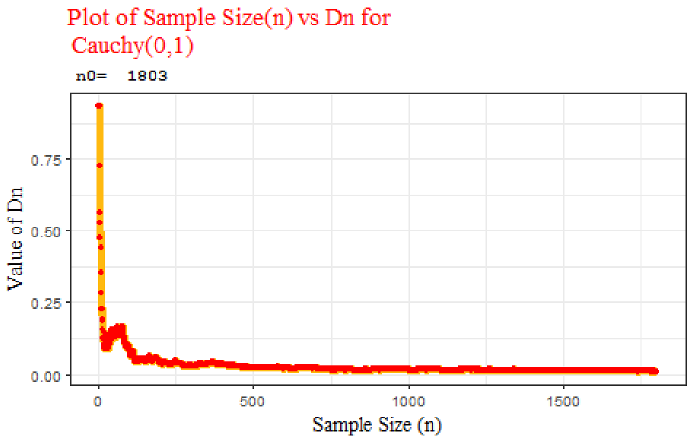
Convergence of D_n for Cauchy :

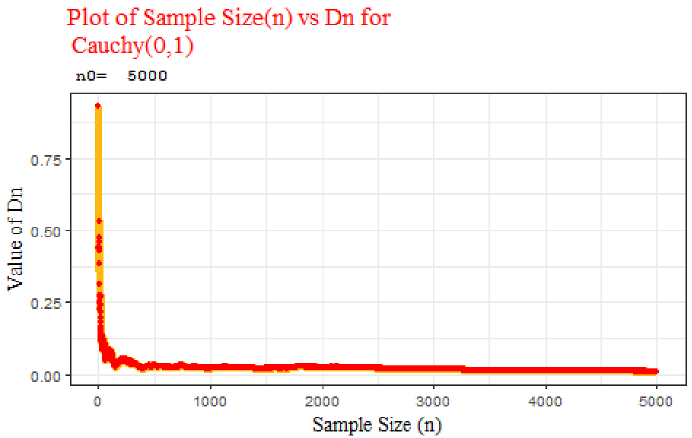




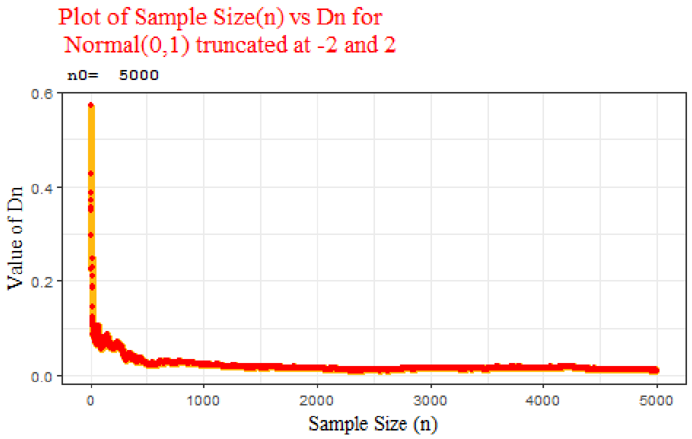


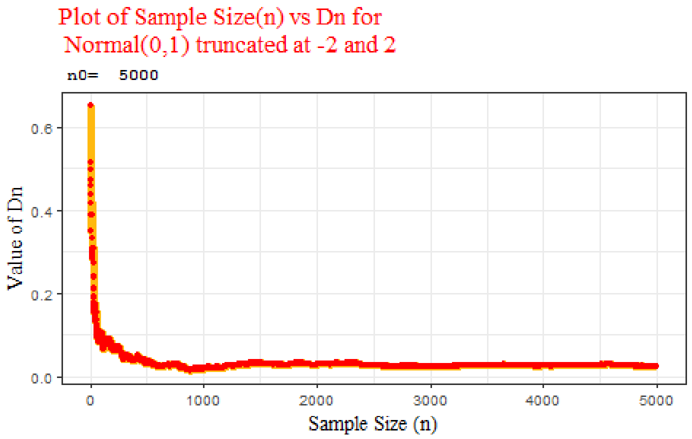


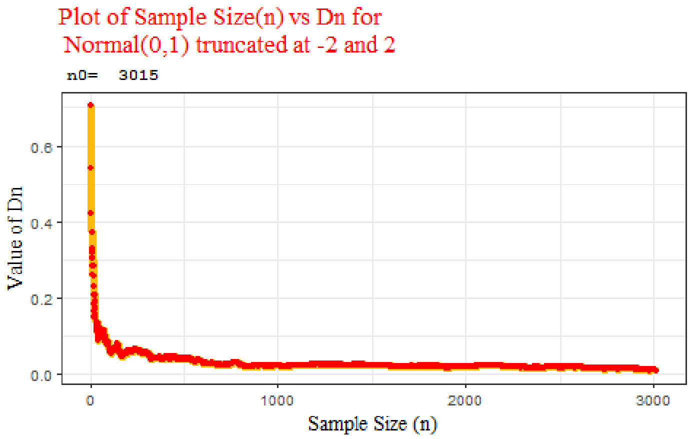


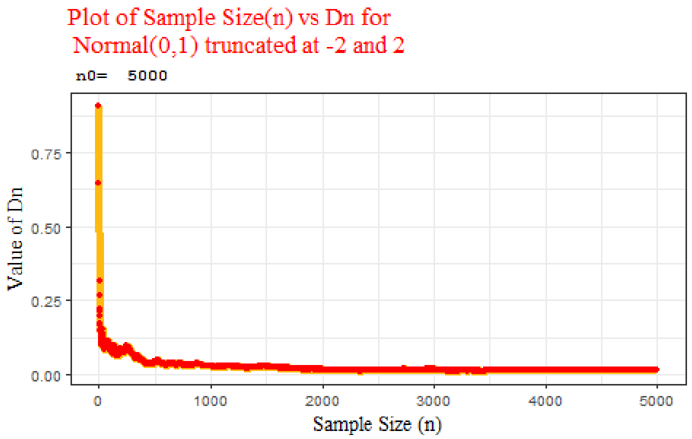


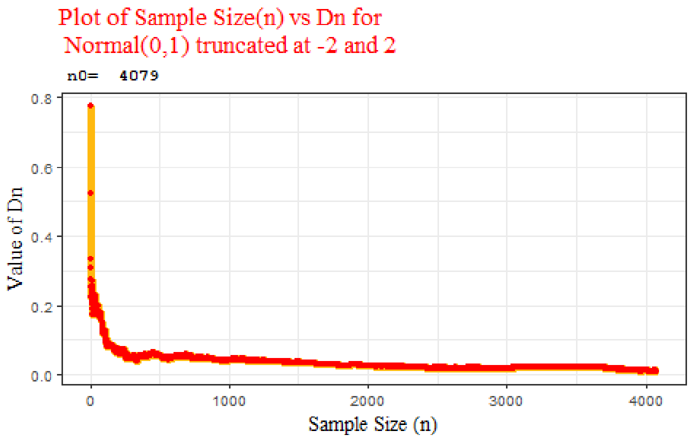
Convergence of D_n for Truncated Normal :

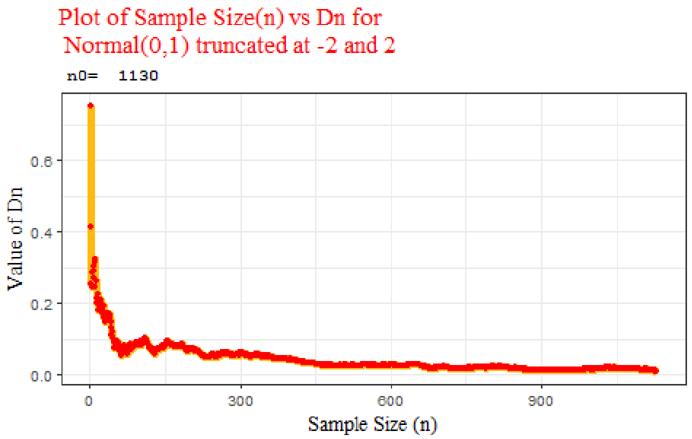




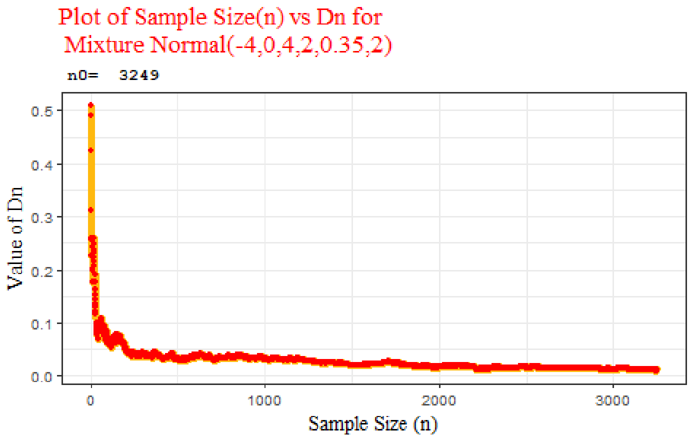


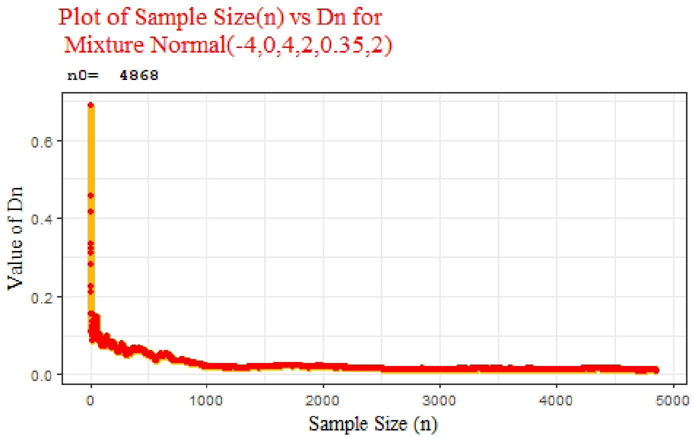


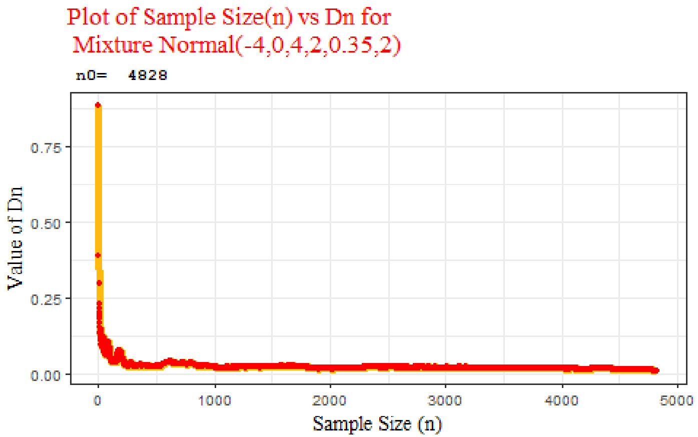


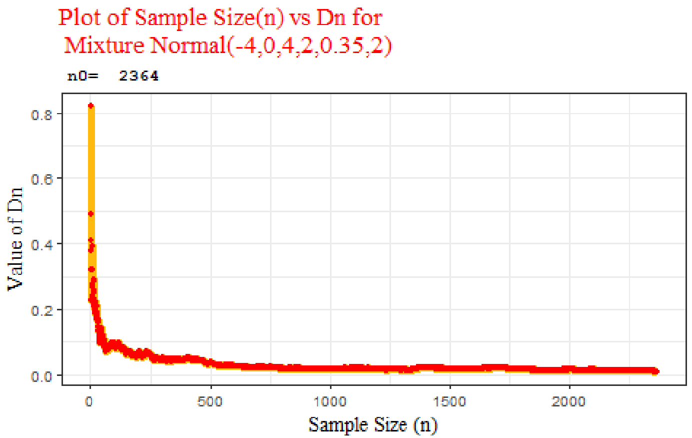


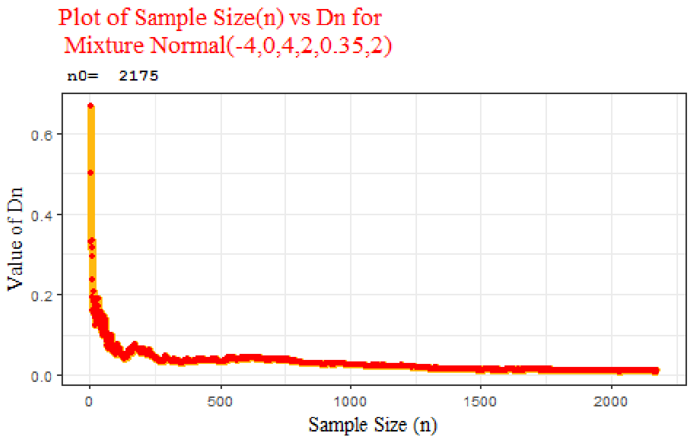
Convergence of D_n for Mixture Normal Distribution :

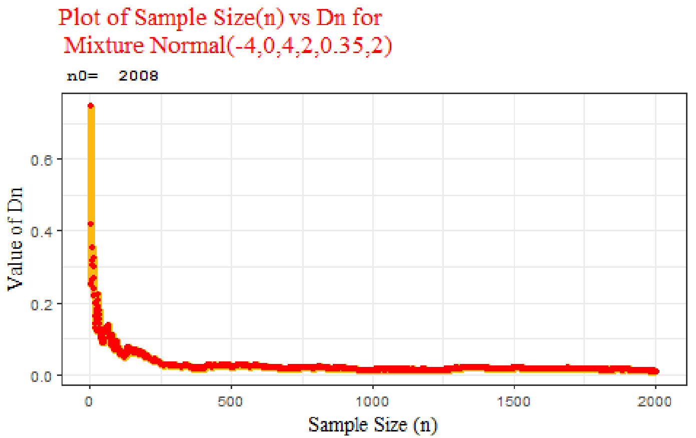




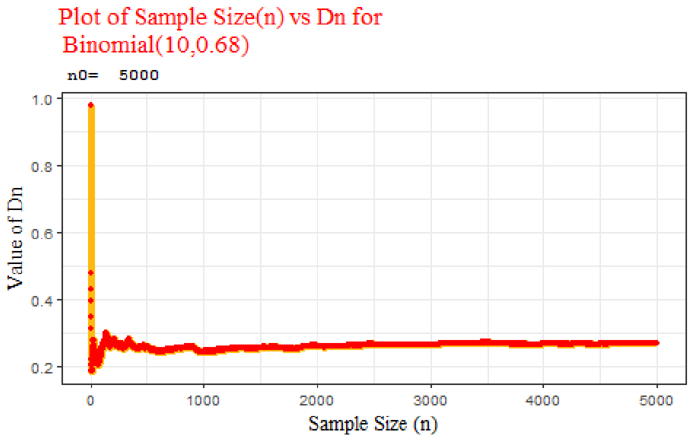


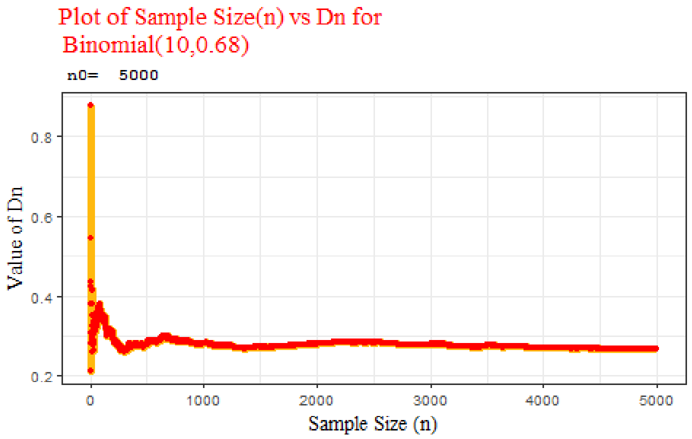


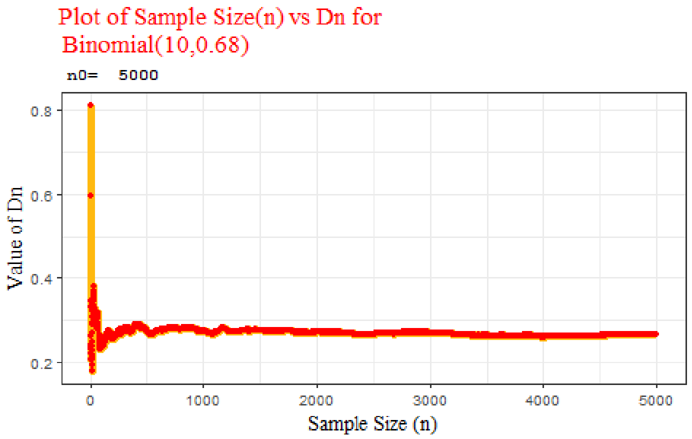


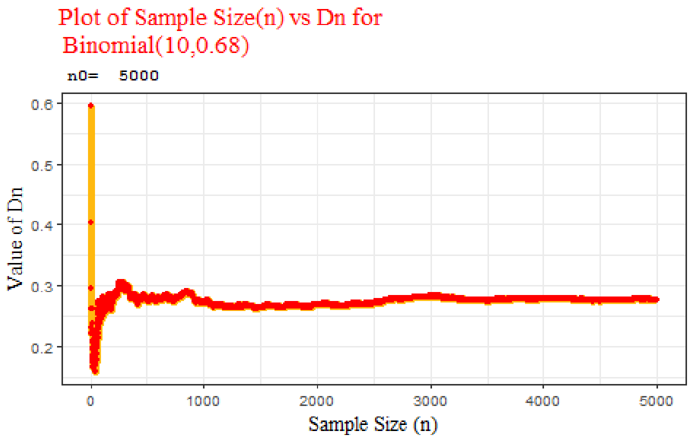


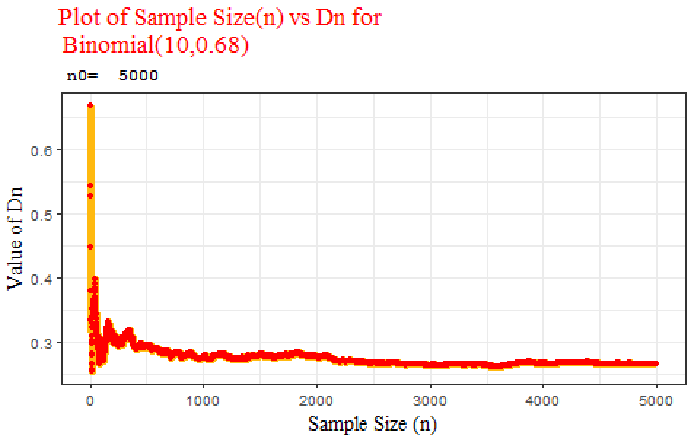
Convergence of D_n for Binomial Distribution :

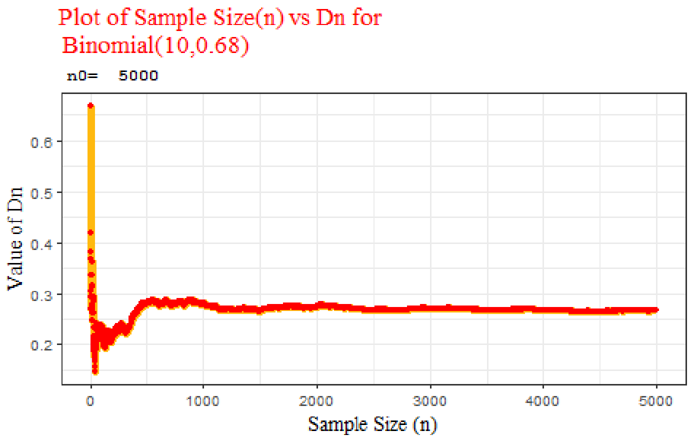




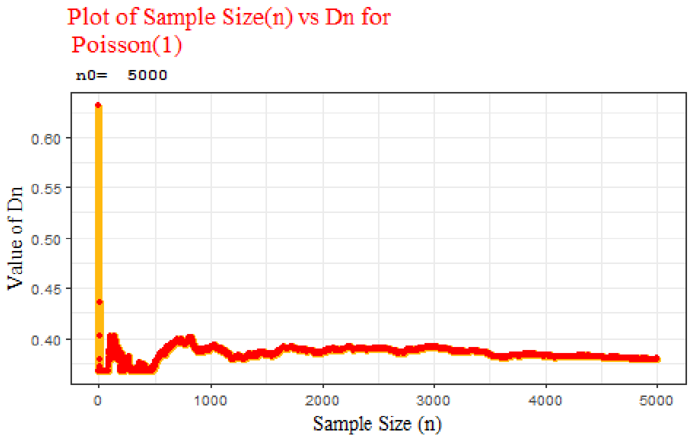


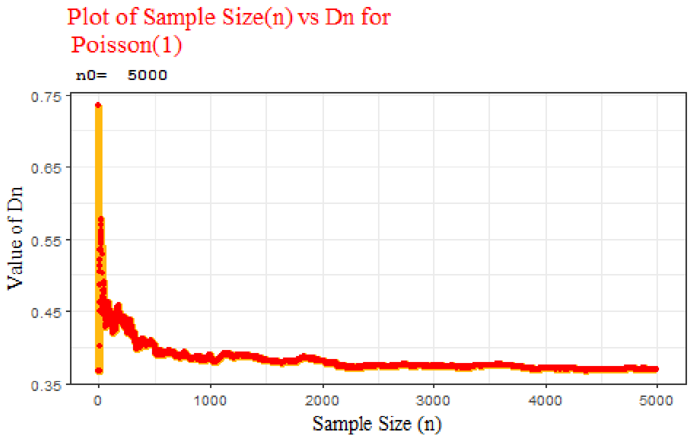


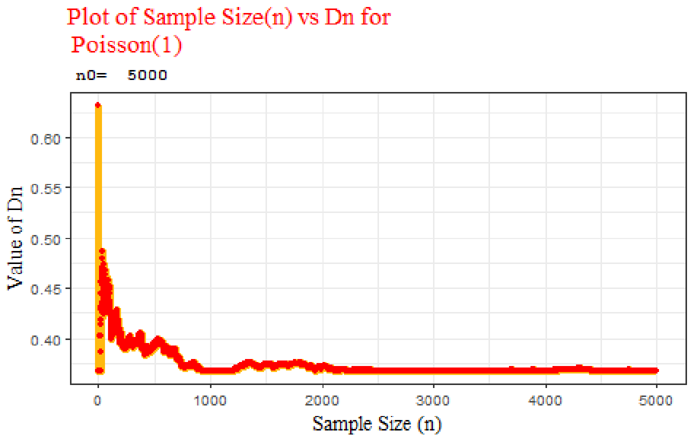


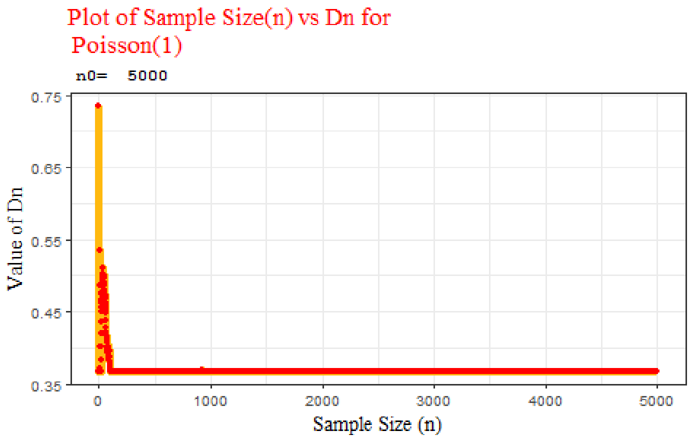


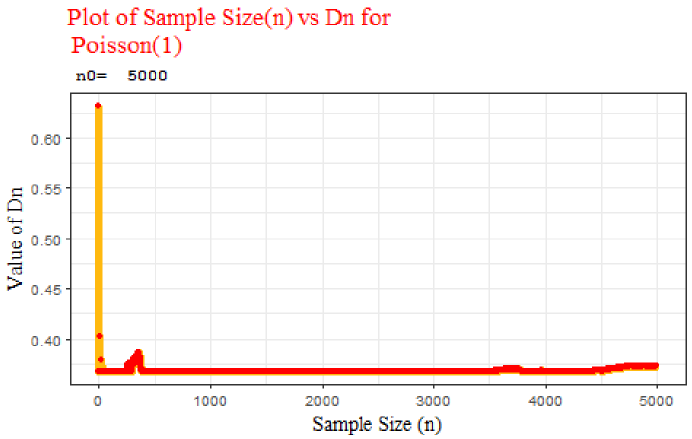
Convergence of D_n for Poisson Distribution :

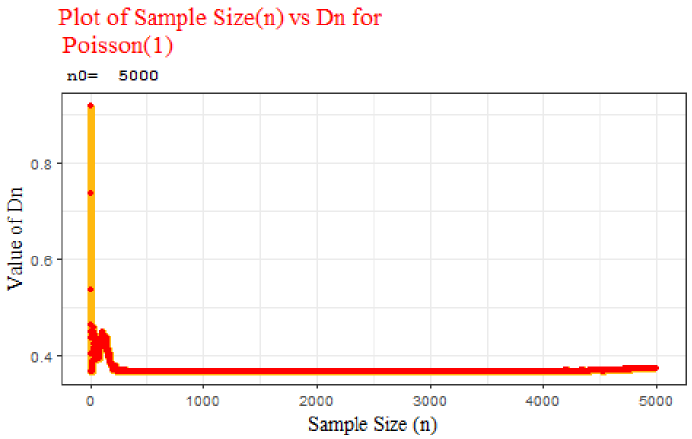






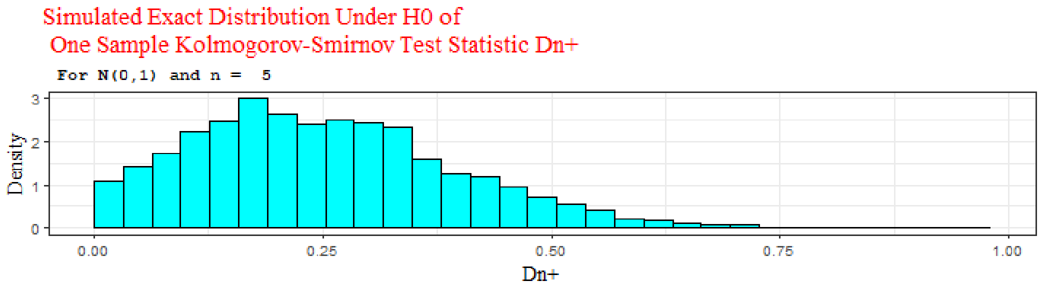


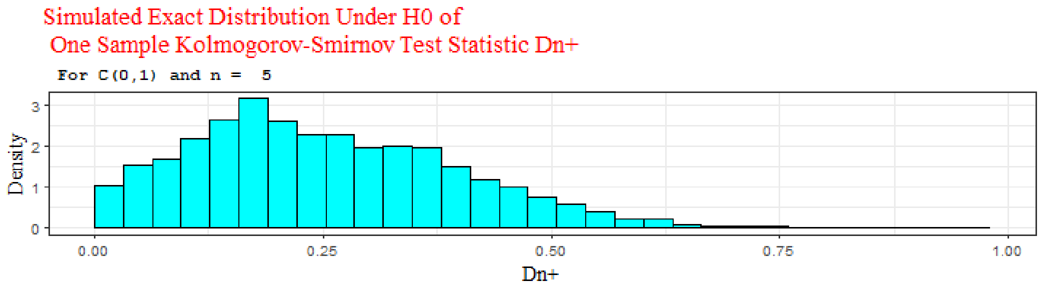


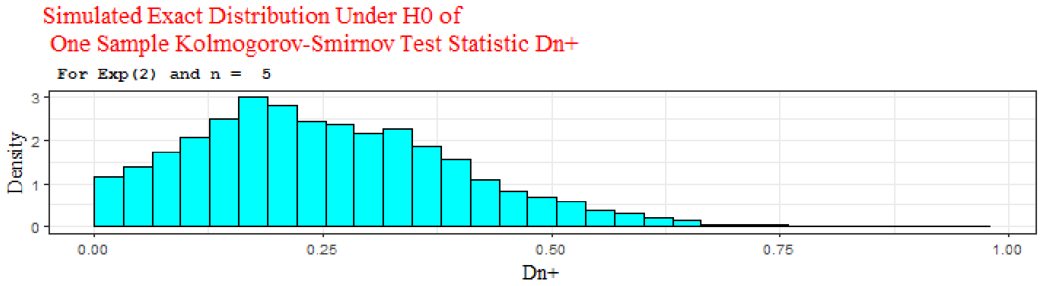


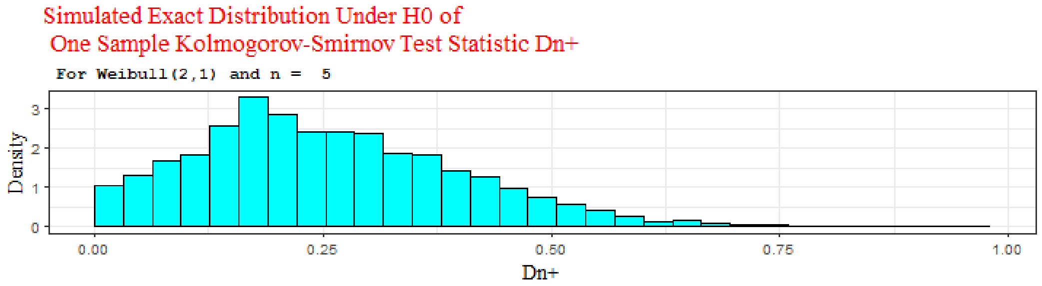
Simulated Exact Distribution of D_n^+ :

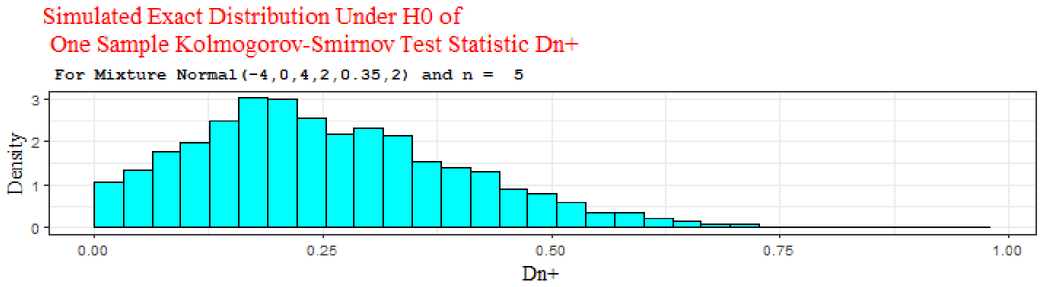
- For Sample Size n=5

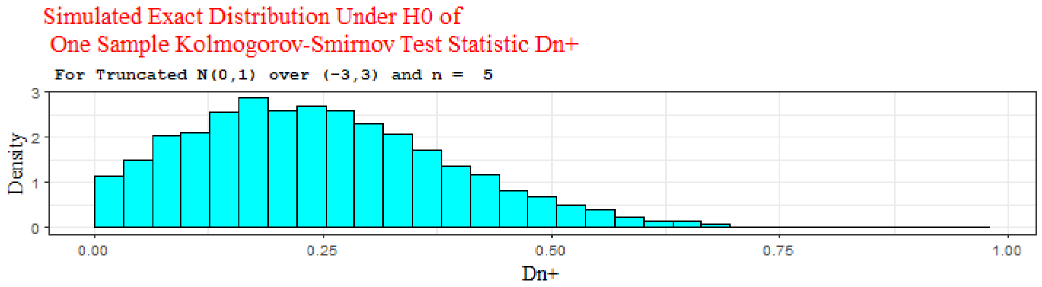




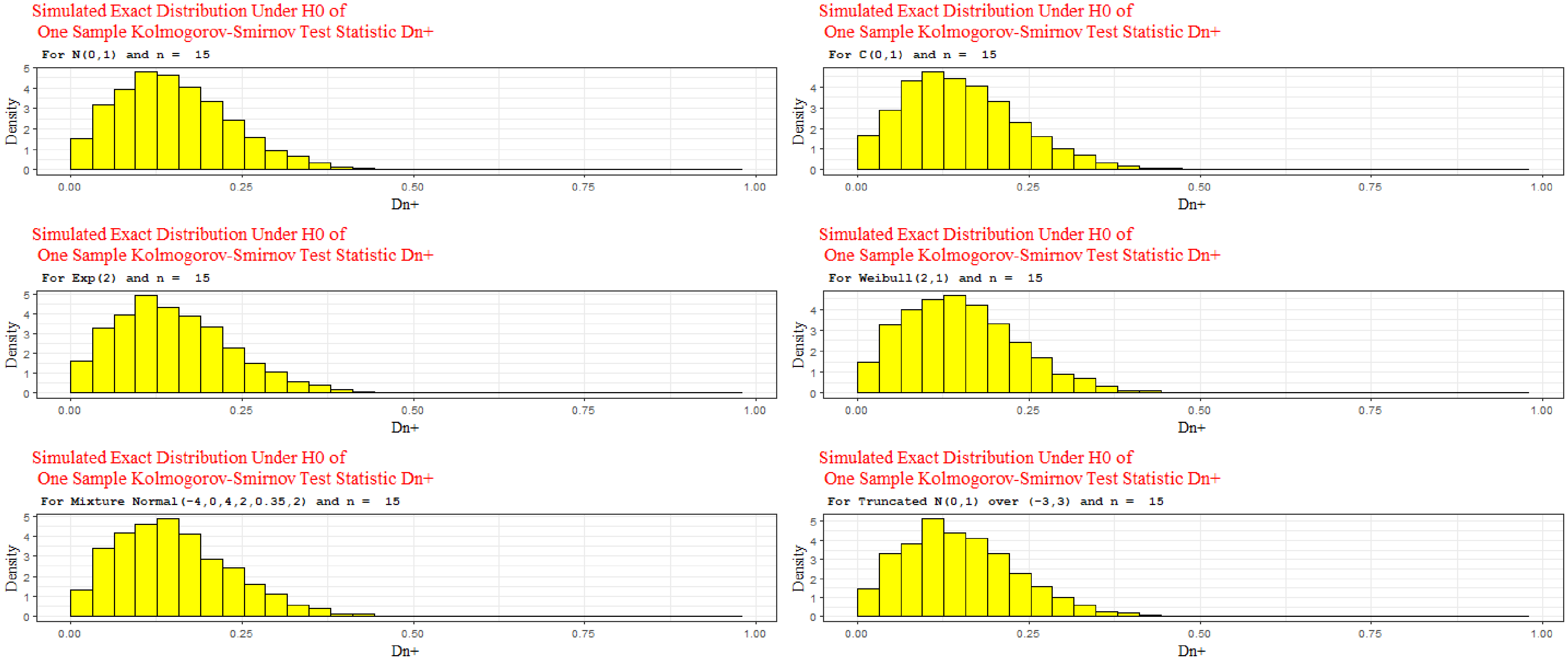








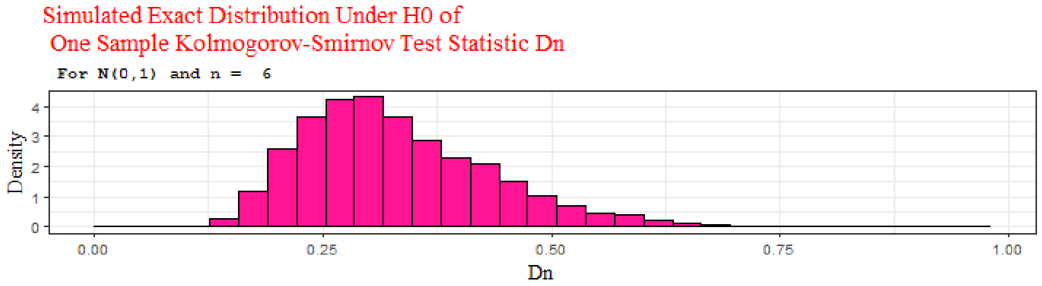
- For Sample Size n=15

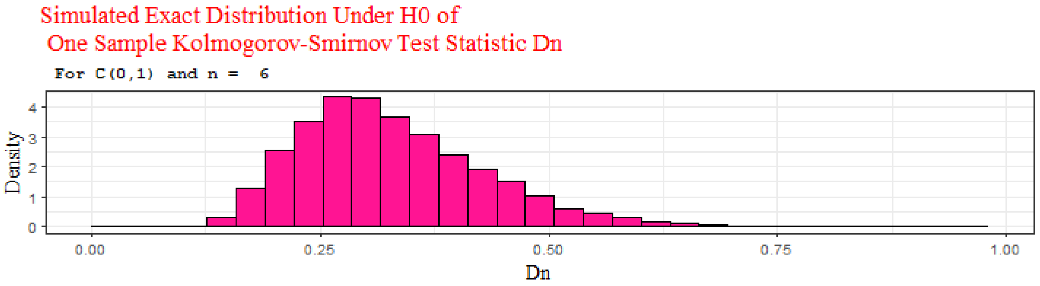


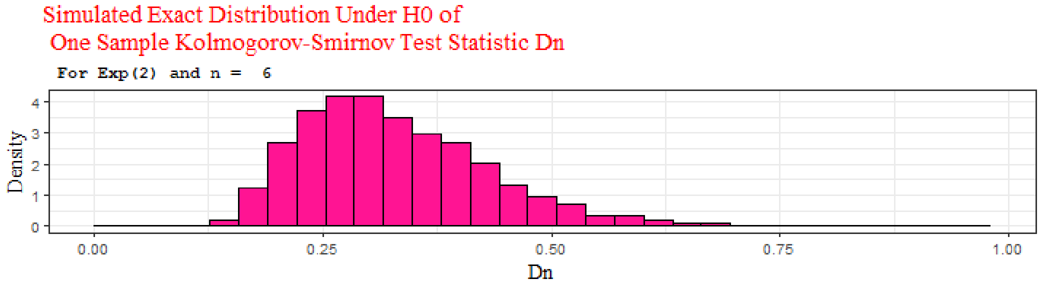
- Empirically, we verified that exact distribution of D_n^+ is "Distribution-Free" under Continuous Parent Population.

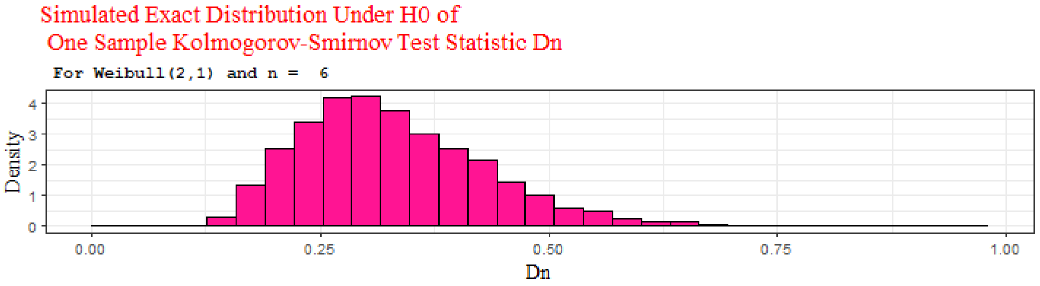
Simulated Exact Distribution of D_n :

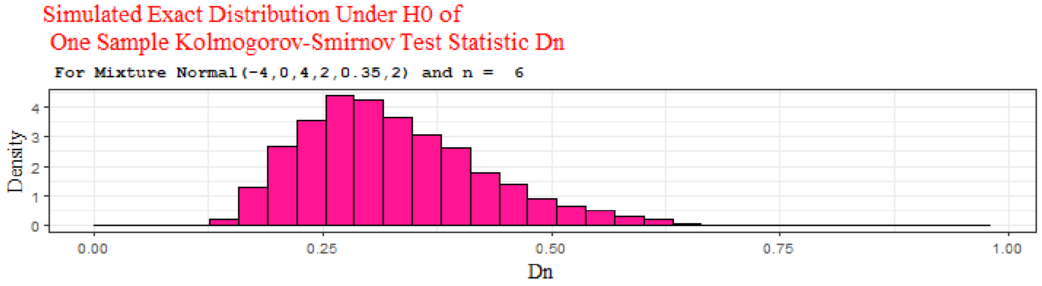
- For Sample Size n=6

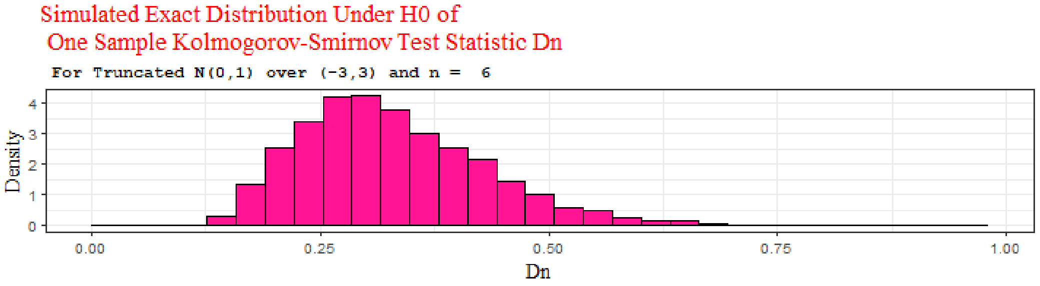




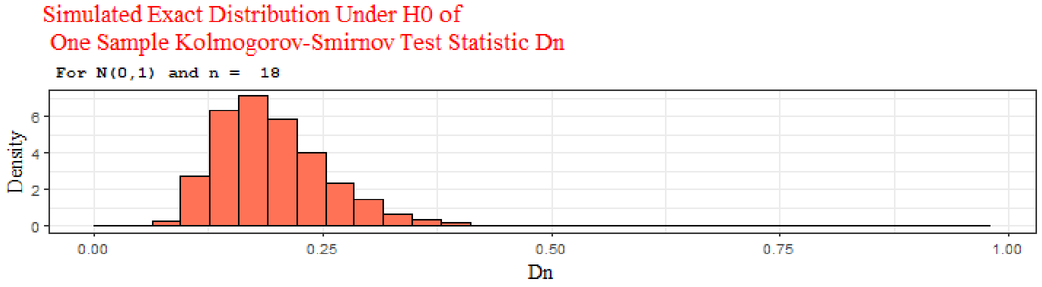


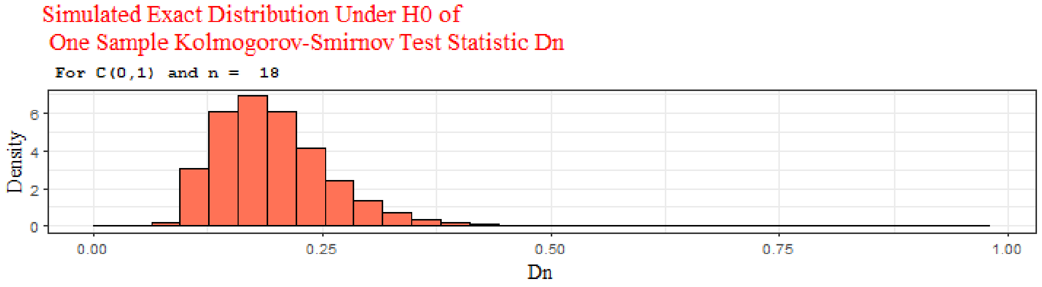


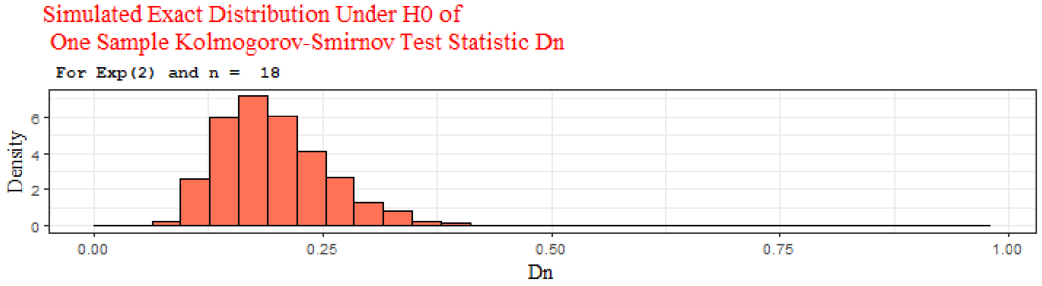


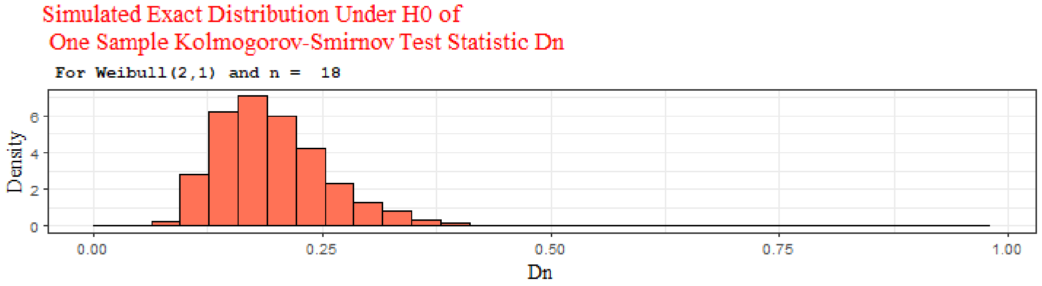


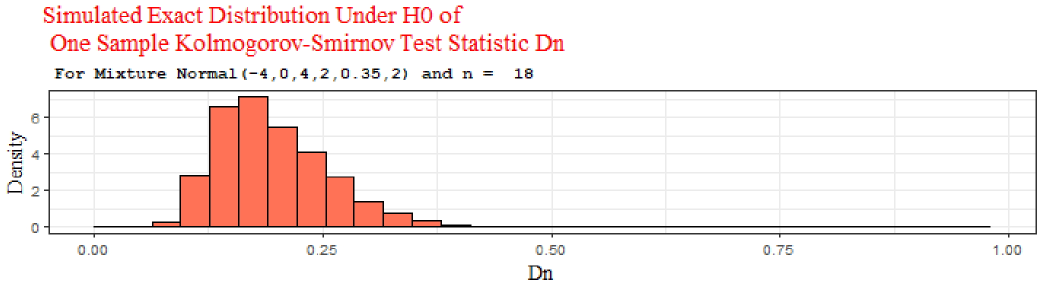
- For Sample Size n = 18

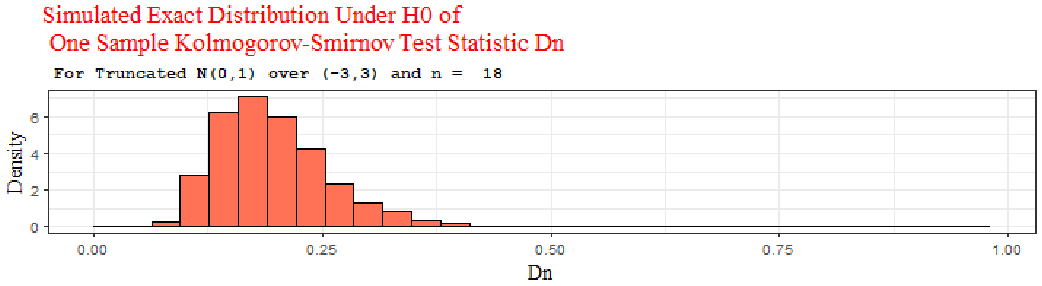








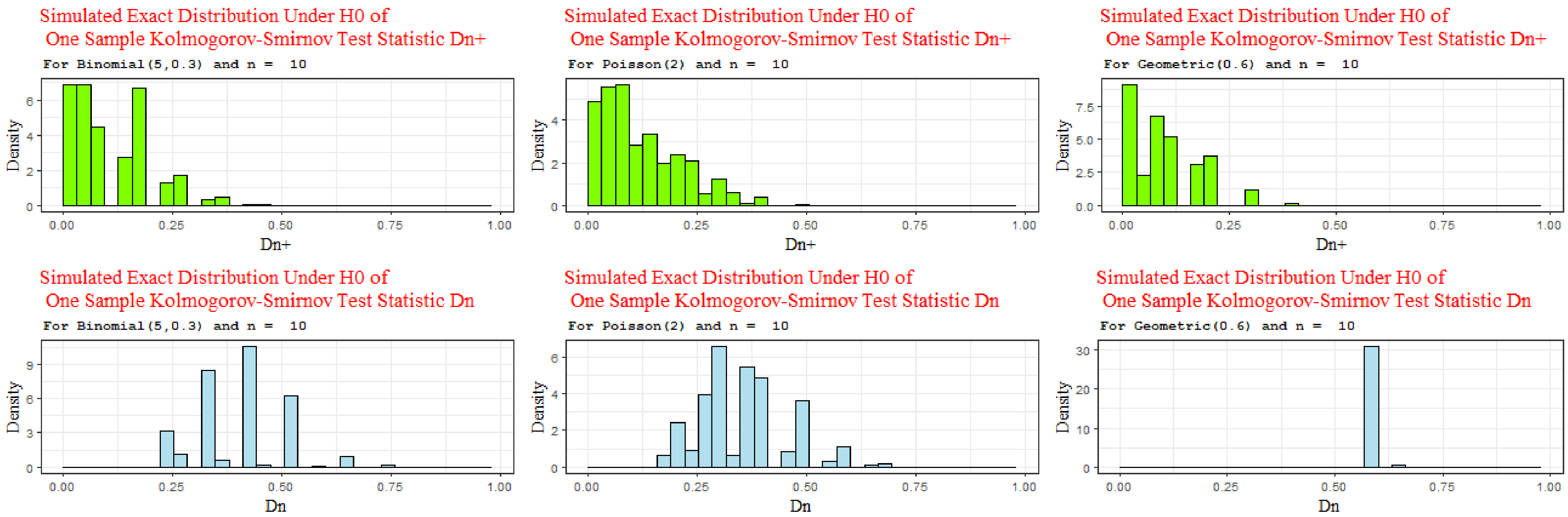




- Empirically, we verified that exact distribution of D_n is "Distribution-Free" under Continuous Parent Population.

What if the Parent Population is Discrete ?

- For Sample Size n = 10



- For Discrete Distribution, there is modified version of Kolmogorov-Smirnov Test Allen, Mark Edward : Kolmogorov-Smirnov test for discrete distributions.

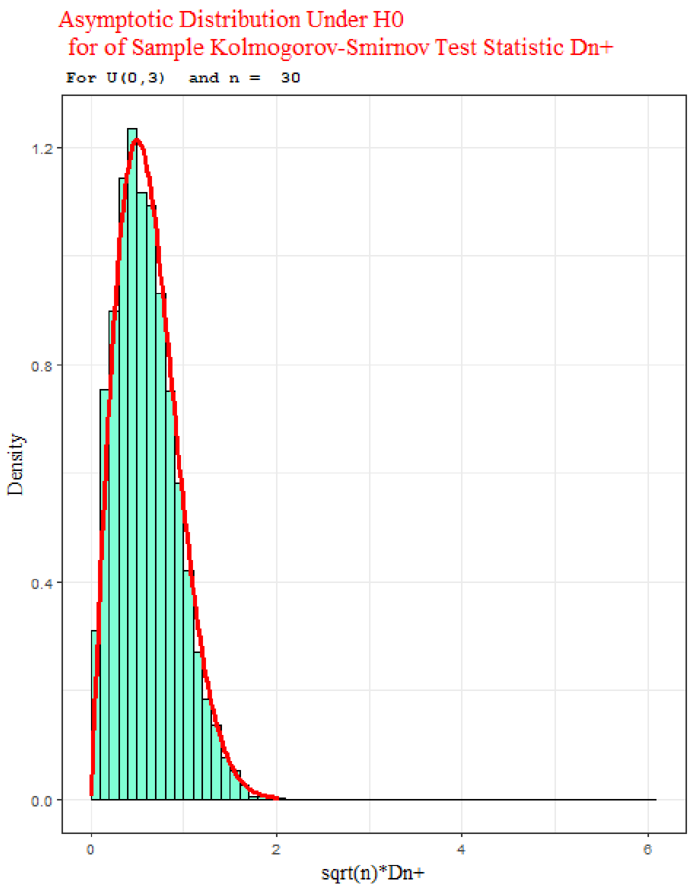
Asymptotic Distribution of D_n^+ and D_n Under H_0 :

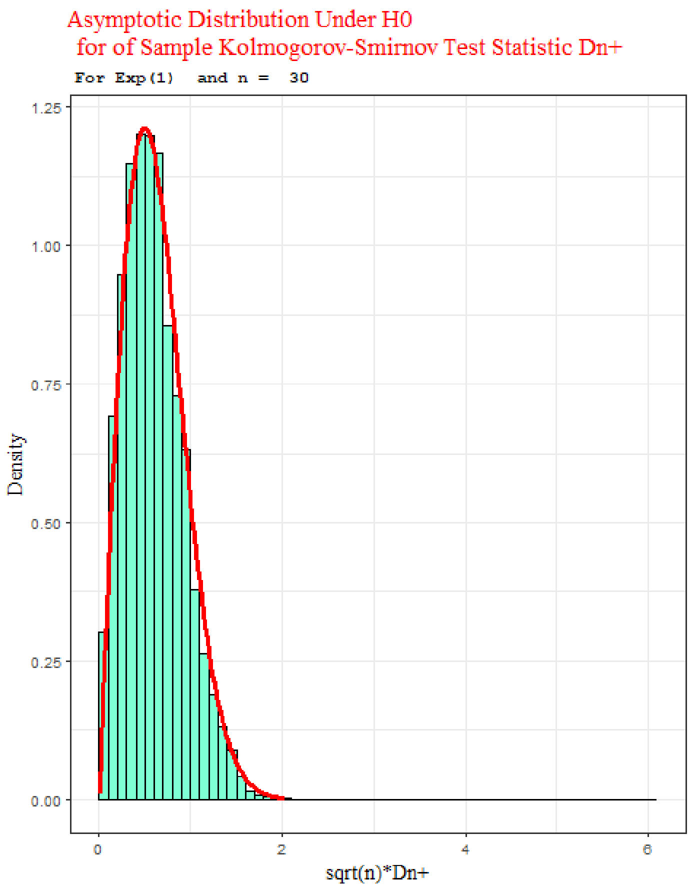
If F is continuous, Then -

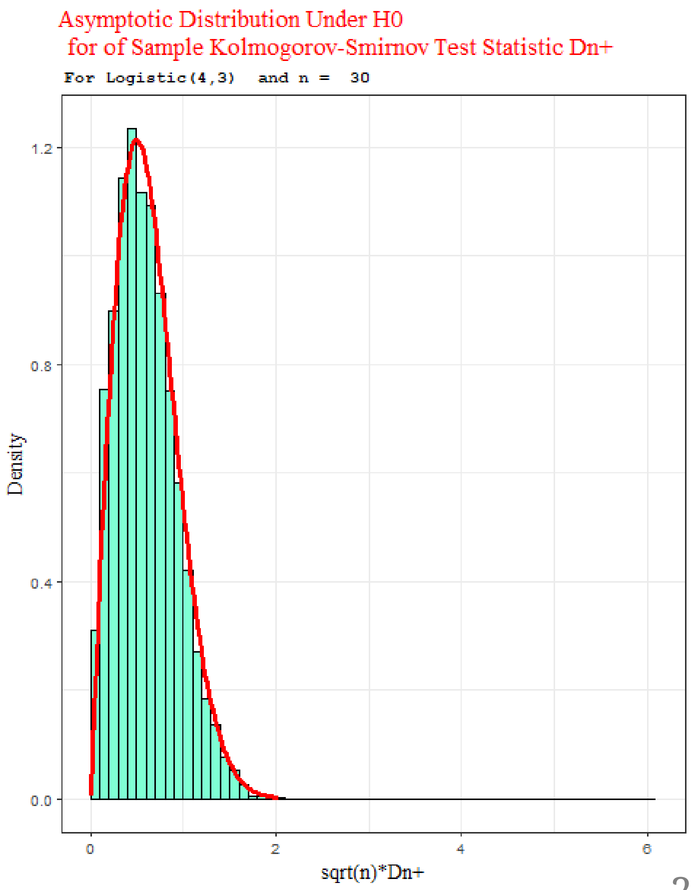
- $\lim\limits_{n\to\infty} P(\sqrt{n}D_n^+ \leq z) = 1 - e^{-2z^2}$ for $z \in R^+$
- $\lim\limits_{n\to\infty} P(\sqrt{n}D_n \leq z) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2z^2}$ for $z \in R^+$
- $V = 4n{D_n}^{+2} \to \chi_2^2$ as $n \to \infty$
- Query : **Are they valid if F is not continuous ?**

Asymptotic Distribution of $\sqrt{n}D_n^+$ under H_0 :

- For Sample Size n = 30

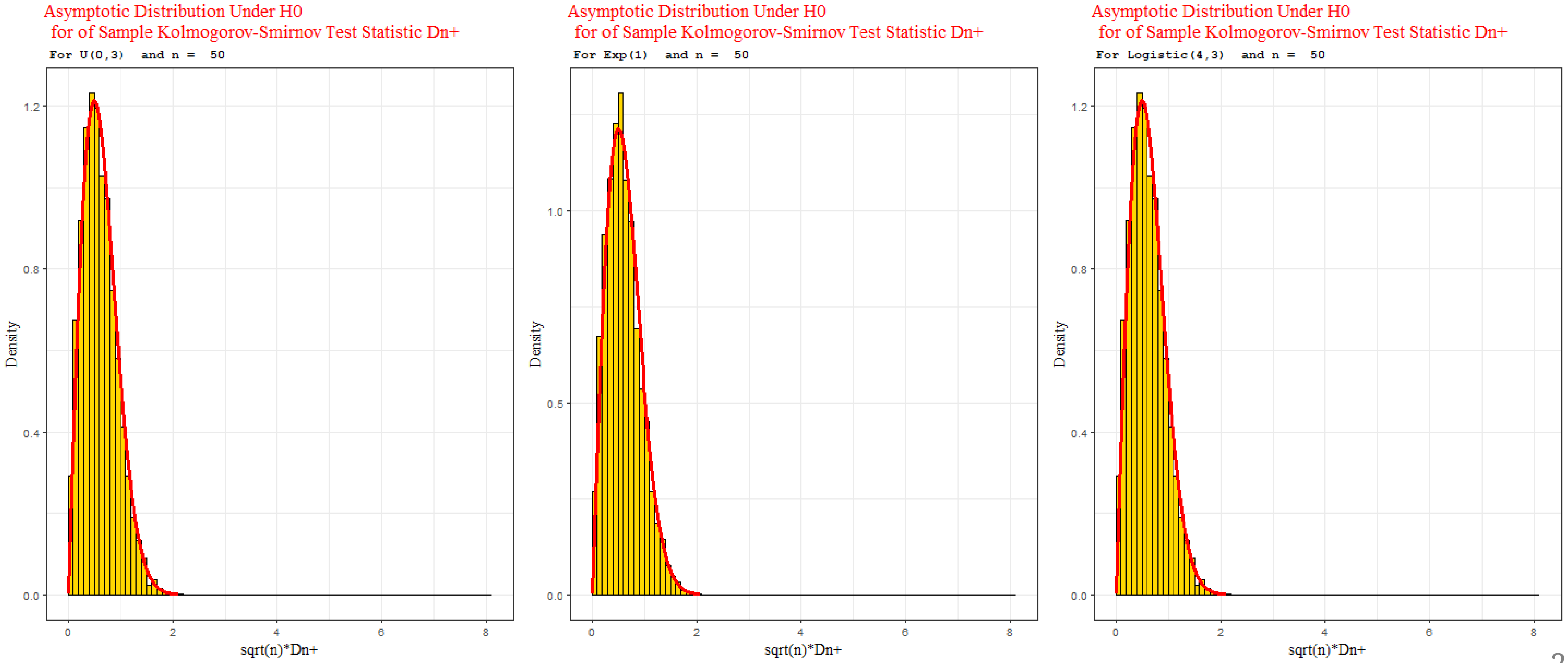




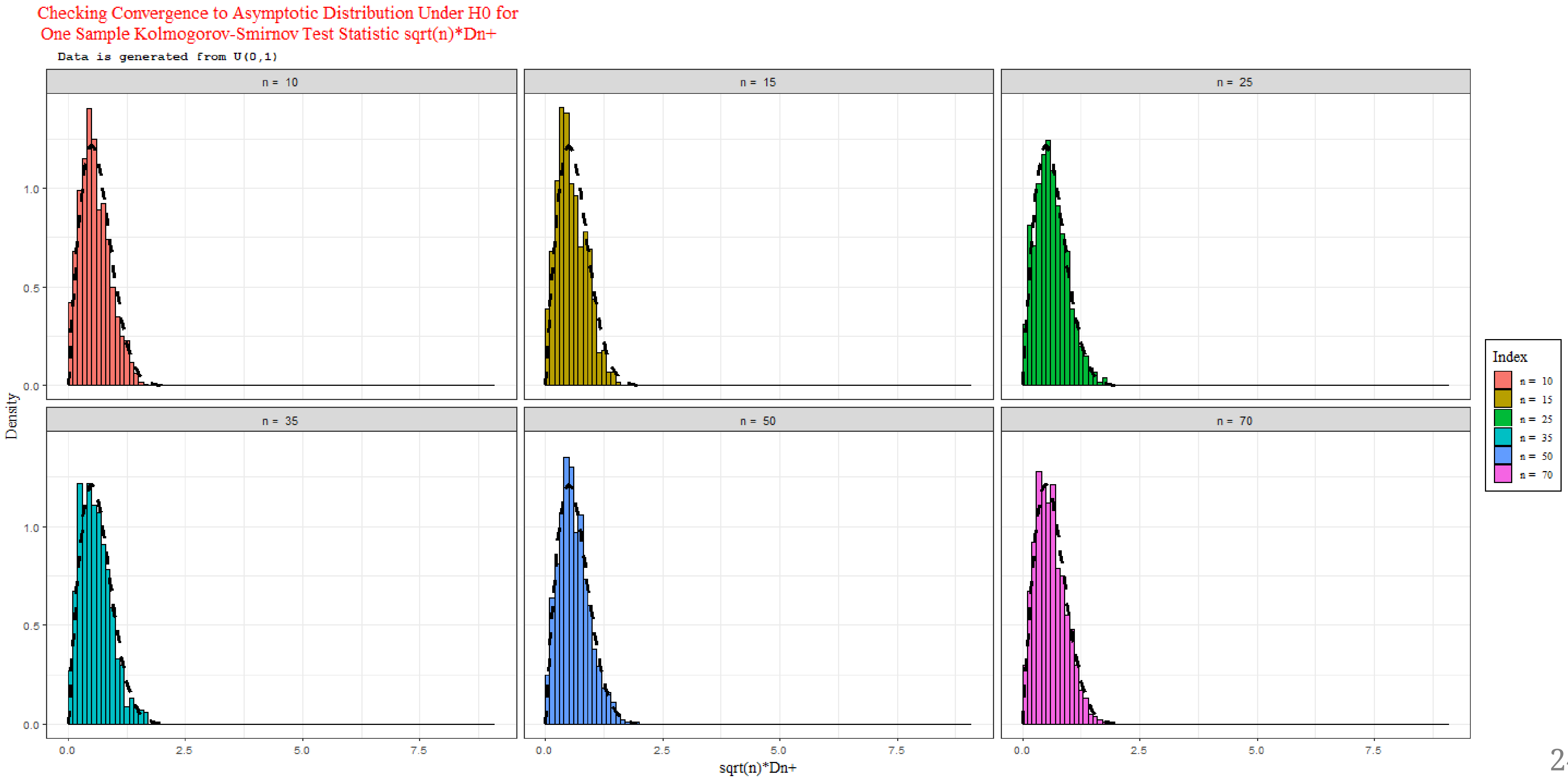


Asymptotic Distribution of $\sqrt{n}D_n^+$ under H_0(Contd.) :

- For Sample Size n = 50

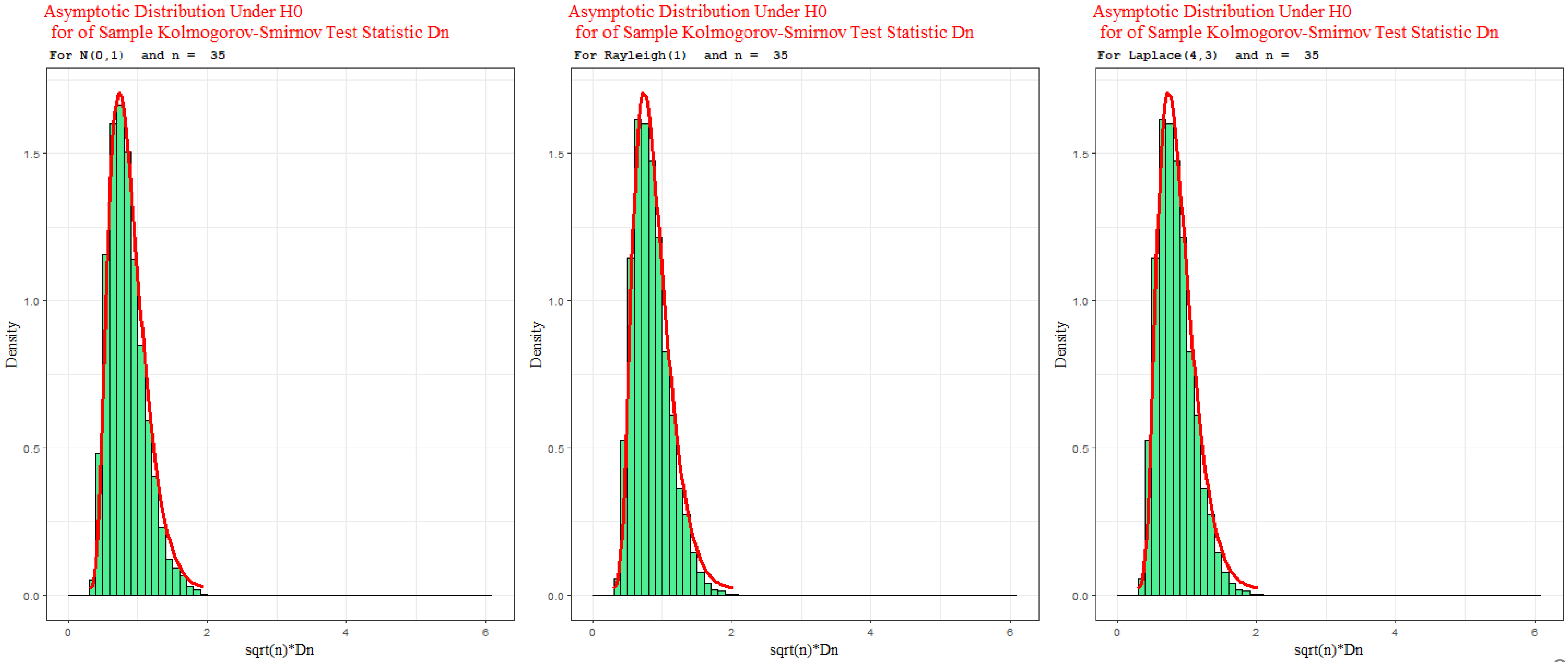


How large is "large" ?



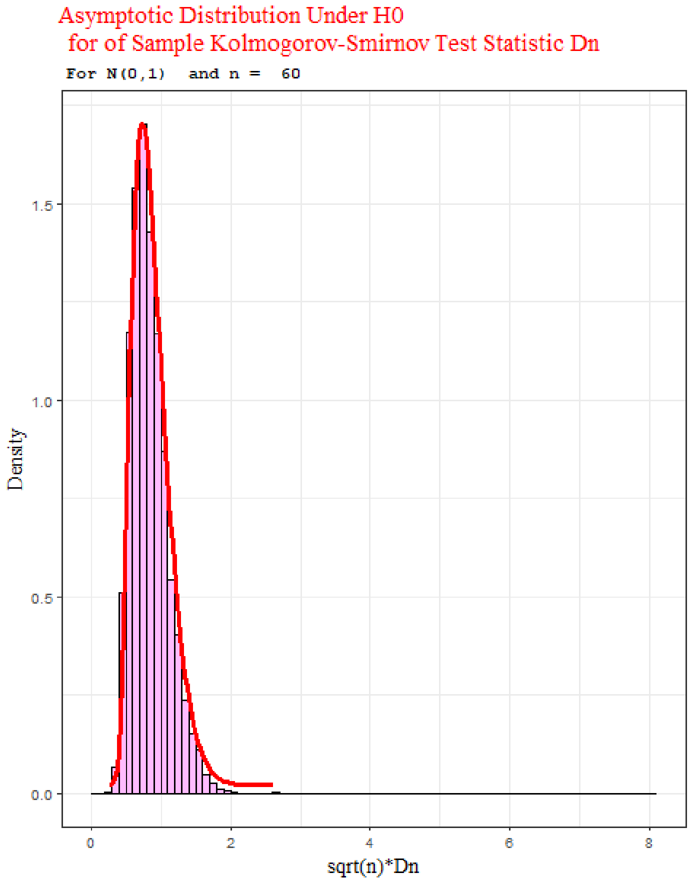
Asymptotic Distribution of $\sqrt{n}D_n$ under H_0 :

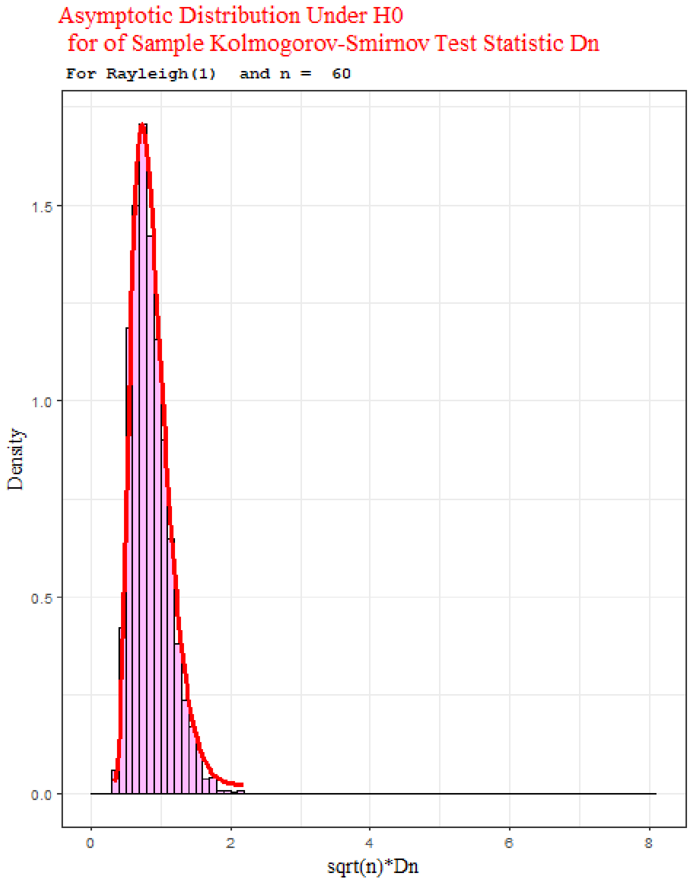
- For Sample Size n = 35

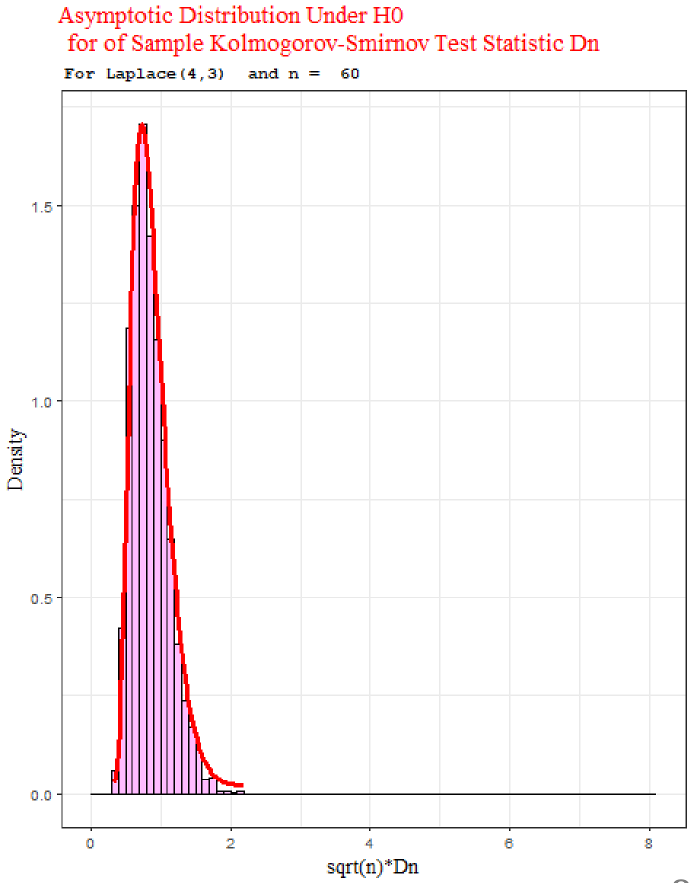


Asymptotic Distribution of $\sqrt{n}D_n$ under H_0 (Contd.) :

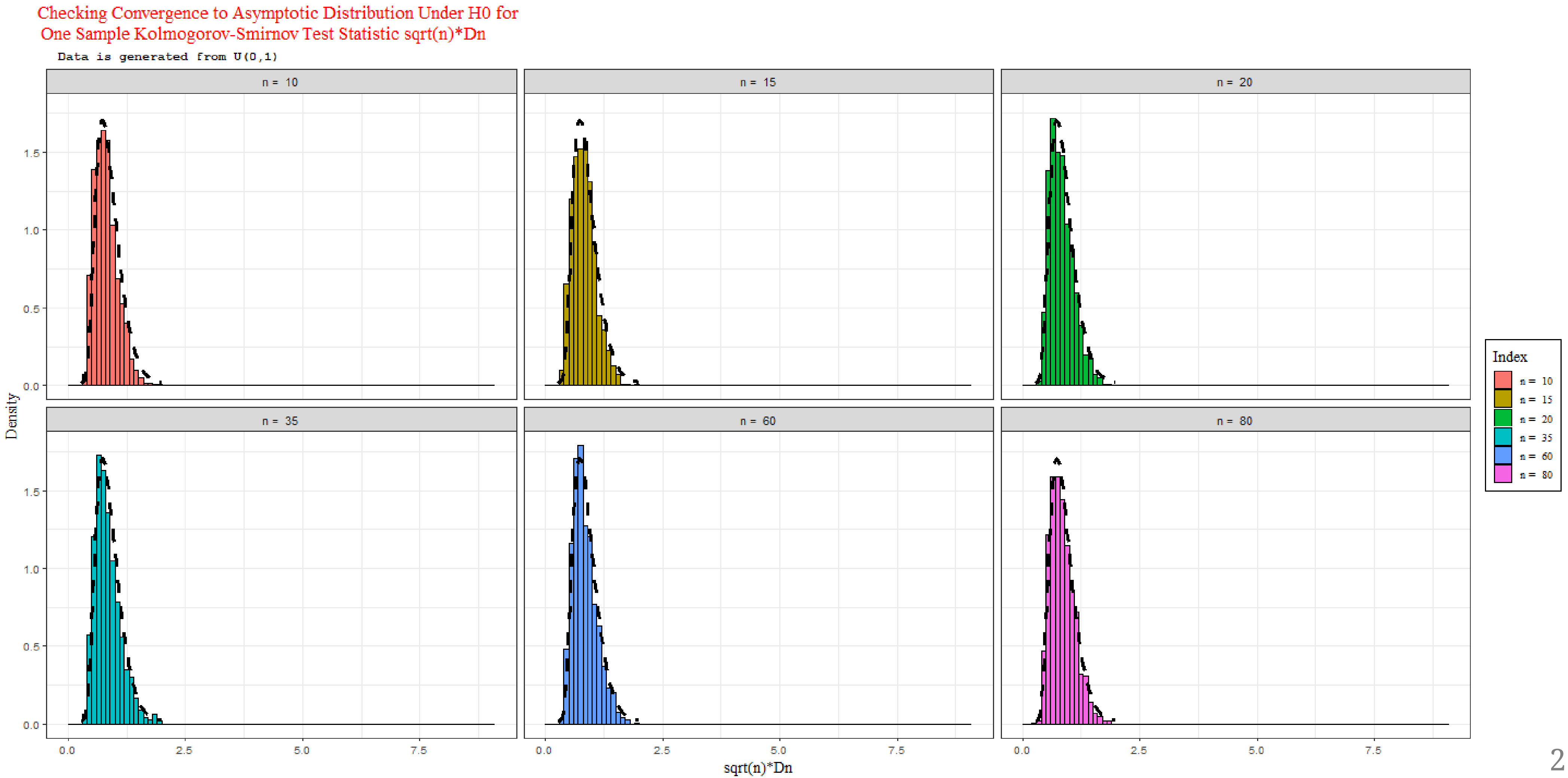
- For Sample Size n = 60







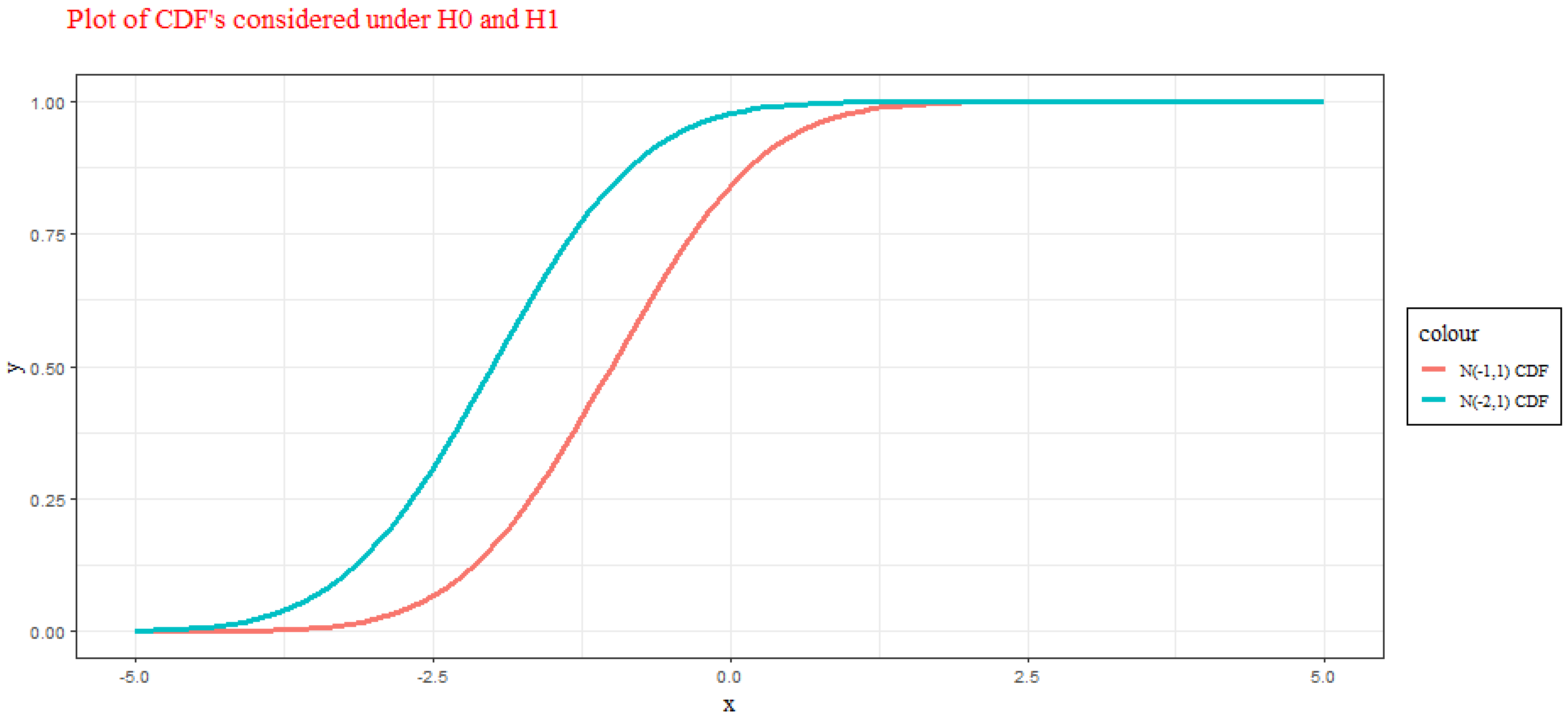
How large is "large" ?



Asymptotic Distribution of $\sqrt{n}D_n^+$ under H1

- To test H_0: F_X = F_0 for all x Vs. H_1: $F_X(x) \geq F_0(x)$ for all x and $F_X(x) > F_0(x)$ with $+ve$ probability.
- We have several choices of alternatives !
- Here, we will consider some particular cases and illustrate them. Later in power comparison we will see more of it.

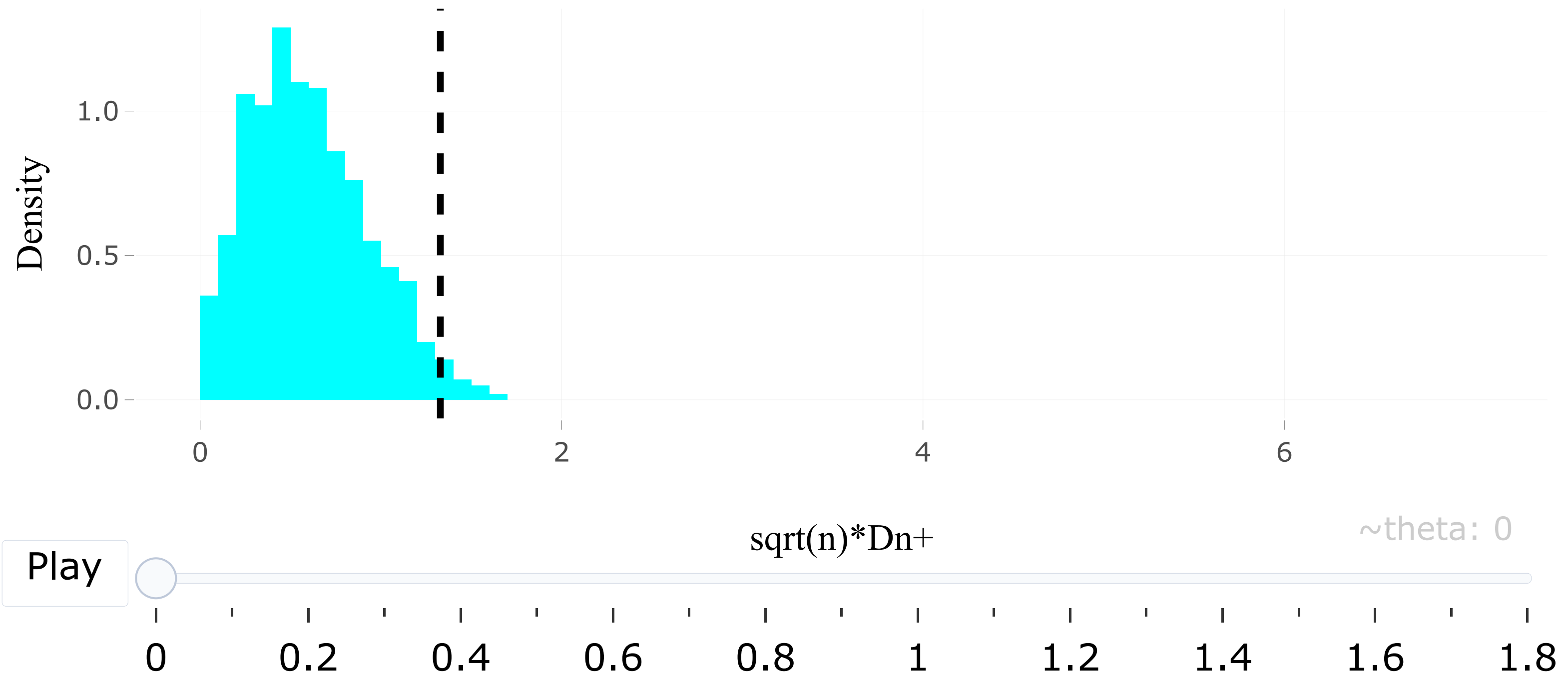
Plot of CDFs for Location Problem :



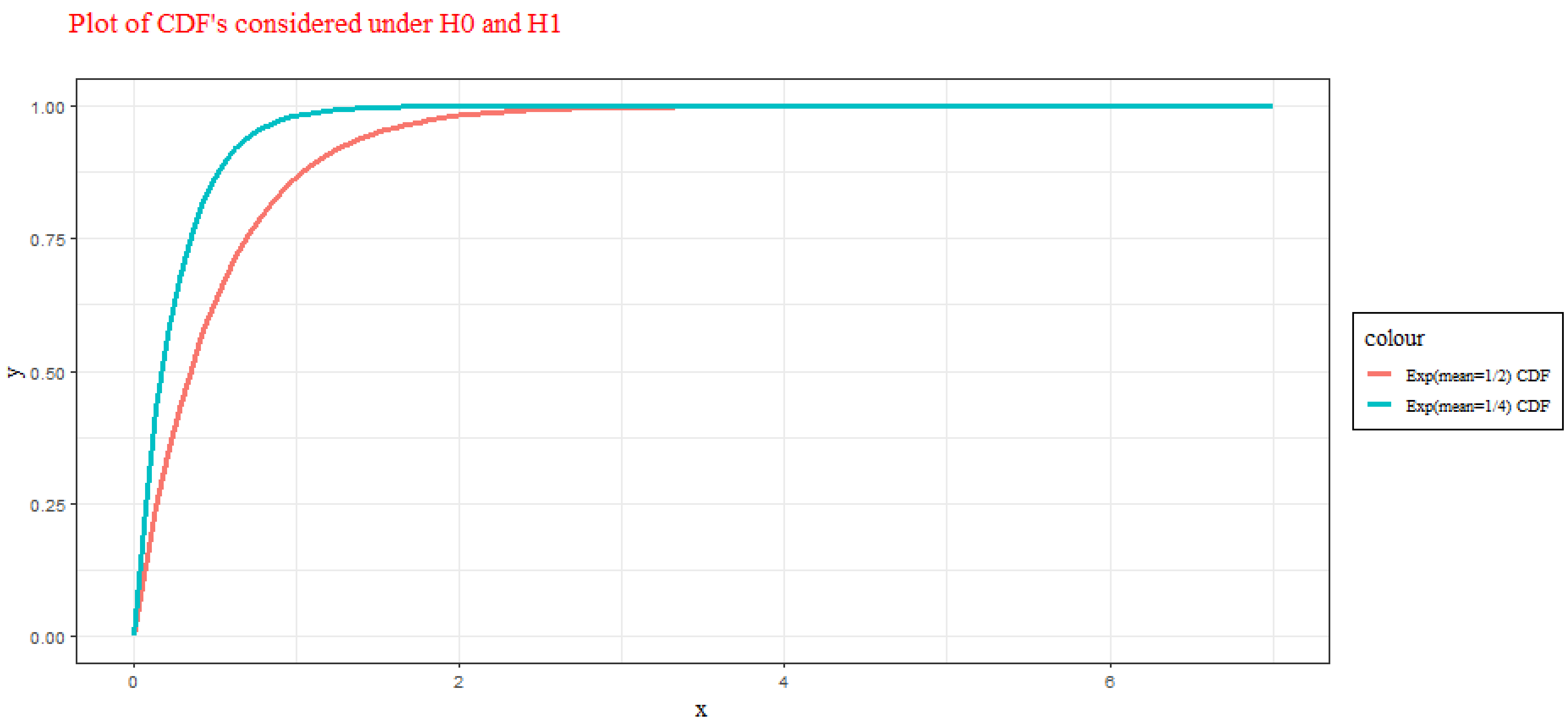
To test H0: X ~ Normal(0,1) vs. H1: X ~ Normal(- μ ,1); μ > 0

- Sample Size(n) = 40

Asymptotic Distribution Under H1 of One Sample Kolmogorov-Smirnov Test Statistic



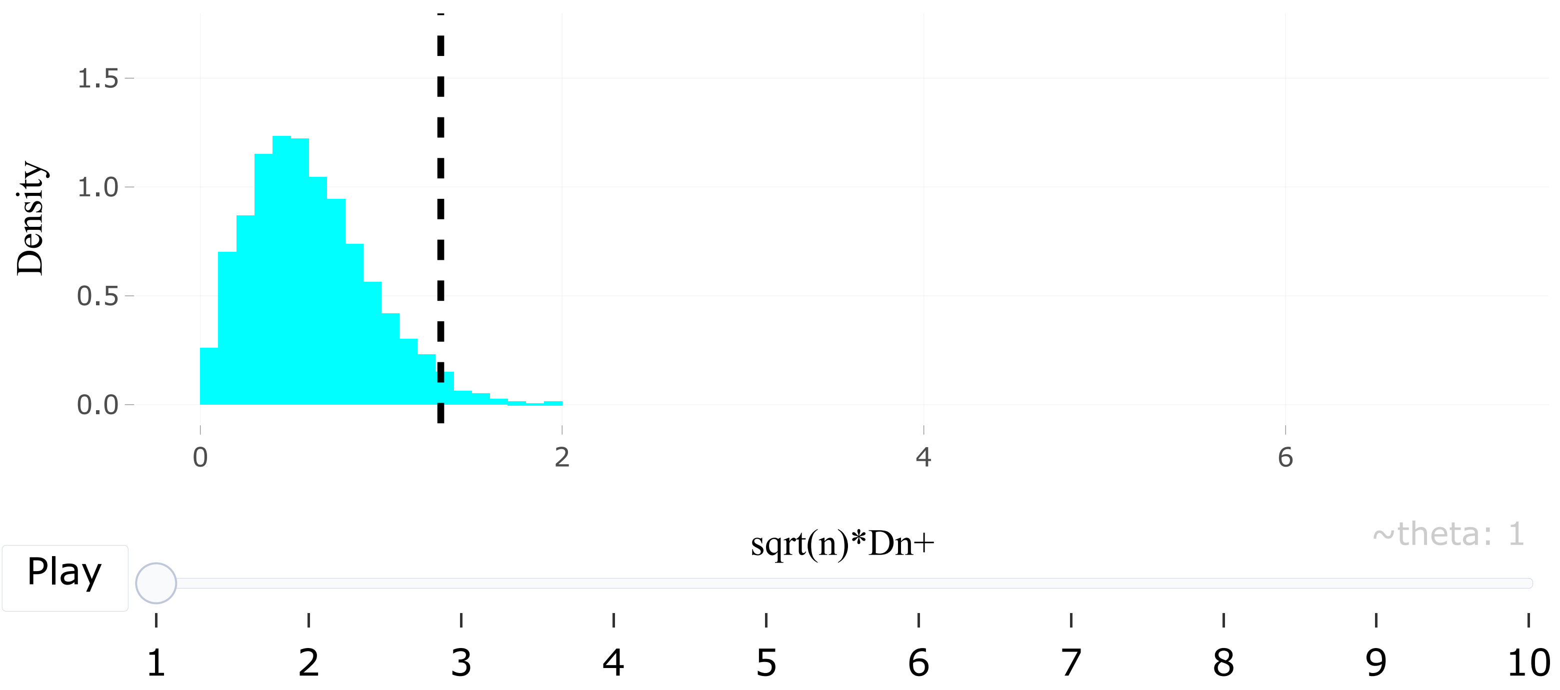
Plot of CDFs for Scale Problem :



To test H0: X ~ Exponential(1) vs. H1: X ~ Exponential(rate = λ); $\lambda > 1$

- Sample Size(n) = 40

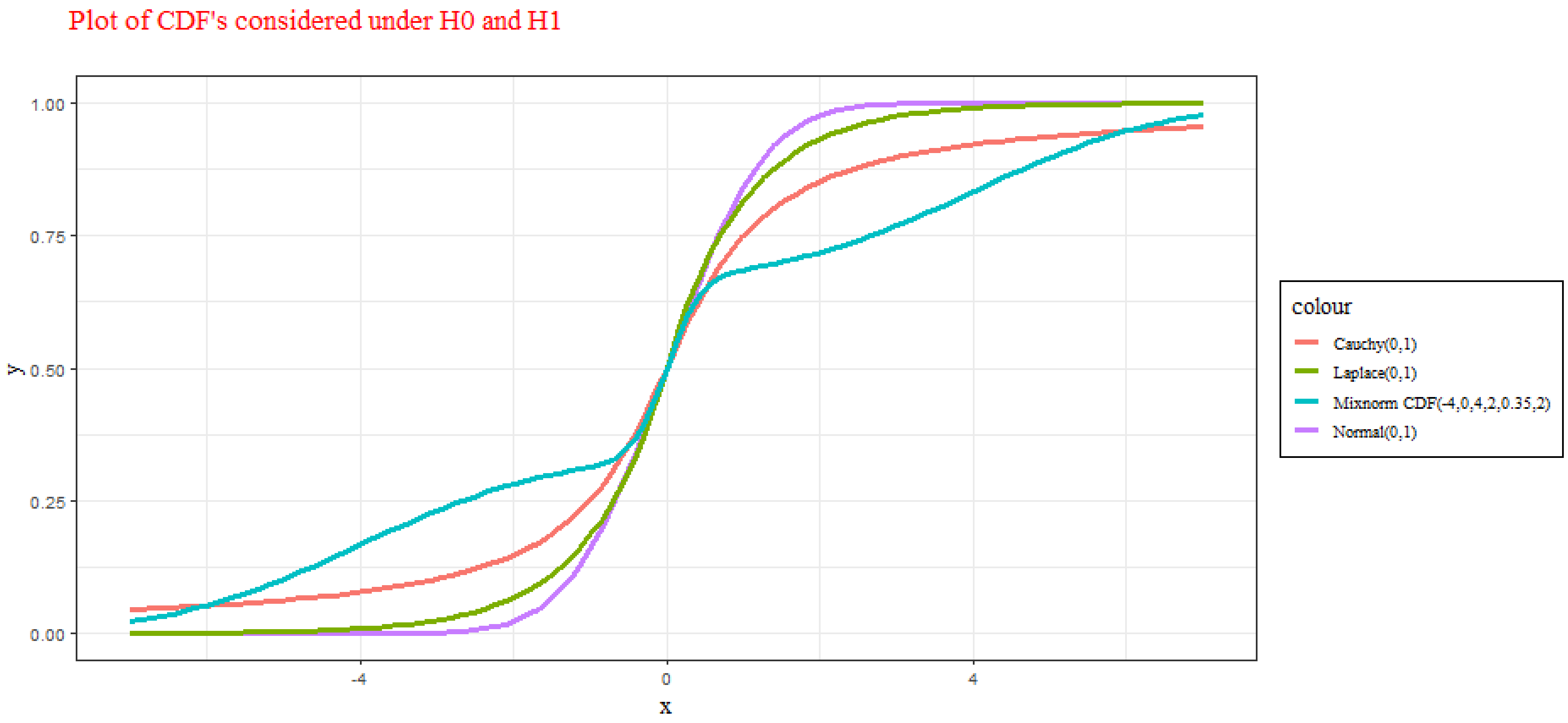
Asymptotic Distribution Under H1 of One Sample Kolmogorov-Smirnov Test Statistic



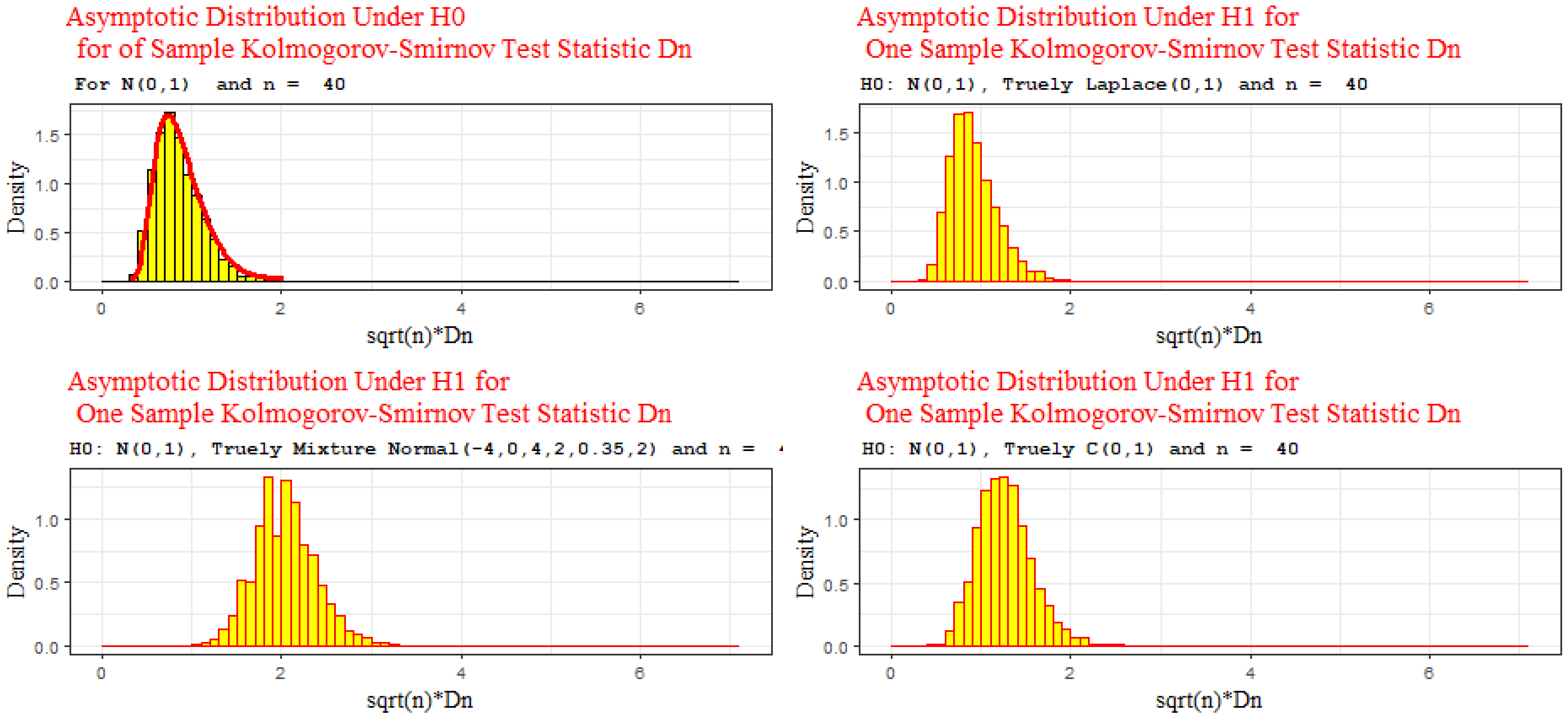
Asymptotic Distribution of $\sqrt{n}D_n$ under H1

- To test H_0: $F_X = F_0$ for all x Vs. H_1: $F_X(x) \neq F_0(x)$ for some x
- Here also, We have several choices of alternatives !

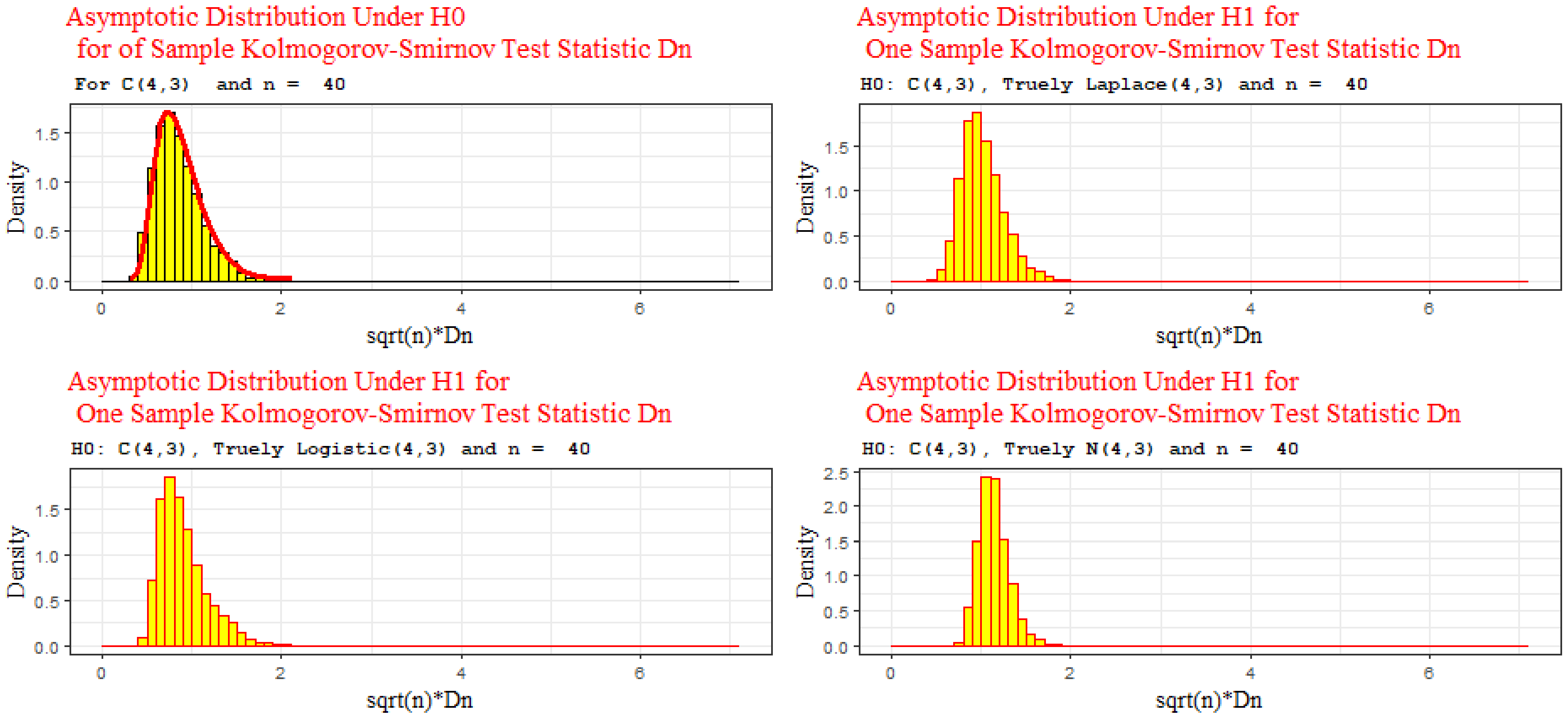
Plot of CDFs For H_0 : X ~ N(0,1) vs. Different Alternatives:



For Testing H_0 : N(0,1) vs different Alternatives



For Testing H_0 : C(4,3) vs different Alternatives



What if the data is Censored?

- We know, if some of observations

$$X_1, X_2, \ldots, X_n$$

  are missing, we say that the data is censored.

- Censoring may happen in 2 ways:

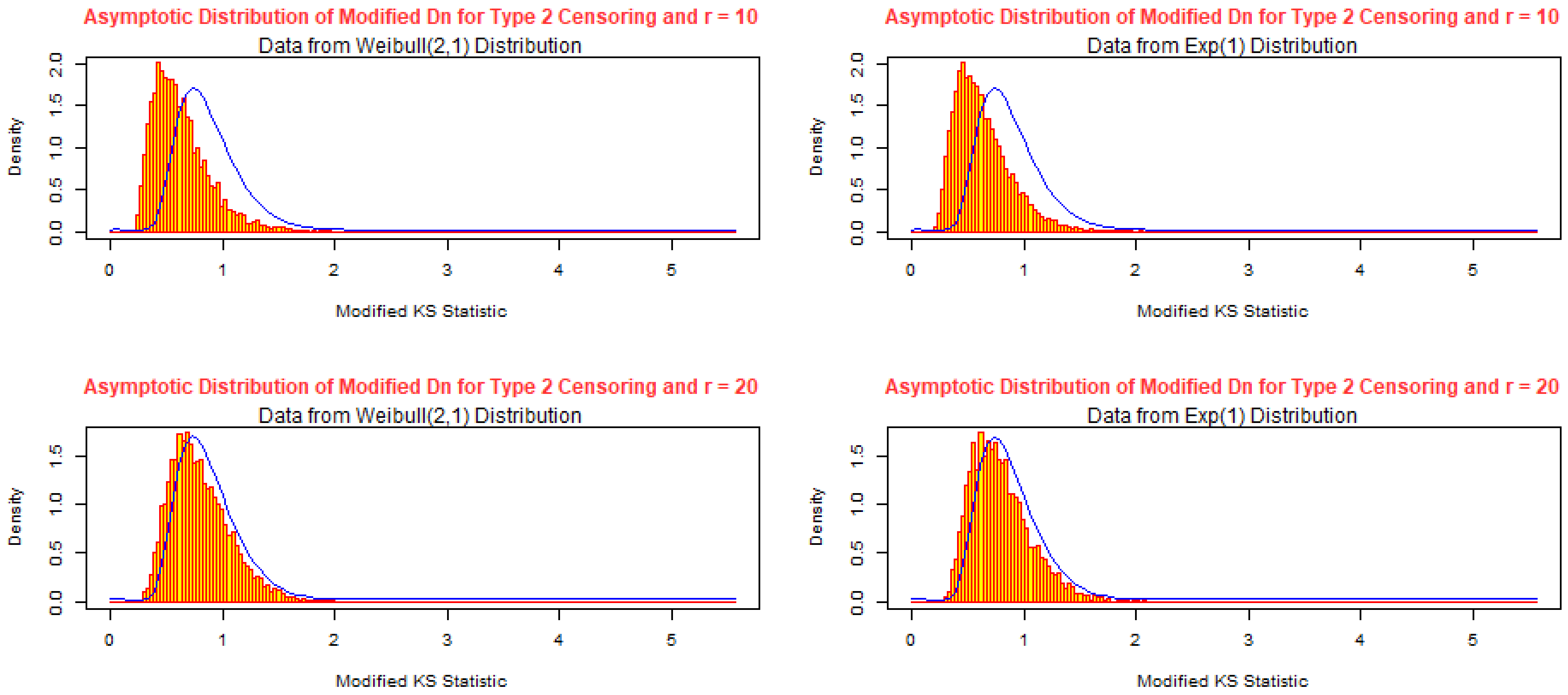
```
1. Type-1 Censoring
2. Type-2 Censoring
```

- If it is decided to follow the experiment say upto r items are failed, is called Type-2 Censoring.
- For Type-2 Censored data, the Kolmogorov-Smirnov statistic for 2-sided test is:

  $_2Dr, n = \sup_{0 \leq z \leq Z_{(r)}} \left|F_n(z) - z\right| = max_{1 \leq i \leq r} \left|\frac{i-0.5}{n} - Z_{(i)}\right| + \frac{0.5}{n}$

- Here, we have simulated distribution of $_2Dr, n$ for different distributions & different values of r.
- More details are available in Ralph B. D'Agostino & Michael A. Stephens : Goodness-Of-Fit-Techniques

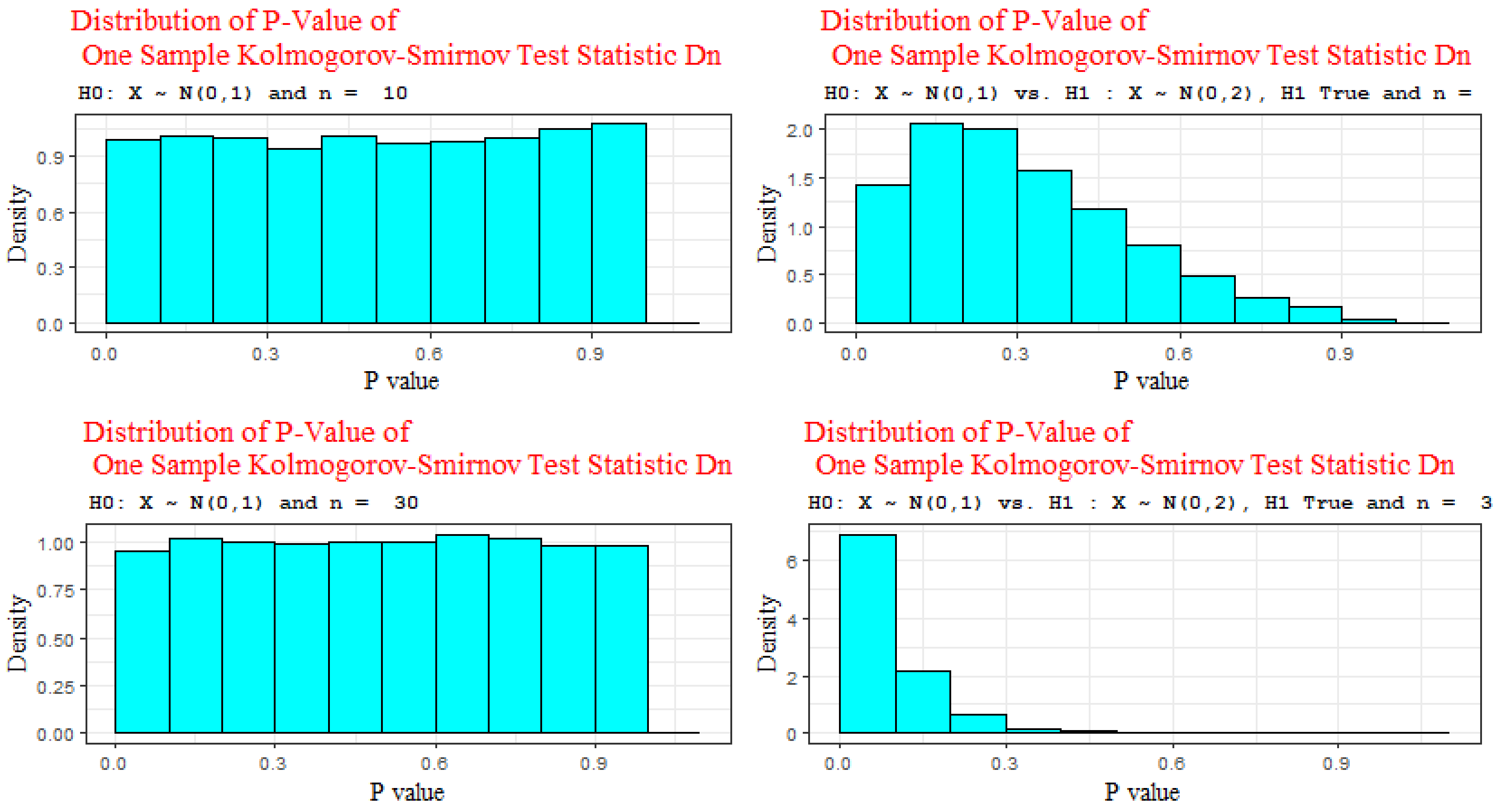
Asymptotic Distribution of ${}_2Dr, n$:



Empirical Size for Two sided Kolmogorov-Smirnov Test :

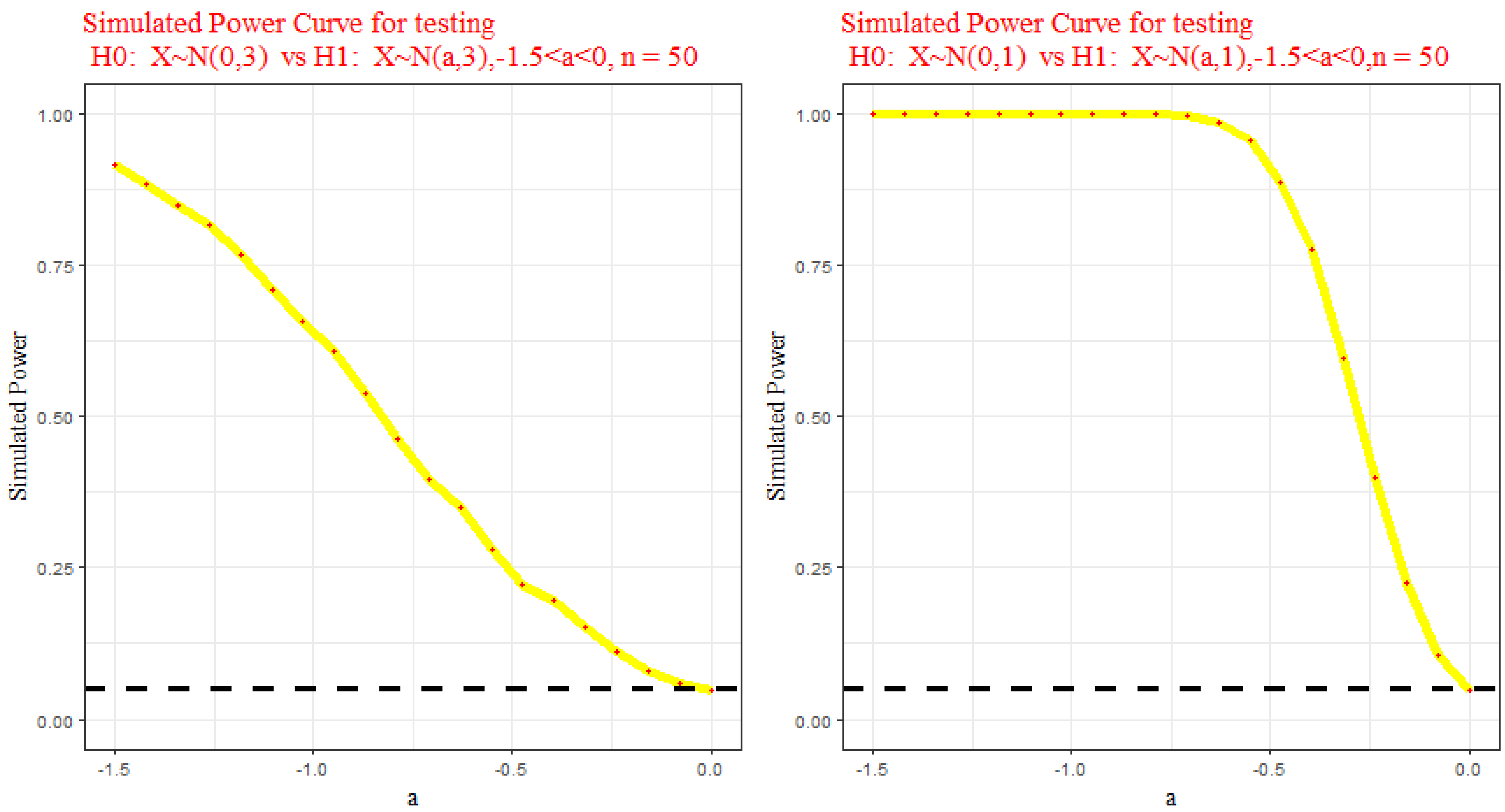
	Normal(0,1)	Logistic(0,1)	Rayleigh(2)
n = 5	0.0436	0.0490	0.0462
n = 20	0.0466	0.0498	0.0484
n = 50	0.0472	0.0520	0.0501
n = 100	0.0506	0.0499	0.0482

Empirical Distribution of P-Value of KS Test :



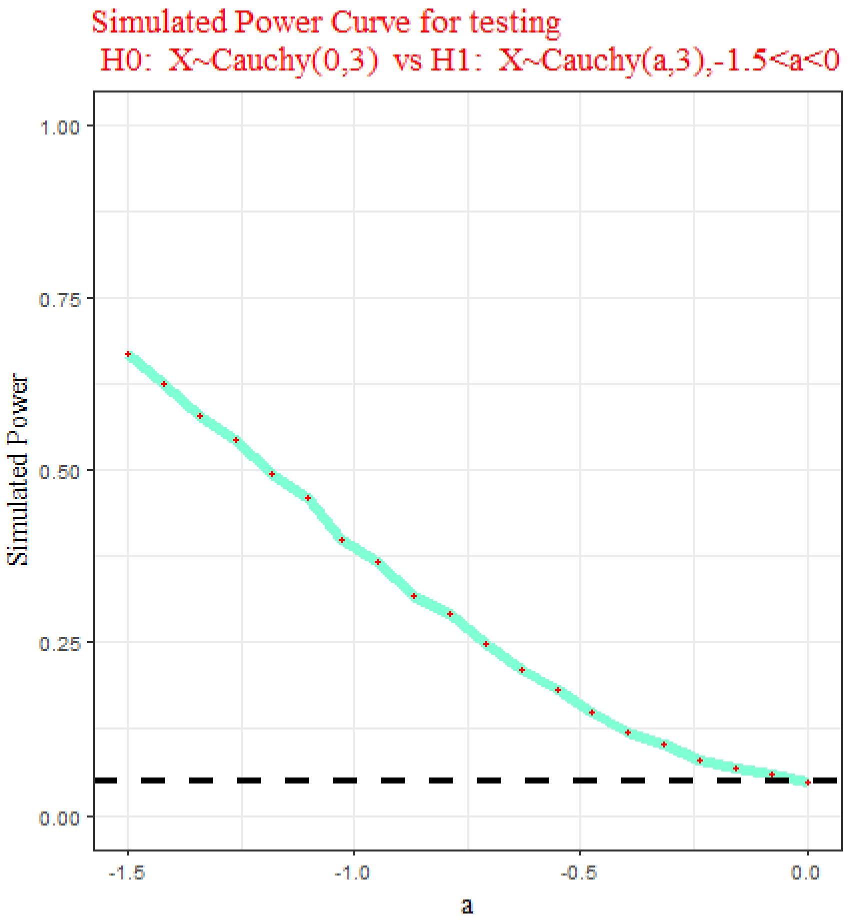
Empirical Power Curve of for One-sided Location Alternative :

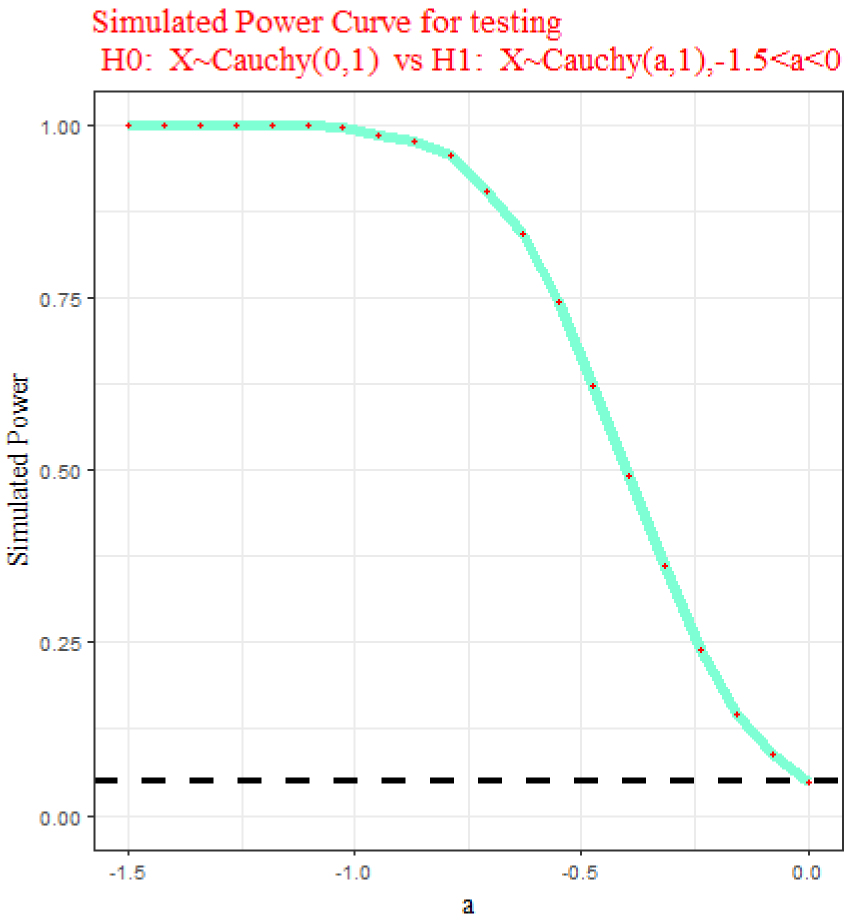
- When the parent Population is Normal



Empirical Power Curve for One-sided Location Alternative :

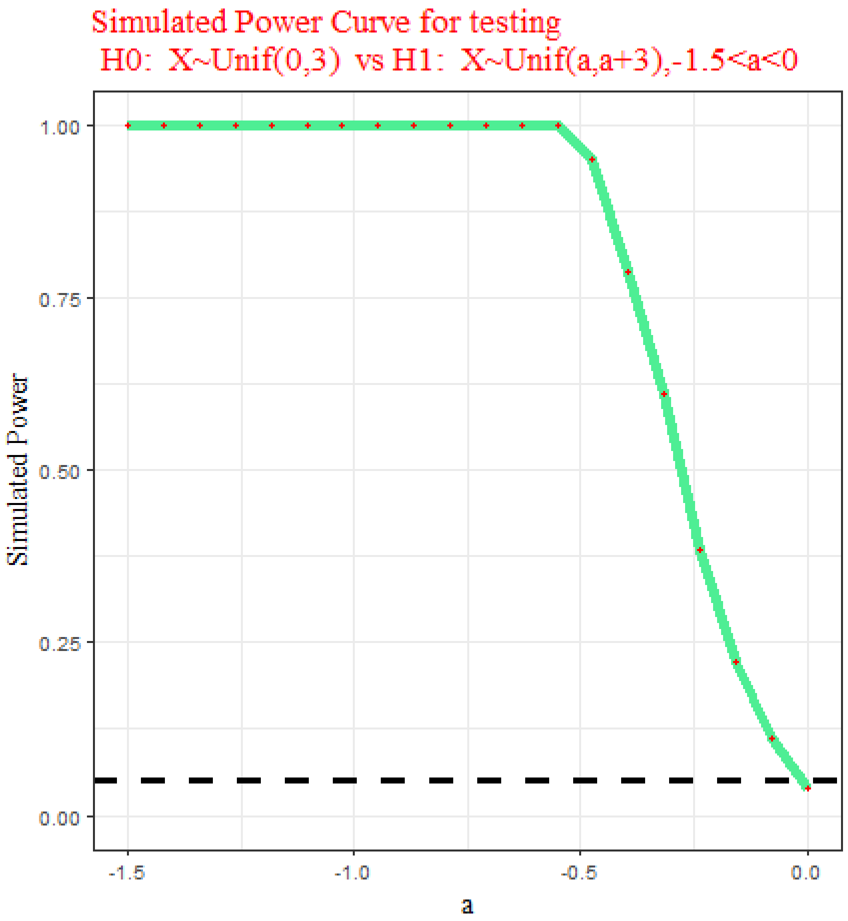
- When the parent Population is Cauchy

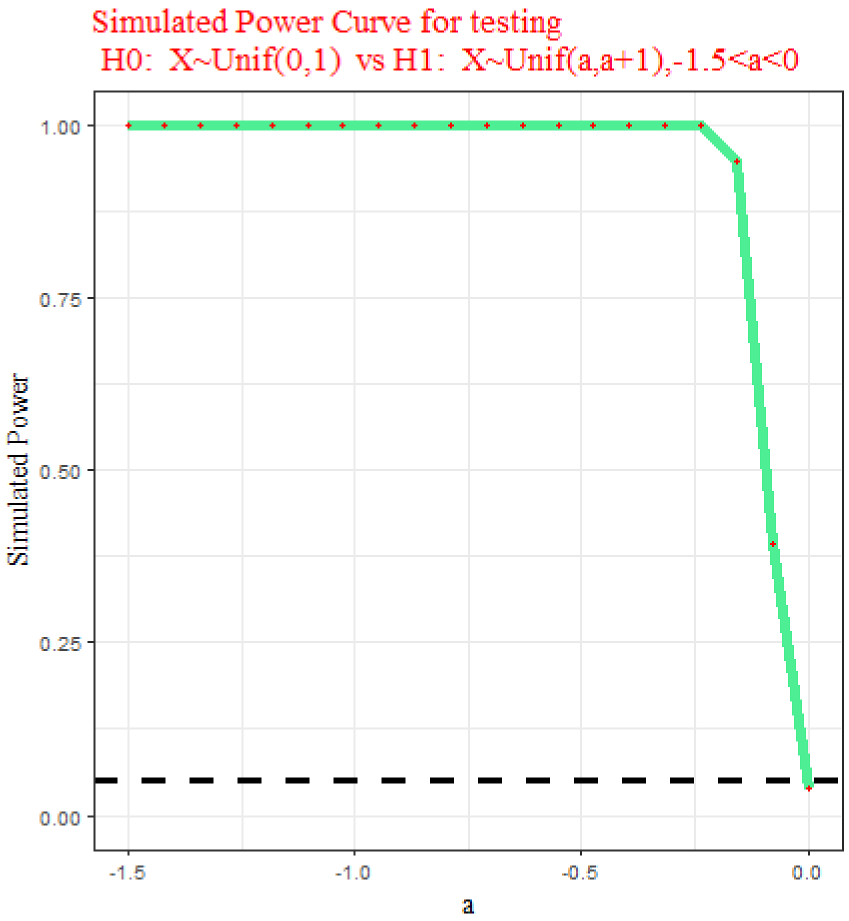




Empirical Power Curve for One-sided Location Alternative :

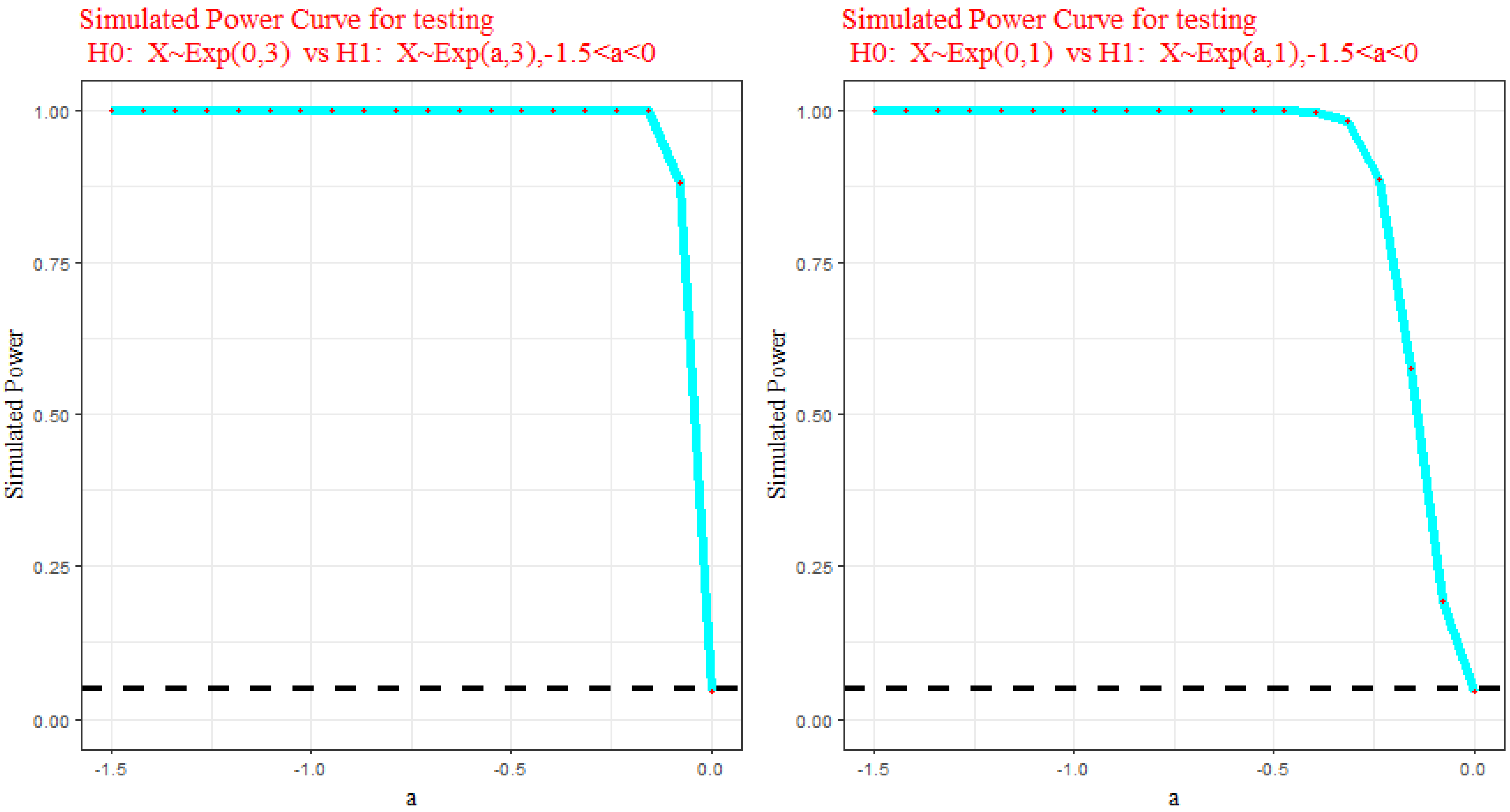
- When the parent Population is Uniform





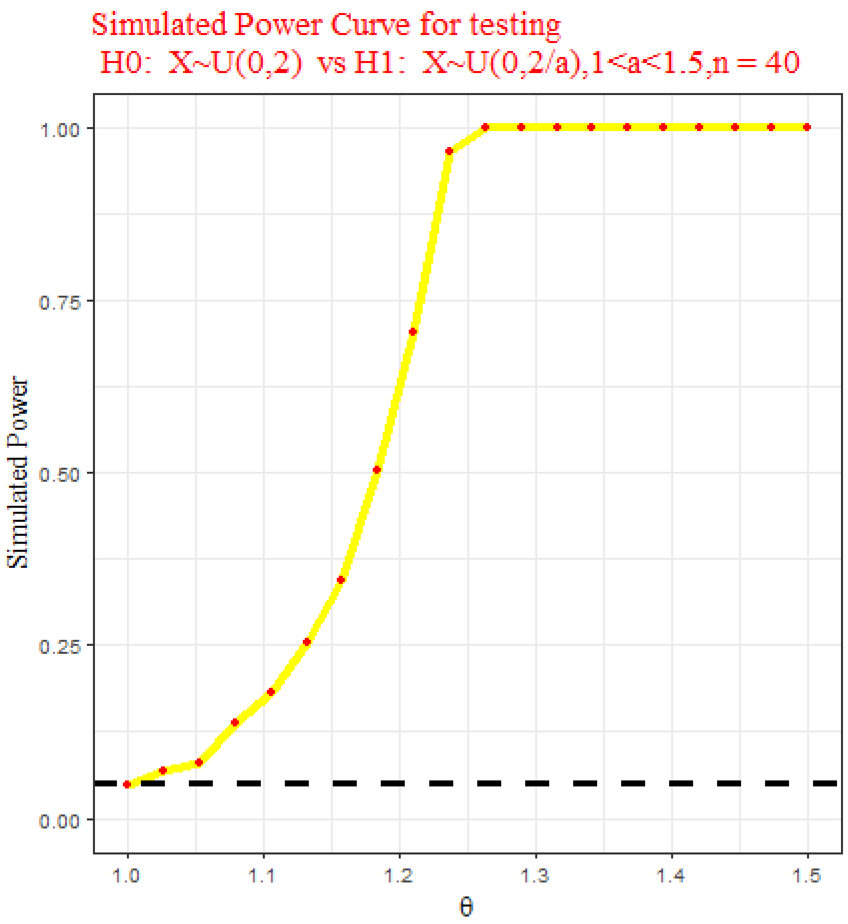
Empirical Power Curve for One-sided Location Alternative :

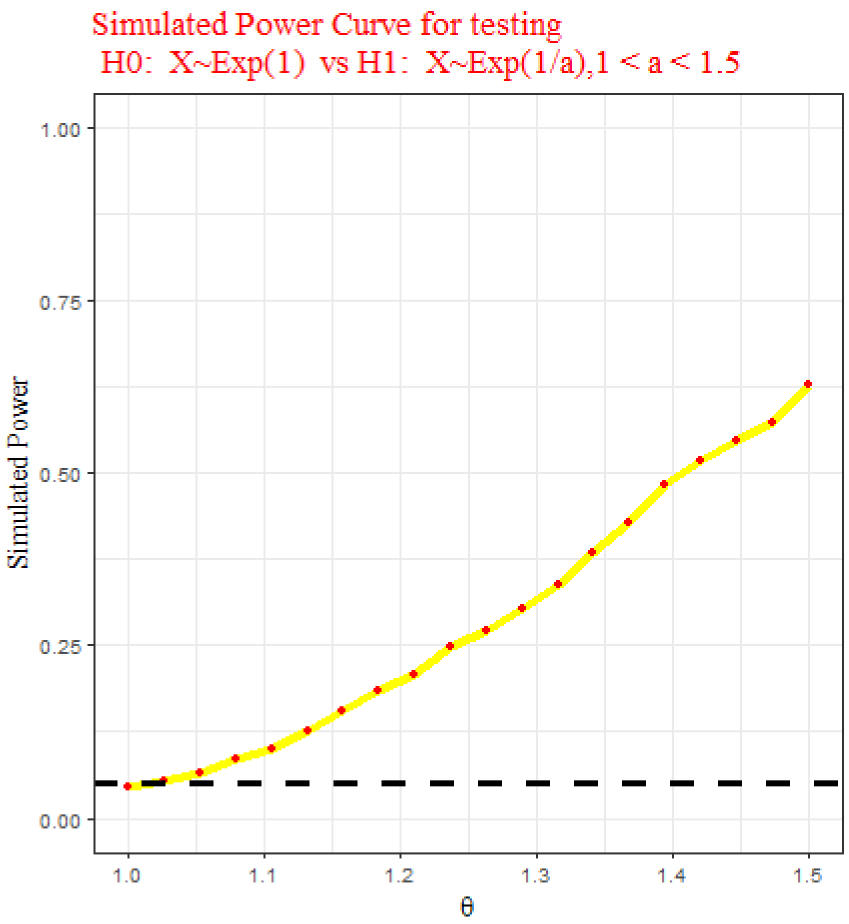
- When the parent Population is Exponential



Empirical Power Curve for One-sided Scale Alternative :

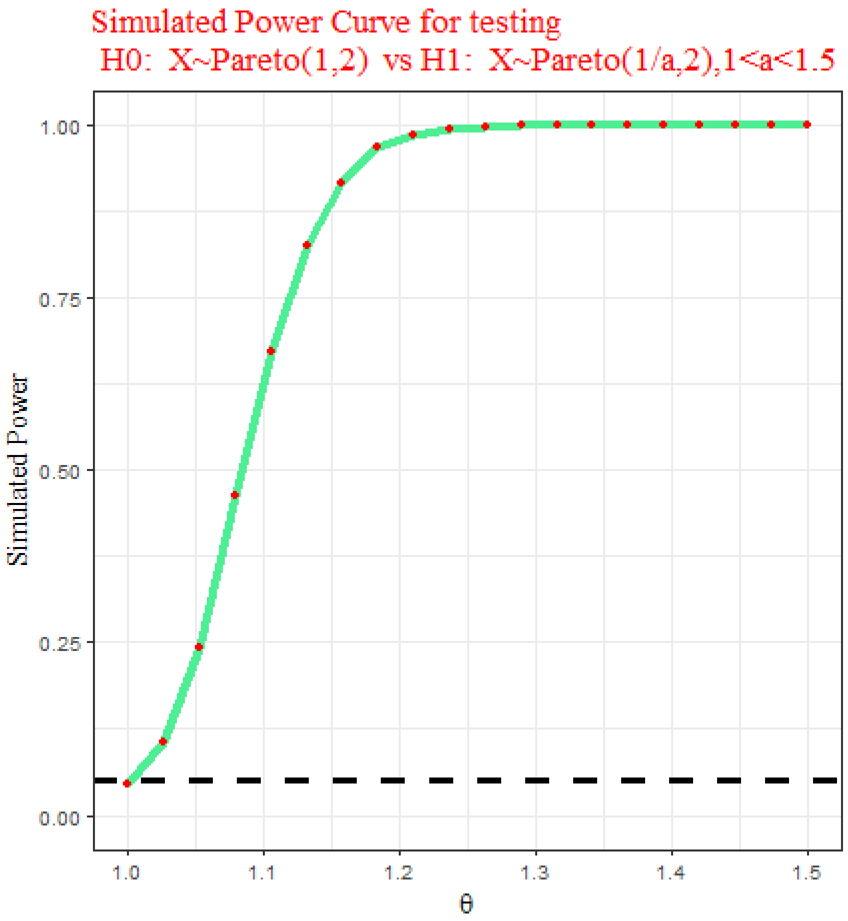
- When the parent populations are Uniform and Exponential





Empirical Power Curve for One-sided Scale Alternative :

- When the parent populations are Rayleigh and Pareto



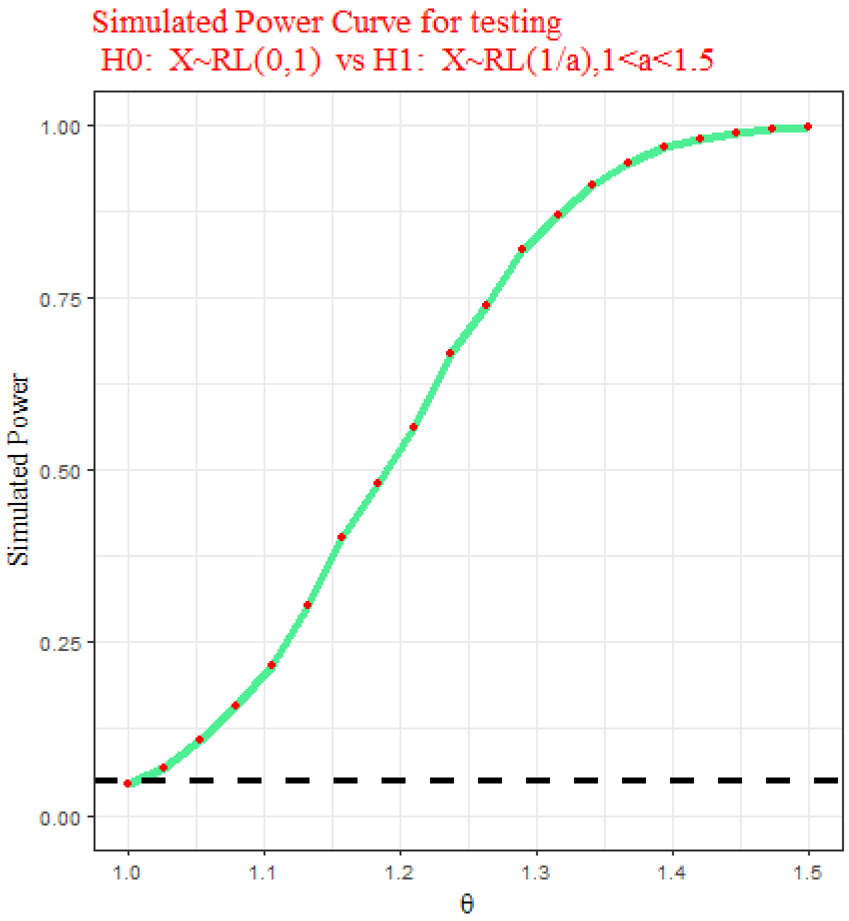


Table of Empirical Power for certain two-sided alternatives(For small sample sizes) :

- For $H_0 : X \sim N(0, 1)$

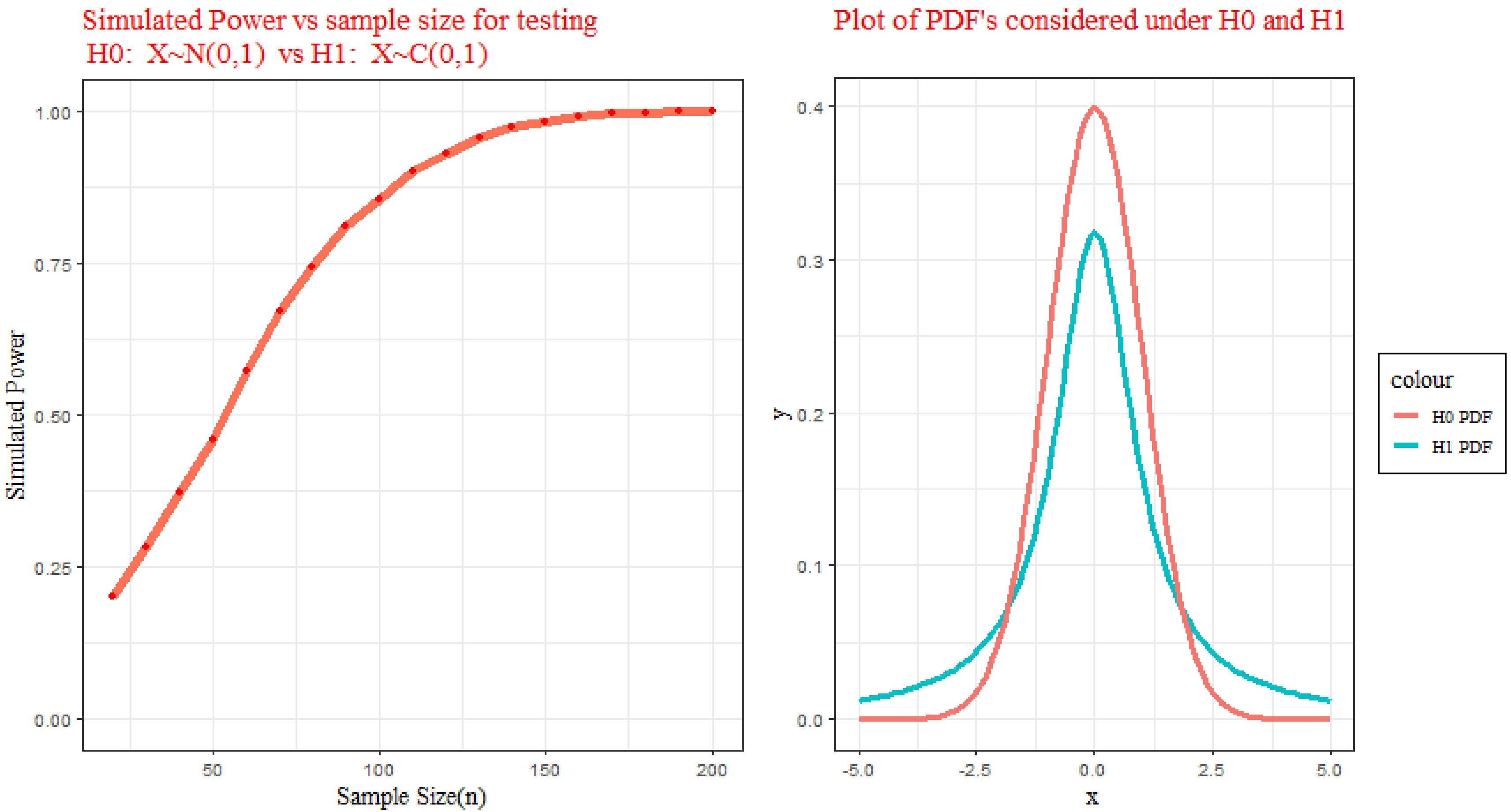
	Cauchy(0,1)	**Laplace(0,1)**	**Logistic(0,1)**	**Truncated Normal(0,1) between(-1,1)**
n = 9	0.1344	0.0526	0.1412	0.0320
n = 12	0.1434	0.0516	0.1640	0.0434
n = 15	0.1618	0.0570	0.1992	0.0508
n = 20	0.1992	0.0580	0.2398	0.0638
n = 25	0.2392	0.0648	0.2806	0.1074

Table of Empirical Power for certain two-sided alternatives(For small sample sizes) :

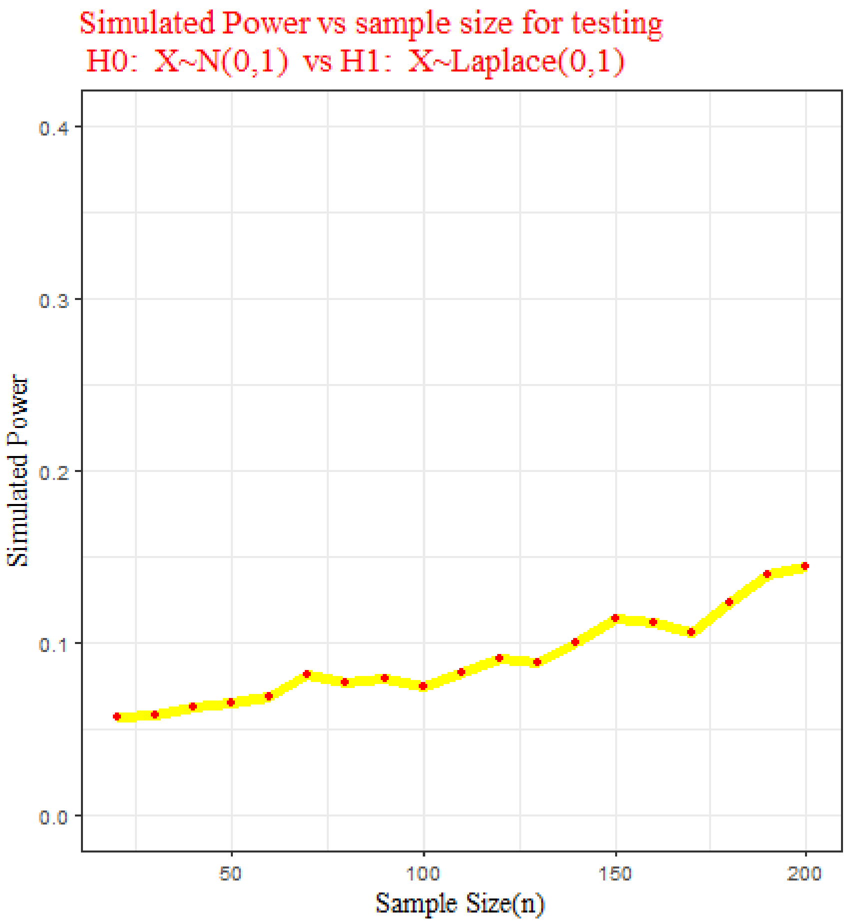
- For $H_0 : X \sim Exp(1)$

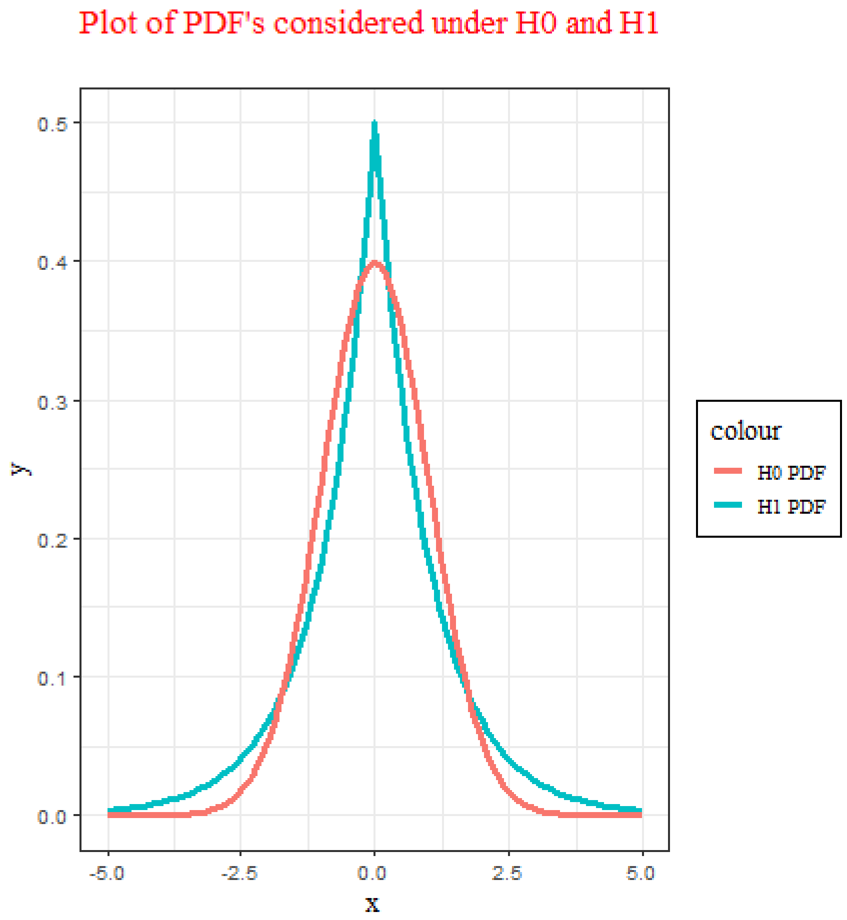
	Rayleigh(1)	Weibull(scale=1,shape=2)	Log Normal(0,1)
n = 9	0.3974	0.0810	0.1464
n = 12	0.5442	0.1258	0.1946
n = 15	0.6878	0.1776	0.2504
n = 20	0.8158	0.2748	0.3082
n = 25	0.9090	0.3732	0.3734

Empirical Power vs sample size for certain two-sided alternatives :

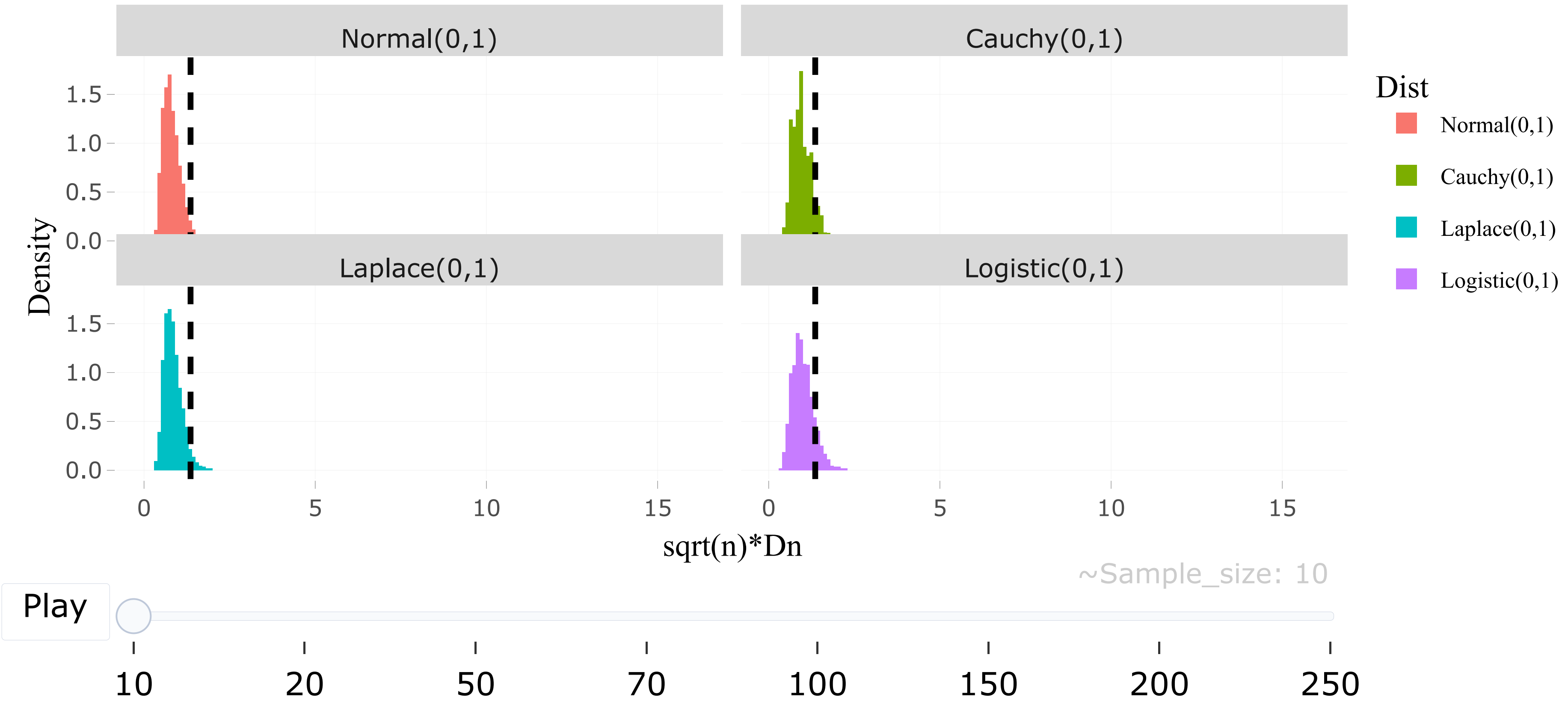


Empirical Power vs sample size for certain two-sided alternatives :

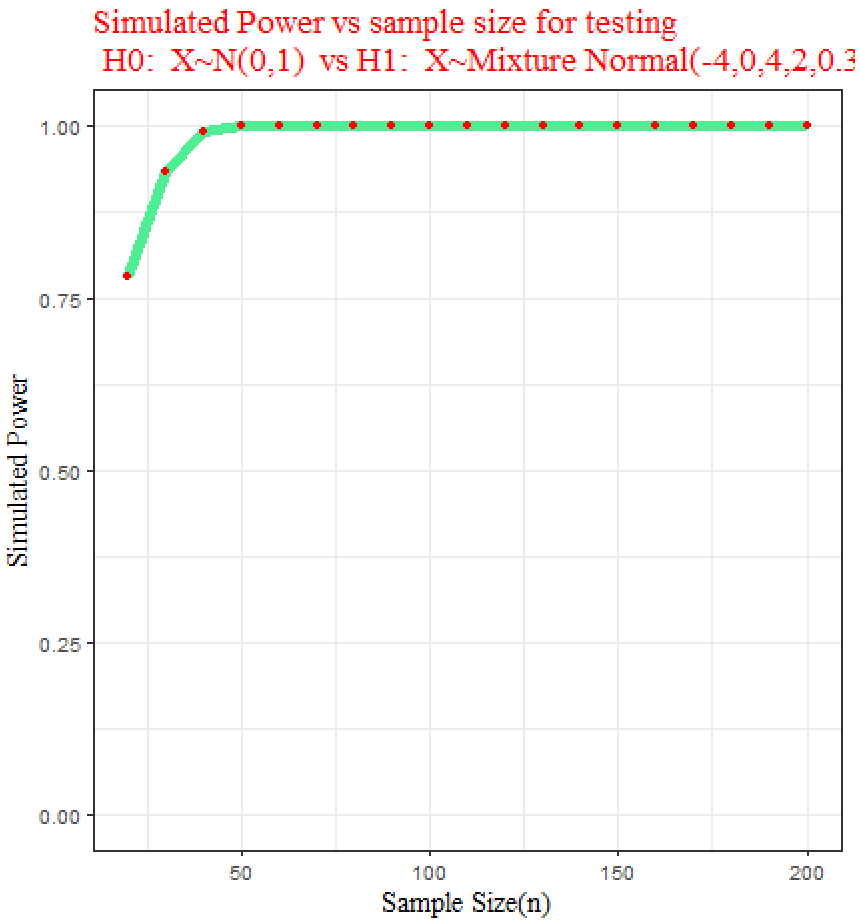


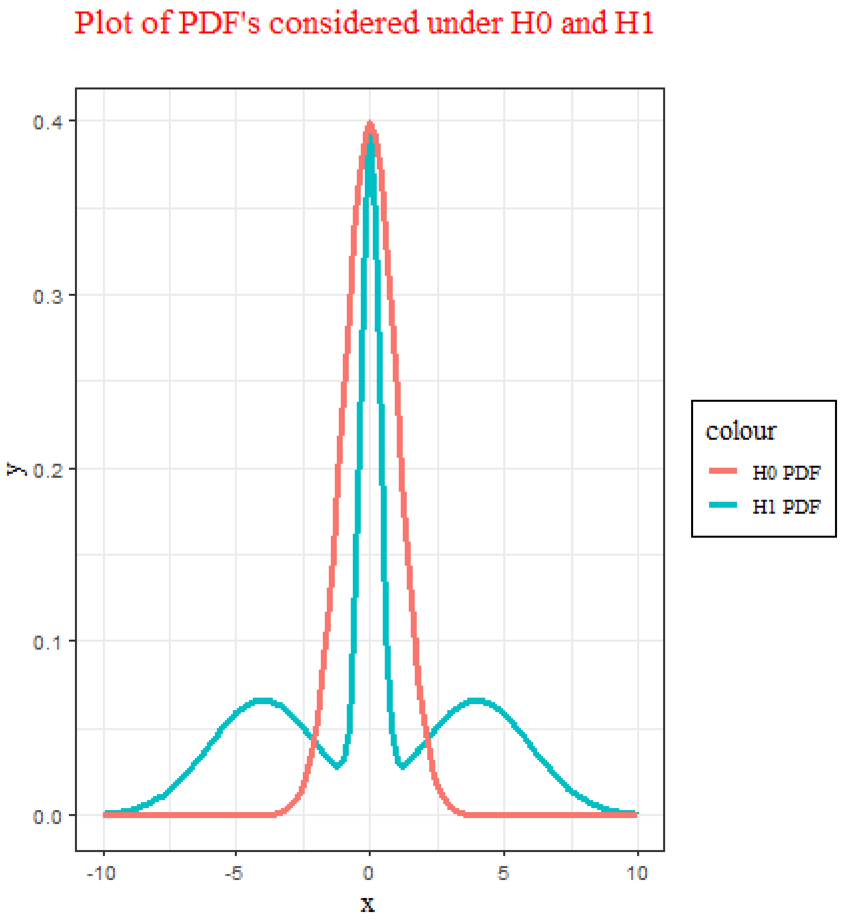


Why is the power so less for Laplace?(Animatic Representation)

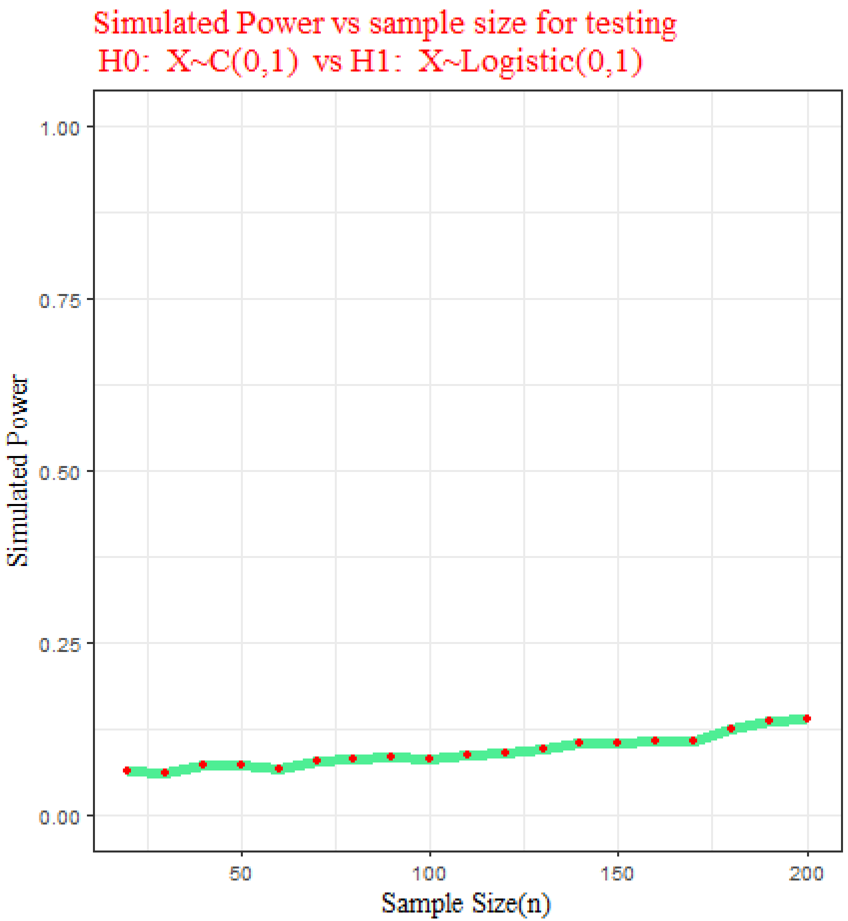


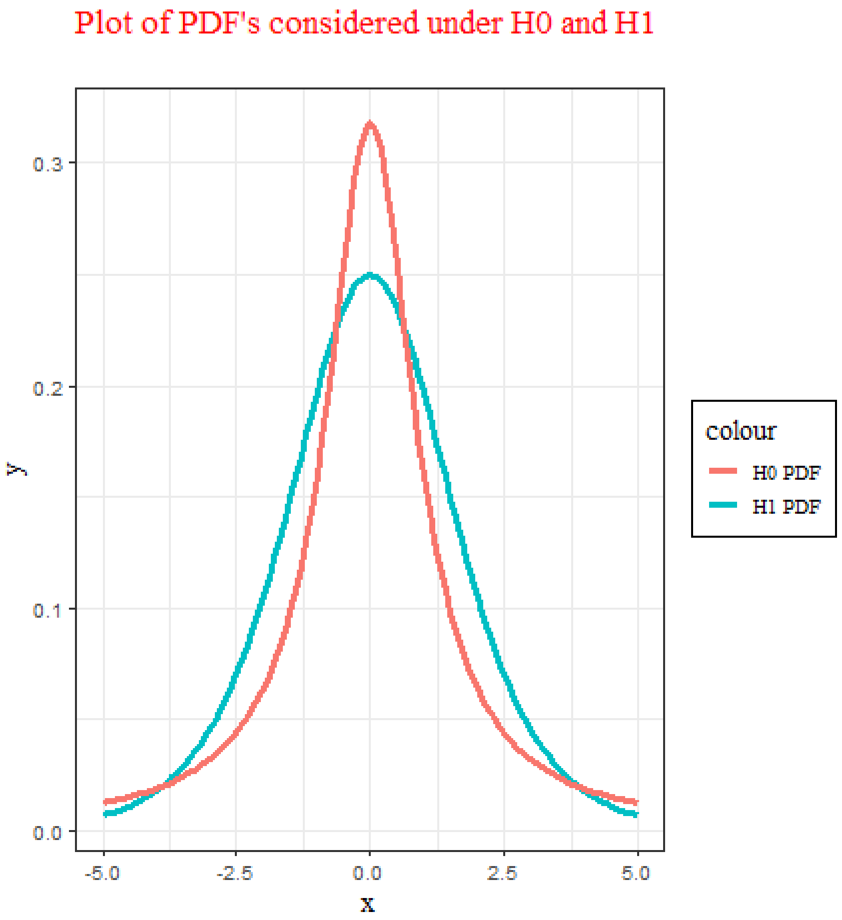
Empirical Power vs sample size for certain two-sided alternatives :



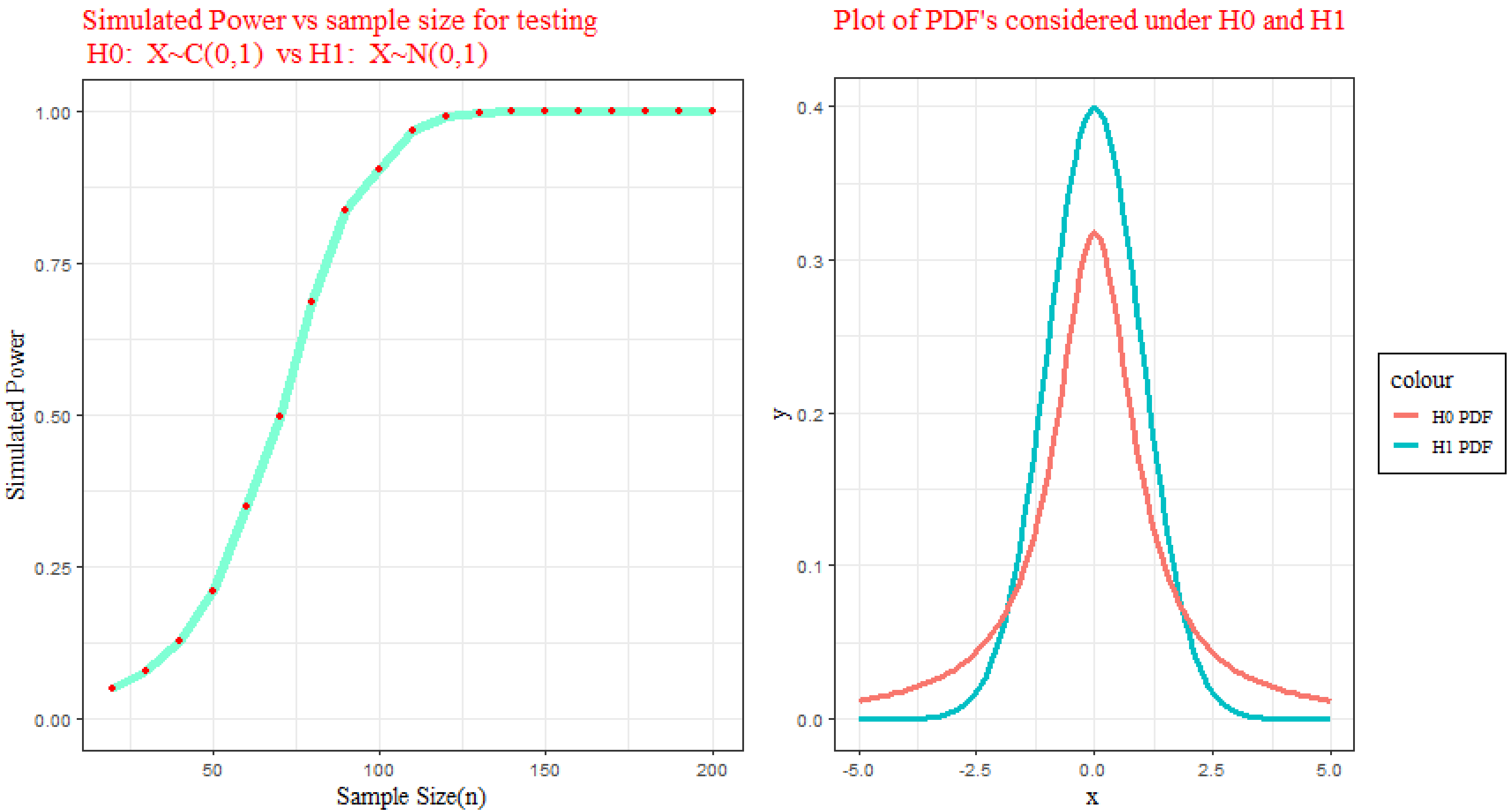


Empirical Power vs sample size for certain two-sided alternatives :

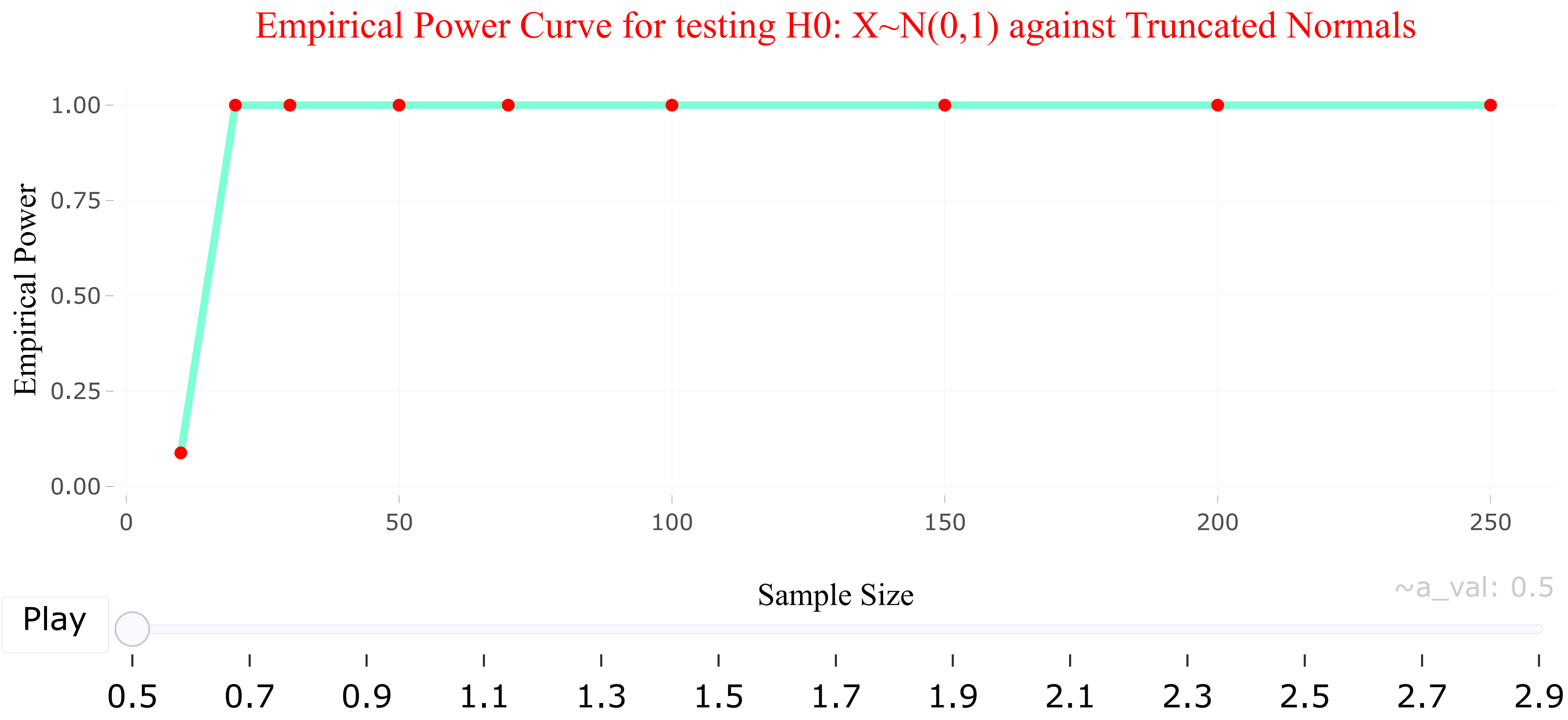




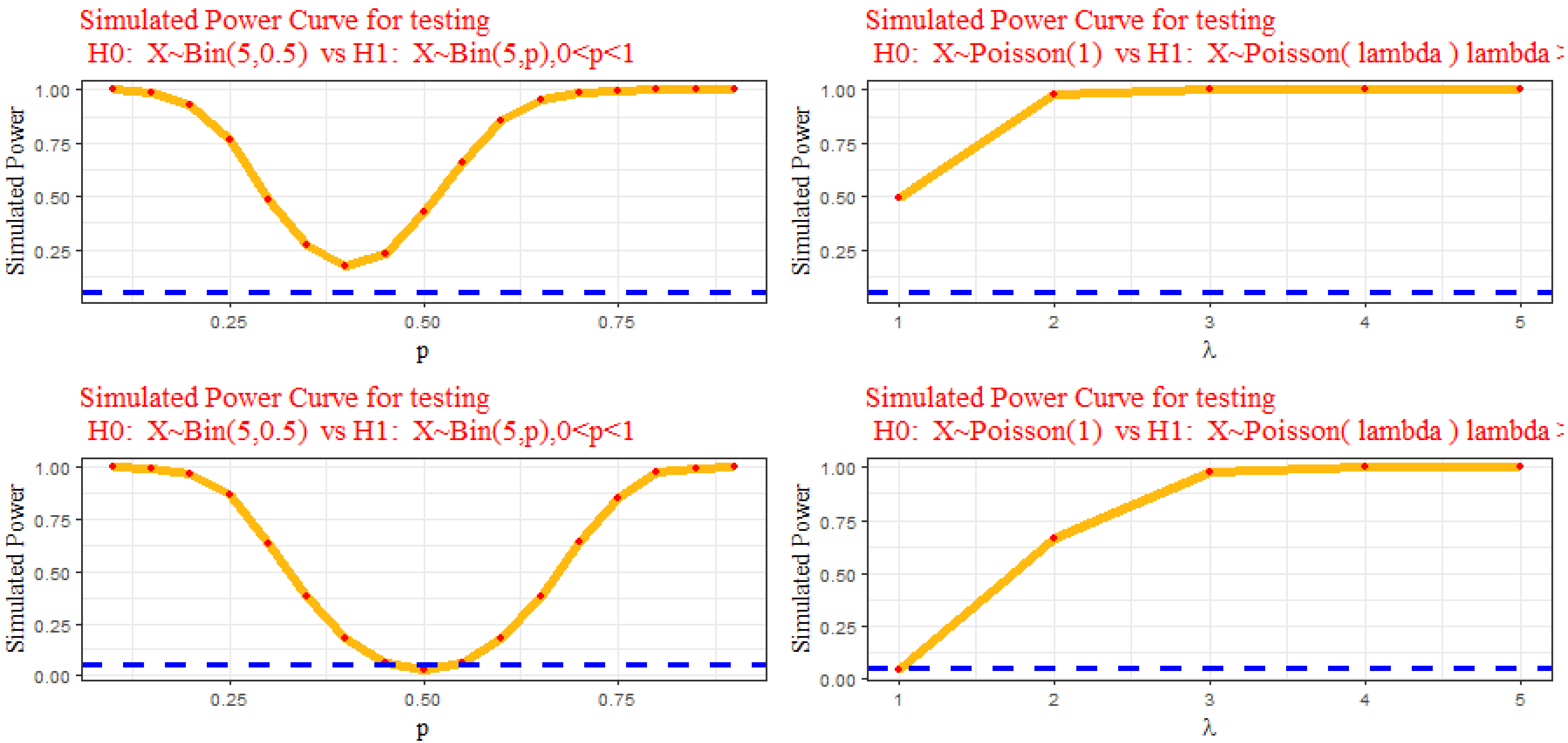
Empirical Power vs sample size for certain two-sided alternatives :



Empirical study how test for N(0,1) behaves when data has come from Truncated Normal(0,1) between (-a,a) :



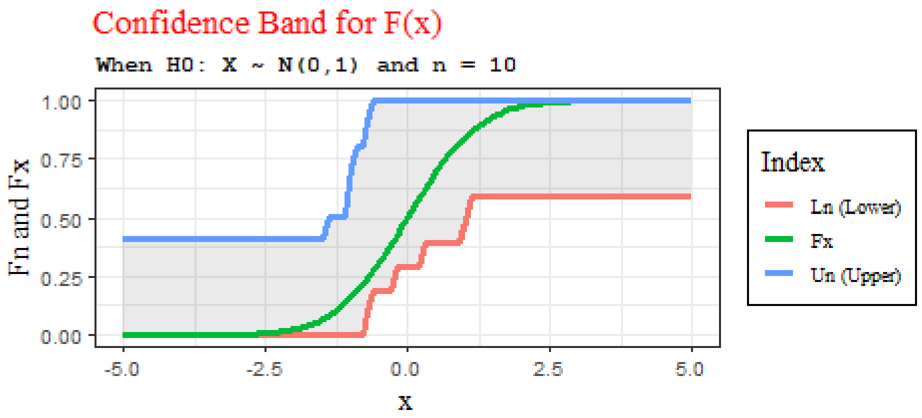
Corrected Power Curve for Discrete Case :

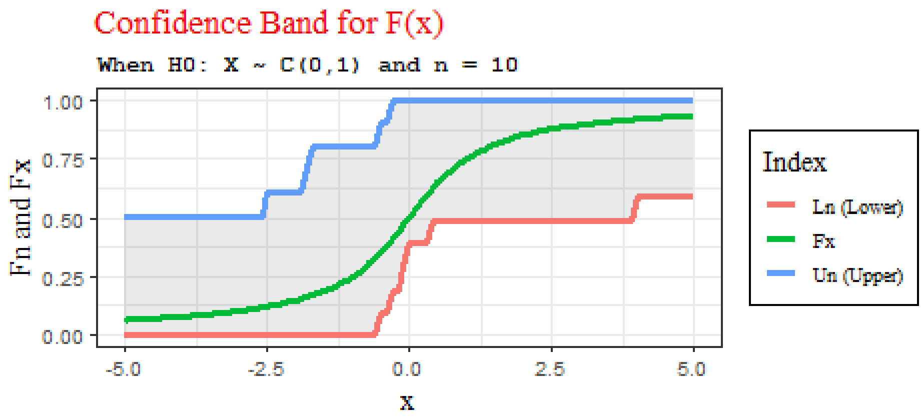


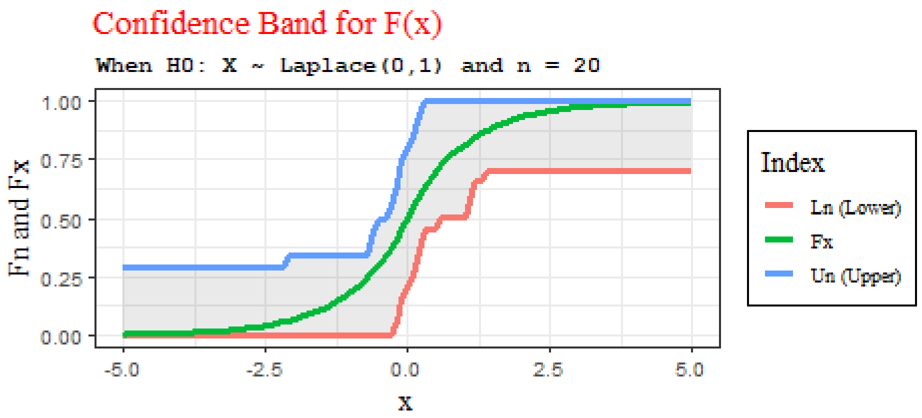
Confidence Band for $F(x)$:

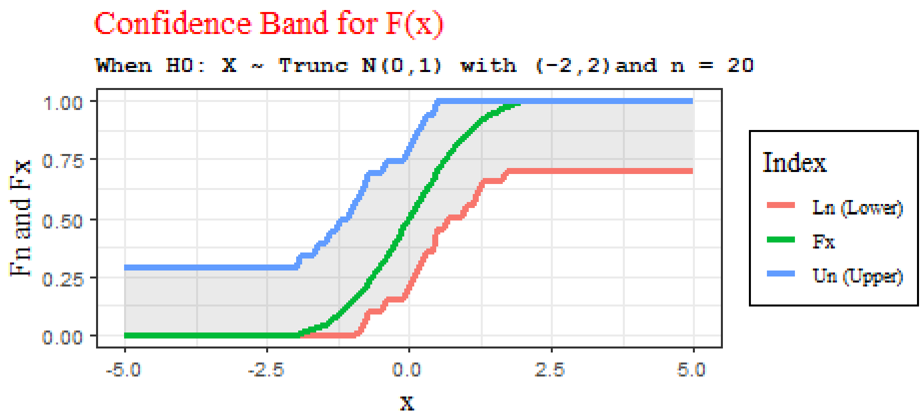
- We know that the confidence band is a random band which covers the Distribution Function F with a preassigned probability.
- Hence that $(1-\alpha)100\%$ confidence band for F can be derived as, $\mathbb{P}(L_n(x) \leq F(x) \leq U_n(x)) = 1-\alpha$ where, $L_n(x) = \max(0, \mathbb{F}_\mathrm{n}(x) - d_{n,\alpha})$ and $U_n(x) = \min(\mathbb{F}_n(x) + d_{n,\alpha}, 1)$

Confidence Band for Some Specific Distributions :

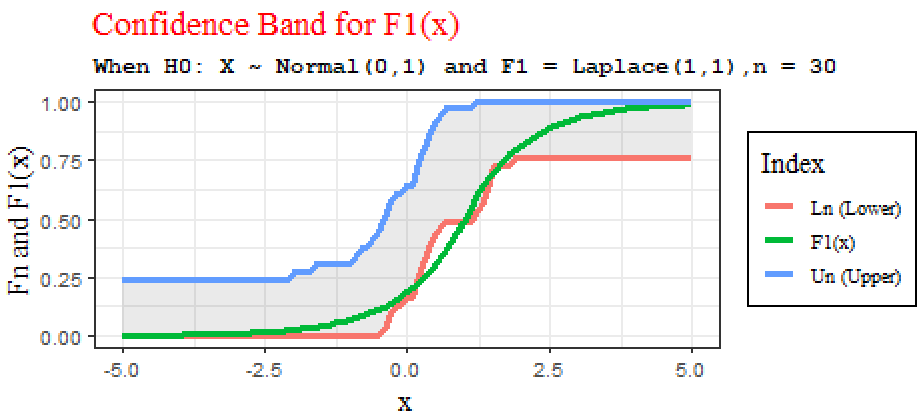


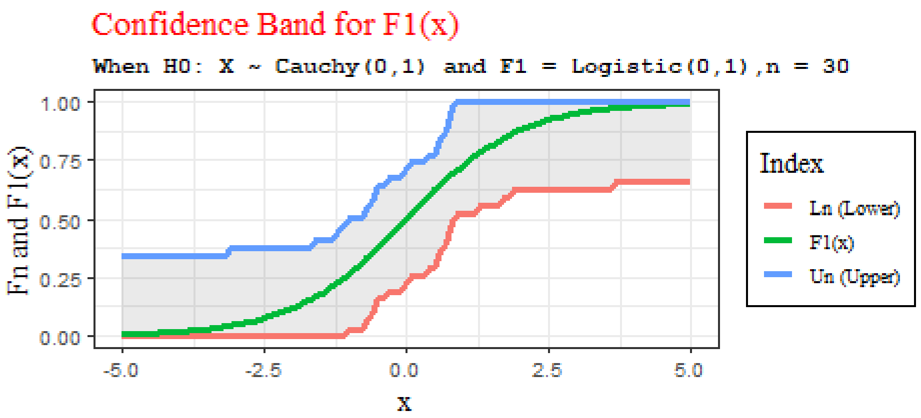


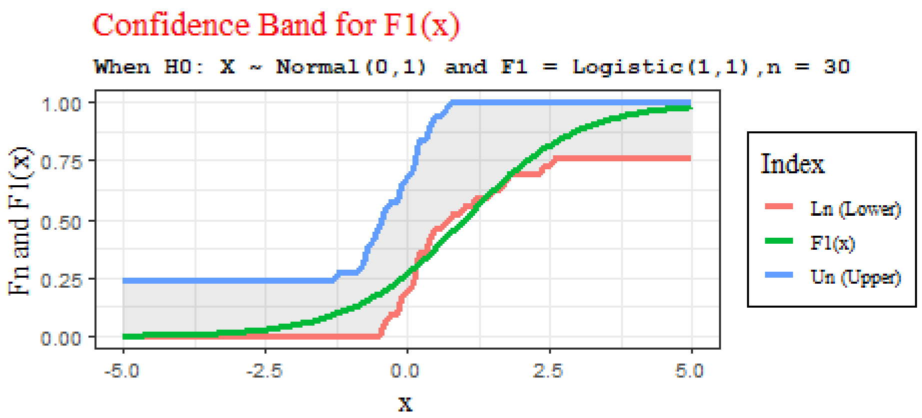


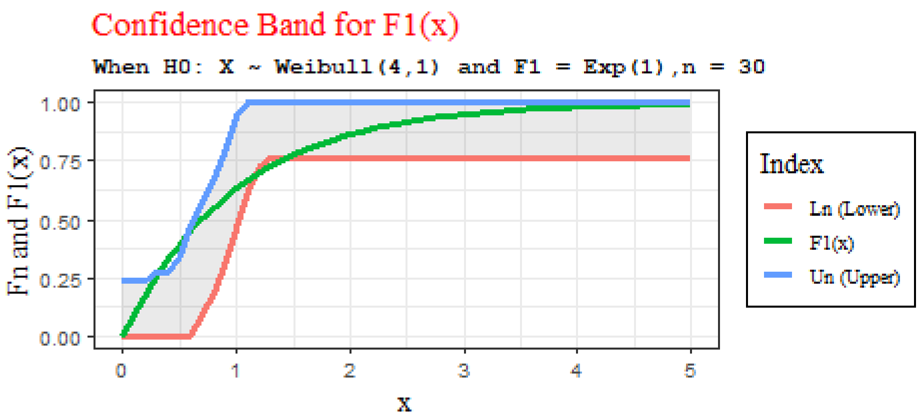


Confidence Band for F_1 when Data is generated from F_0 :









Coverage of Confidence Band For Some Particular Distributions :

- Theoretically Coverage of a Confidence Band is defined as - $\mathbb{P}_{\mathcal{H}_o}(\mathbb{F}_n(x) - d_{n,\alpha} \leq F_1(x) \leq \mathbb{F}_n(x) + d_{n,\alpha};\ for\ all\ x \in \mathrm{r})$
- When F_1 is the same CDF as the CDF under H_0 , then it is called Confidence Band.

	Laplace(0,1)	**Normal(0,1)**	**Logis(0,1)**	**Cauchy(0,1)**
Laplace(0,1)	0.963	0.962	0.934	0.968
Normal(0,1)	0.954	0.964	0.926	0.968
Logis(0,1)	0.814	0.814	0.963	0.968
Cauchy(0,1)	0.834	0.834	0.948	0.963

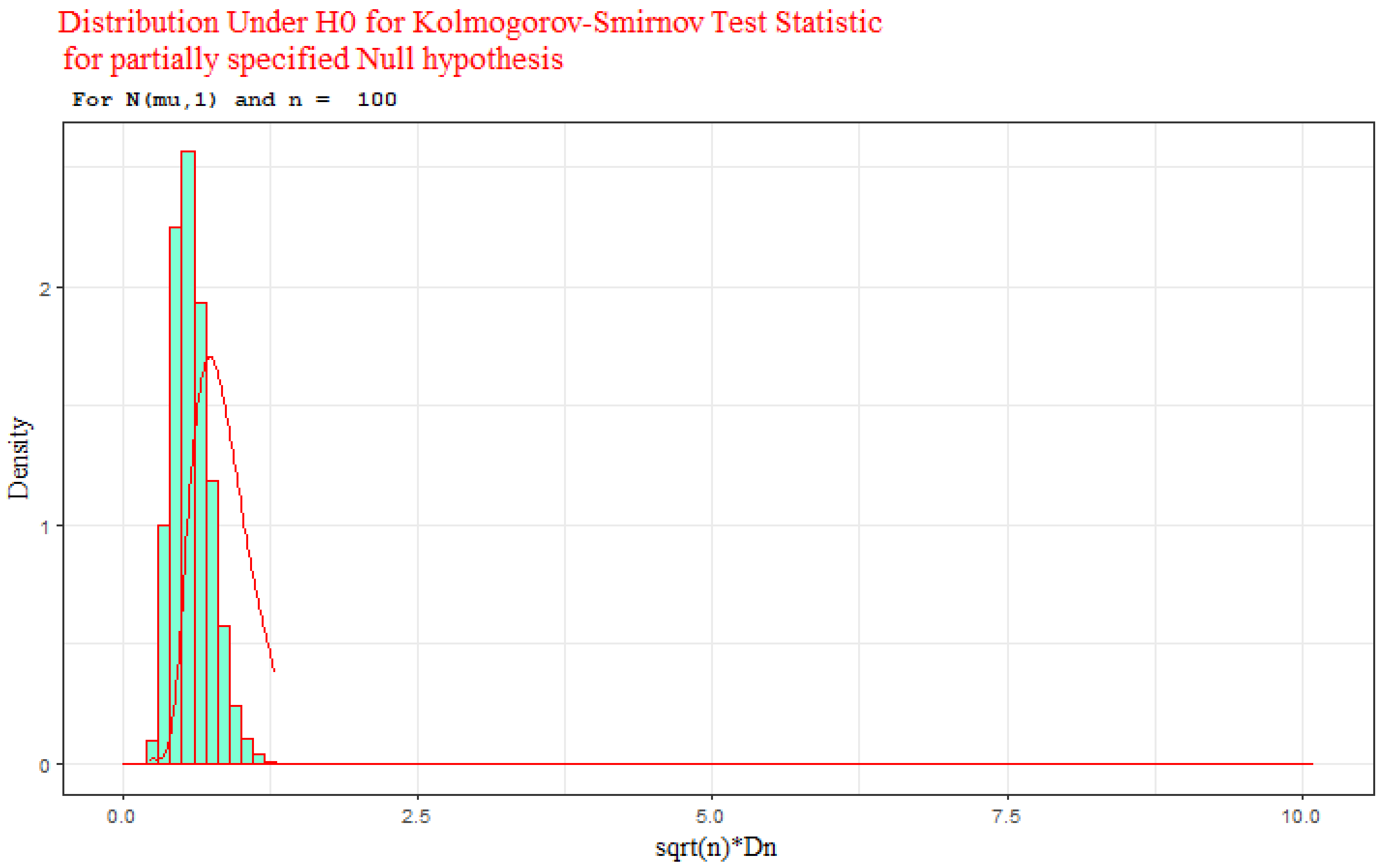
Kolmogorov-Smirnov test for Partially Specified Null Hypothesis :

- Till now, what ever observations we made was based on taking the null hypothesis to be completely specified.
- Let us consider the problem when the null hypothesis is not completely specified as the above cases.
- In particular let us consider the null as,

$$\mathcal{H}_o : X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1); \mu \in \mathbb{R}$$

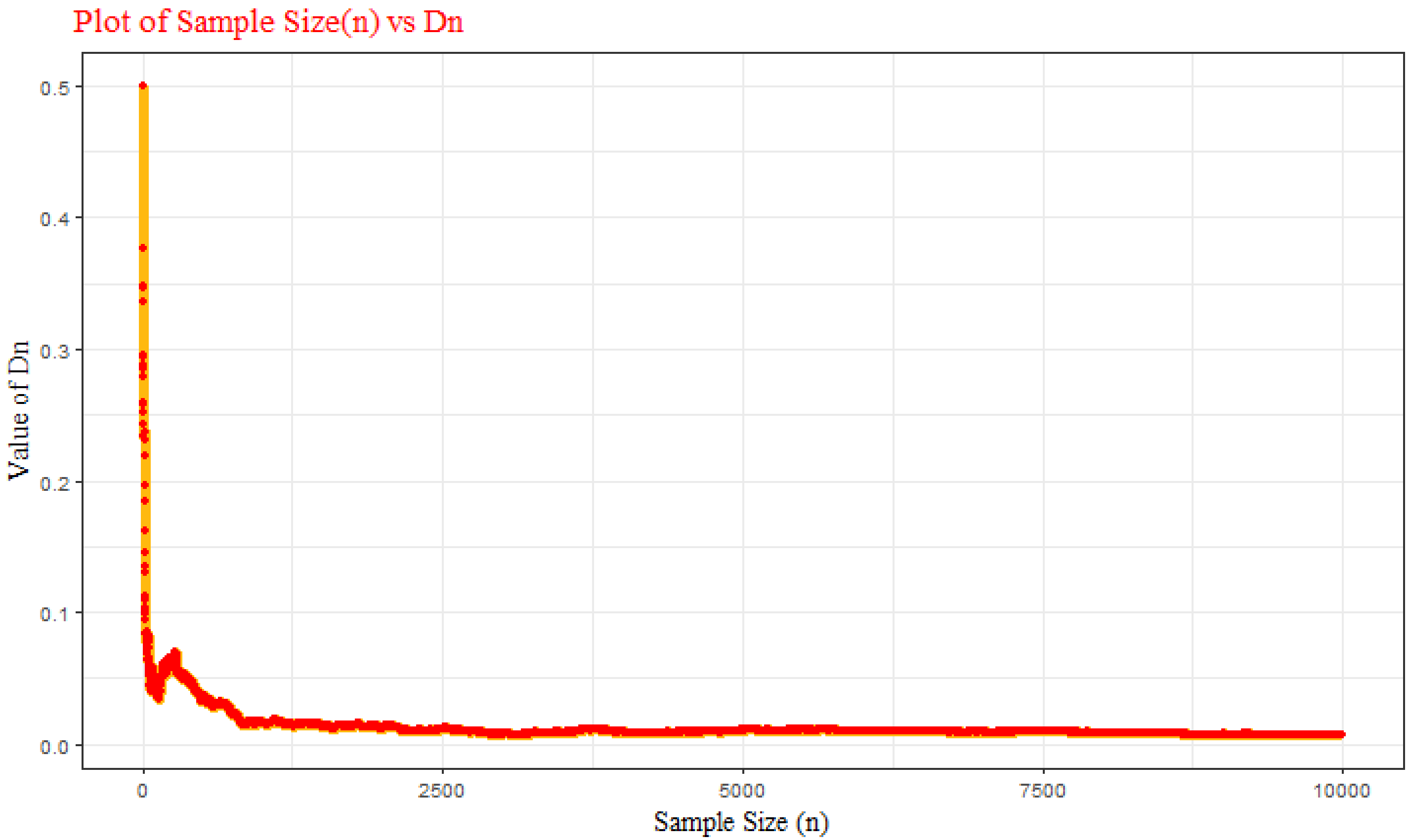
- We first see compute the D_n^+ under $\mathcal{H}_o$ statistic for a sample $X_1, \ldots X_{100}$ drawn from $N(\mu, 1)$ where μ is unknown.
- Since we don't know μ we estimate μ by the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{100} X_i$ and take $\hat{F}_o = N(\bar{X}_n, 1)$ to calculate $D_n^+ = \sup_x(\hat{F}_o(x) - \mathbb{F}_n(x))$.

Kolmogorov-Smirnov test for Partially Specified Null Hypothesis :



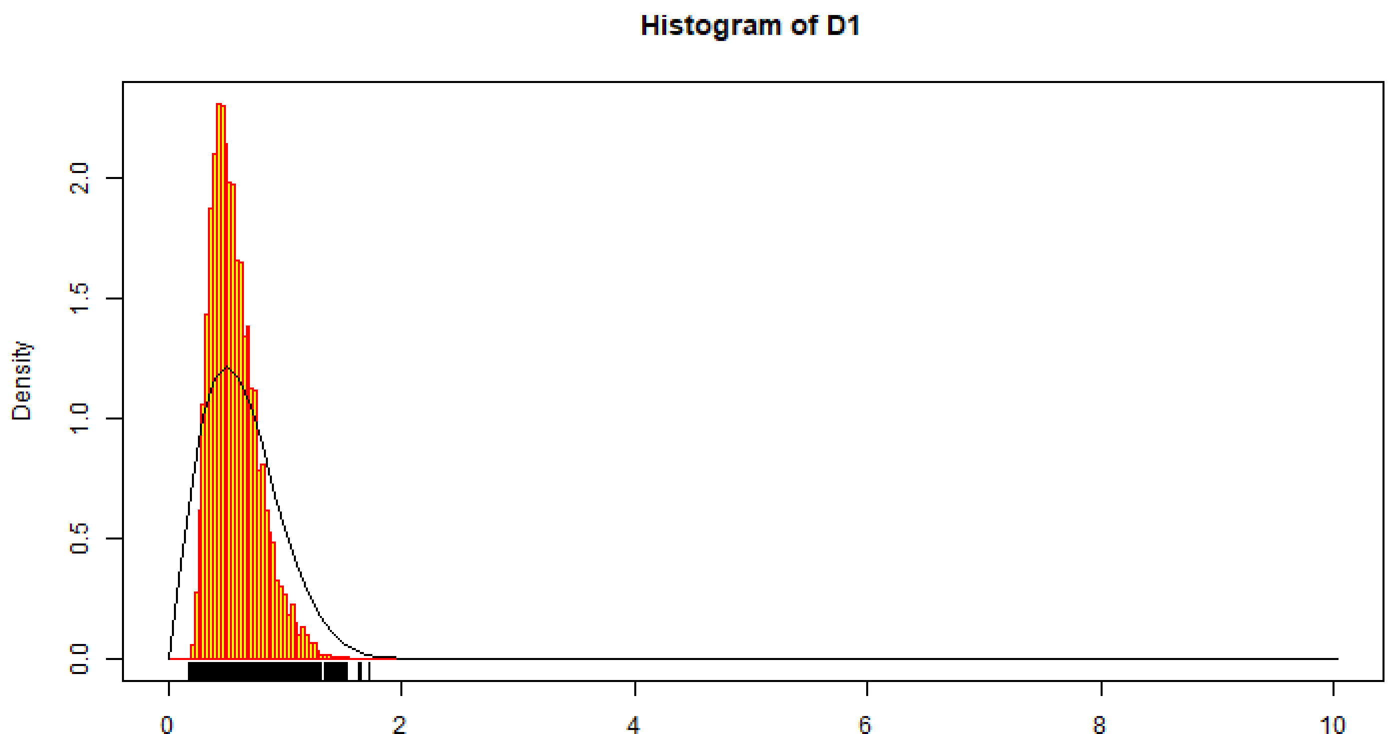
Convergence of D_n under Partially Specified Null Hypothesis :

- From the heavily positive skewness of the distribution of D_n^+ under the Partially Specified Null Hypothesis we get the indication that even though we don't know the exact null distribution, the statistic D_n tend to go to 0, detecting Normality in the data.



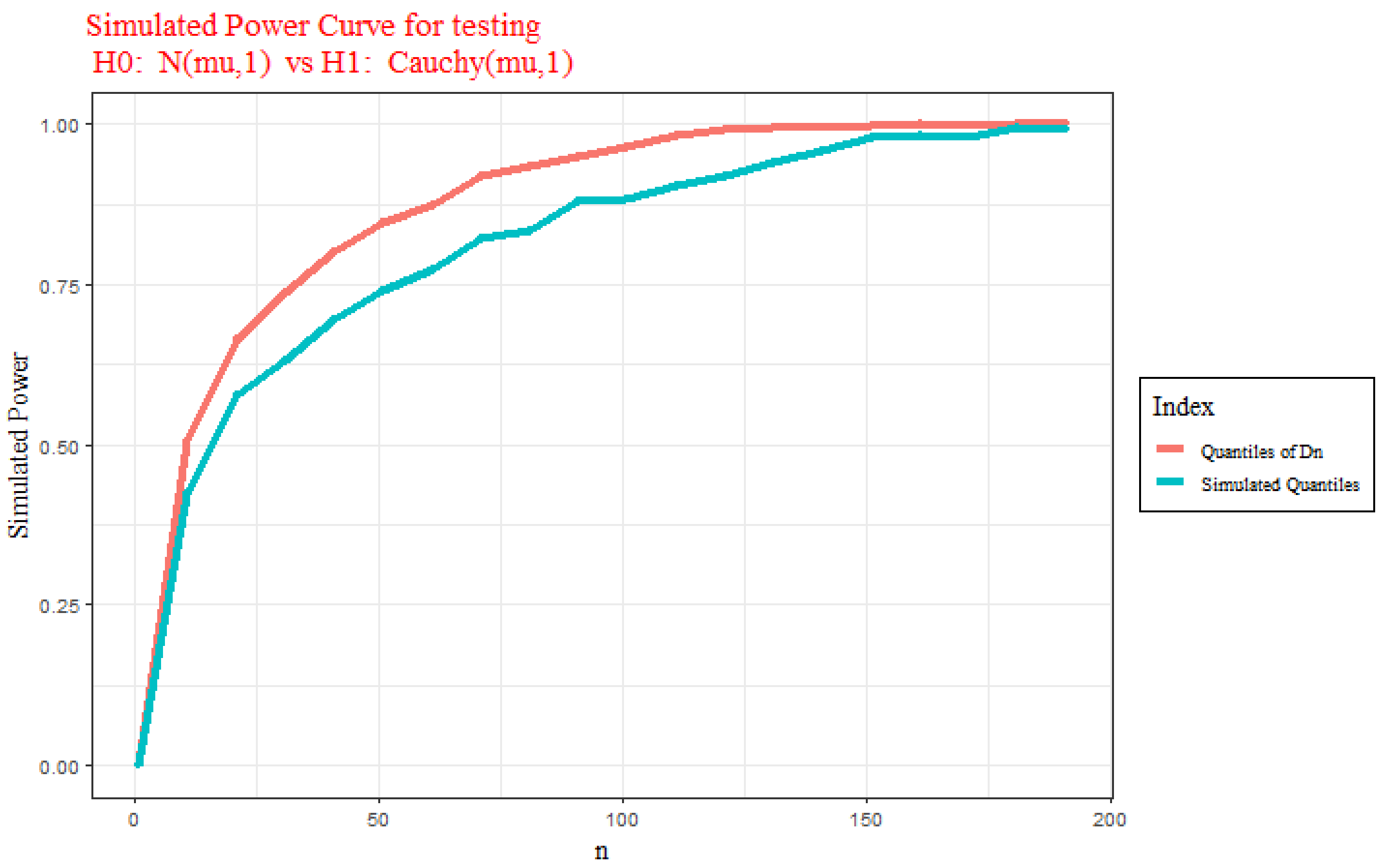
Kolmogorov Statistic under Partially Specified Null Hypothesis for Exponential with unknown mean

- Here we consider the Null Hypothesis to be $\mathcal{H}_o : X_1, \ldots, X_n \overset{iid}{\sim} Exponential(mean = \theta)$.
- We estimate the rate by $\bar{X_n}^{-1}$(reciprocal of the sample mean).We plot the distribution of D_n^+ and the check the convergence of D_n with the estimated value of θ.

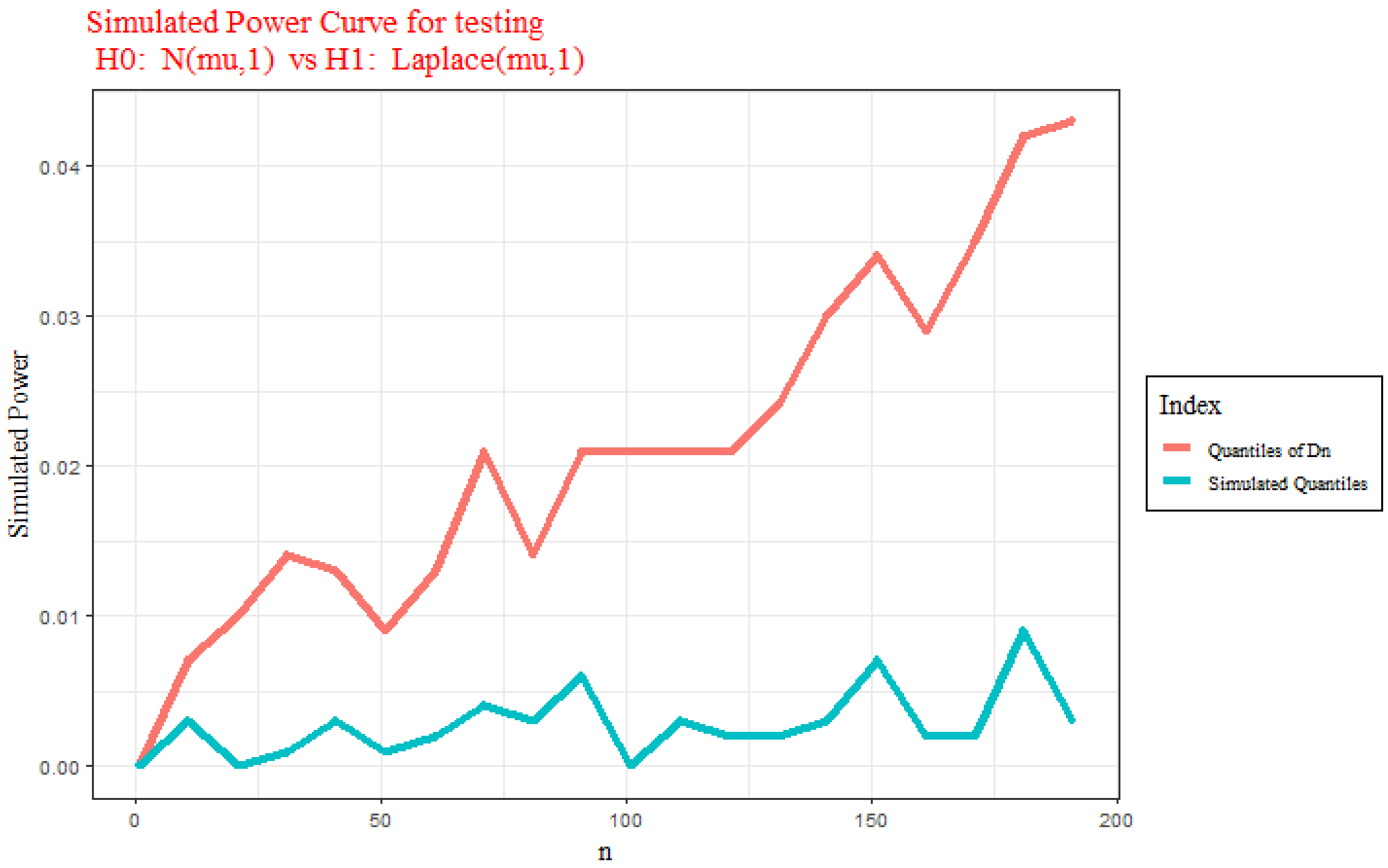


Power Curve for the Partially Specified Null Hypothesis :

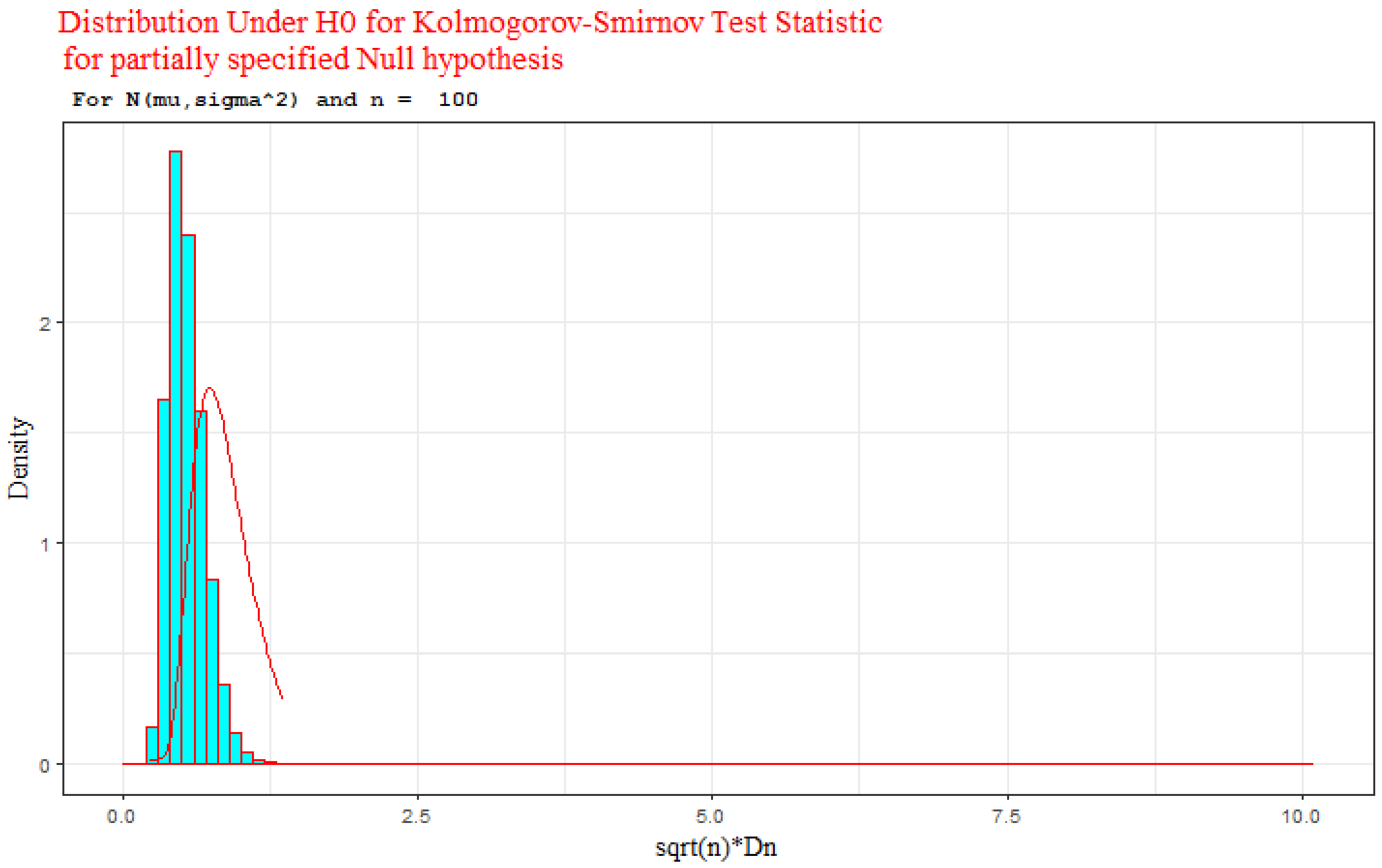
- Normal vs Cauchy for unknown location



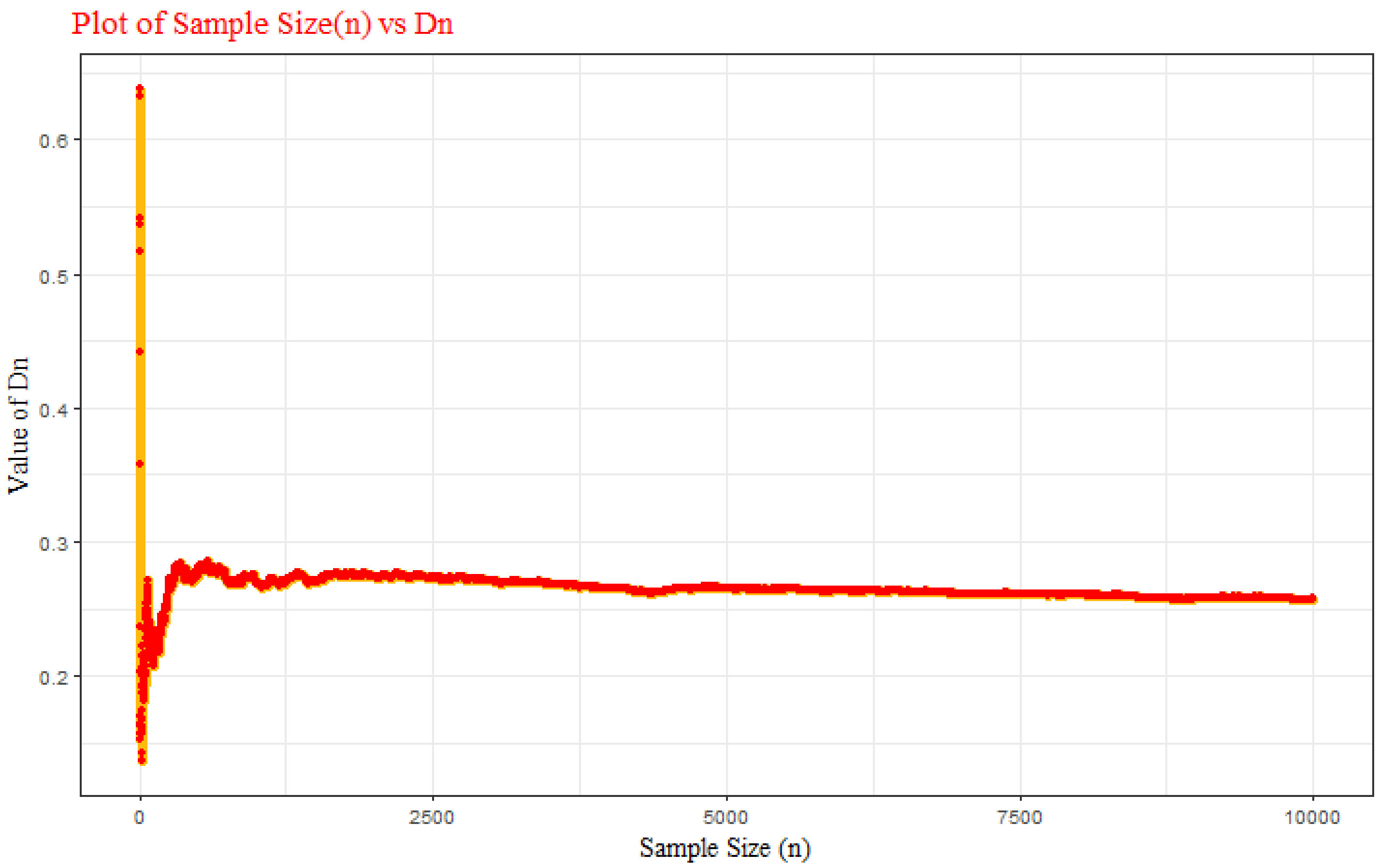
Normal vs Laplace for unknown location :



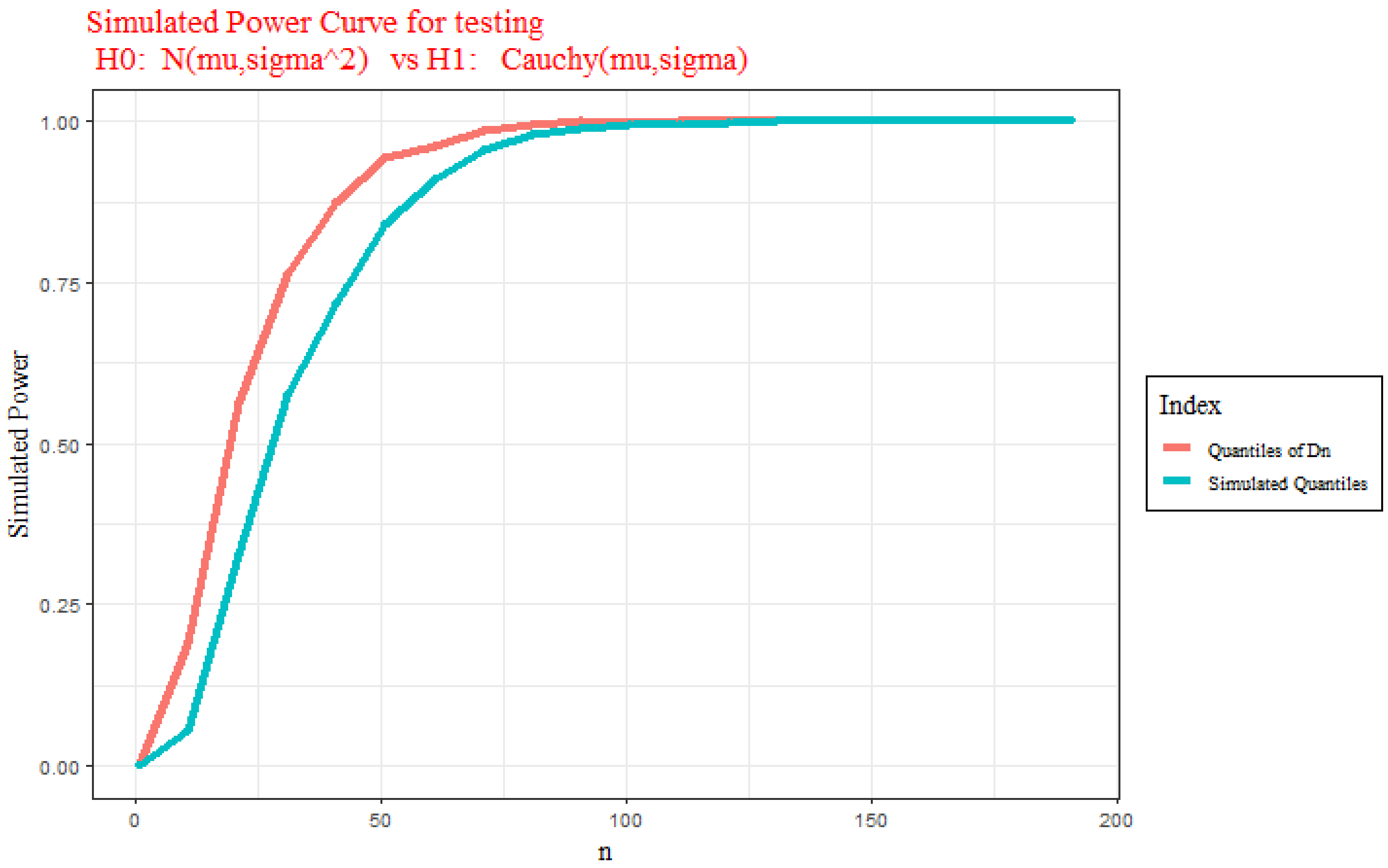
Null Hypothesis with both the Location and Scale unspecified :



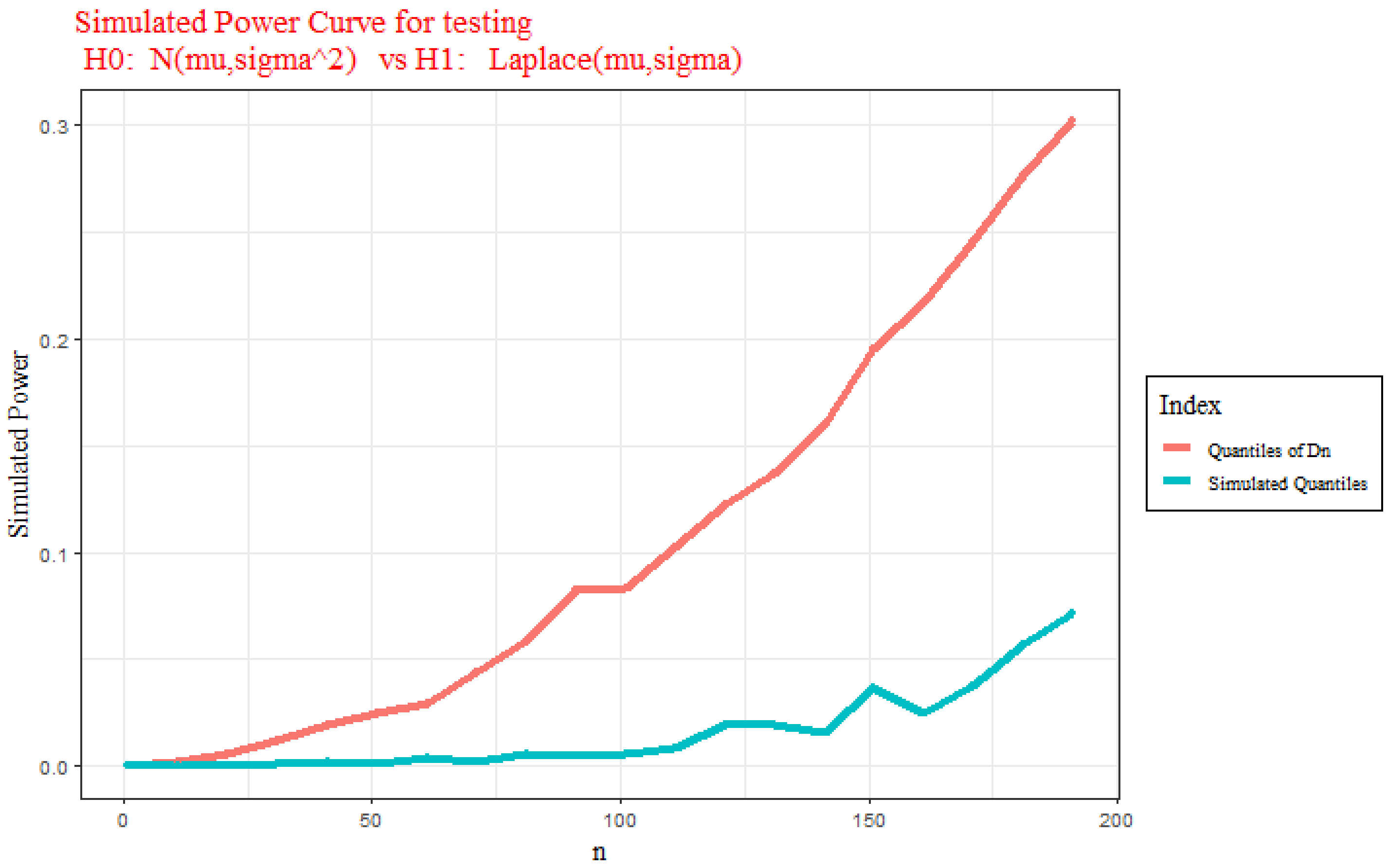
Convergence for D_n for Normal with unknown mean and Variance :



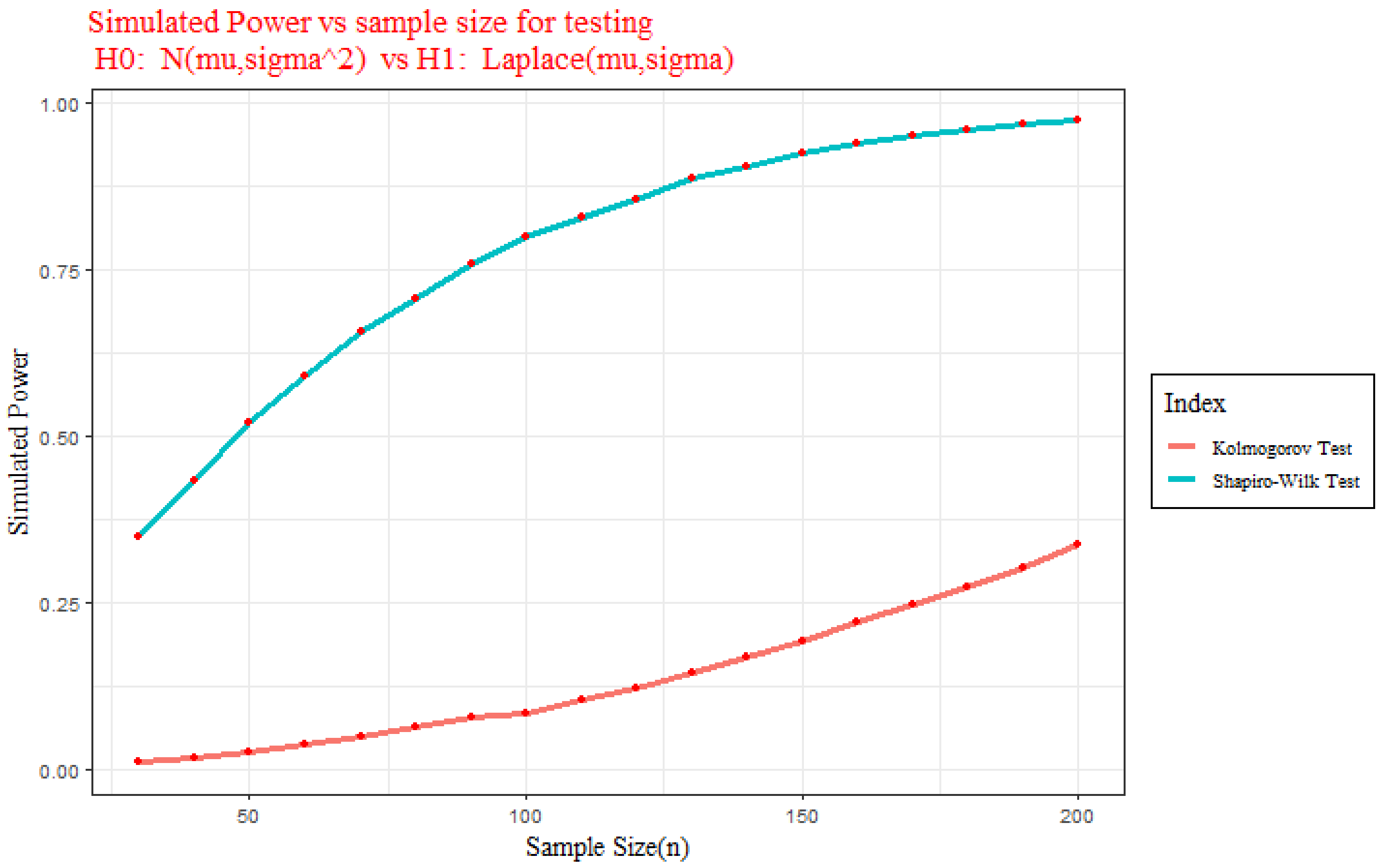
Power Function for partially specified Hypothesis :



If alternative is Laplace :

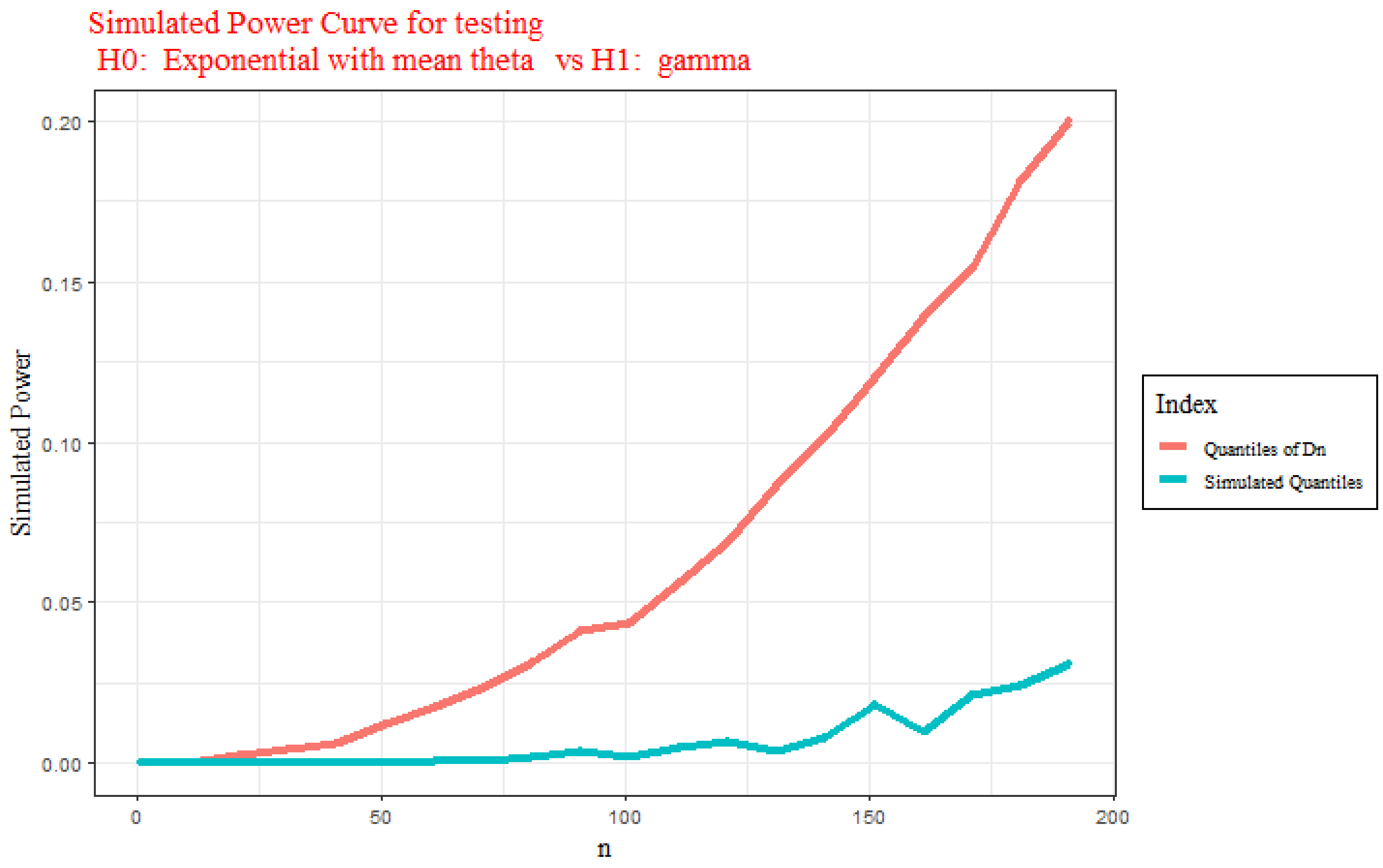


Test of Normality : Shapiro Wilk Test



Power Curve for the Partially Specified Case for Exponential

- Taking the alternative as Gamma we have the following empirical power curve



Violation of Independence Assumption of Kolmogorov-Smirnov Test :

- Kolmogorov-Smirnov Test assumes that $X_1, X_2, \ldots., X_n$ are **independent**.
- What if the random variables have identical distribution but they are **not independent**.
- Under the violation of independence assumptions, is D_n still distribution free ?
- We will consider three different situations !

Some Examples of Dependent Sequence of Random Variables :

> - Dependent Exponential(1) Random Variables :

- Consider sequence of independent **Exponential(1)** random variables $X_1, X_2, \ldots.$. Now, we construct the following sequence of random variables- Y_1 = X_1, Y_2 = $2Min(X_1, X_2)$,..... are dependent **Exponential(1)** r.v.

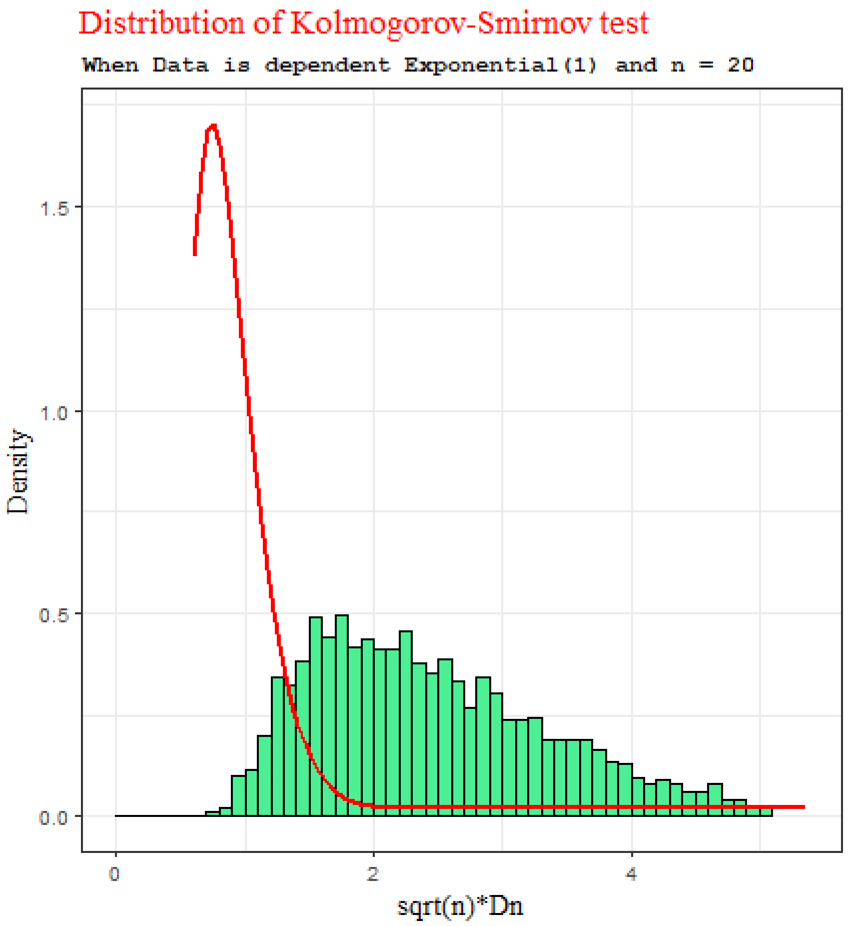
> - Dependent Normal(0,1) Random Variables :

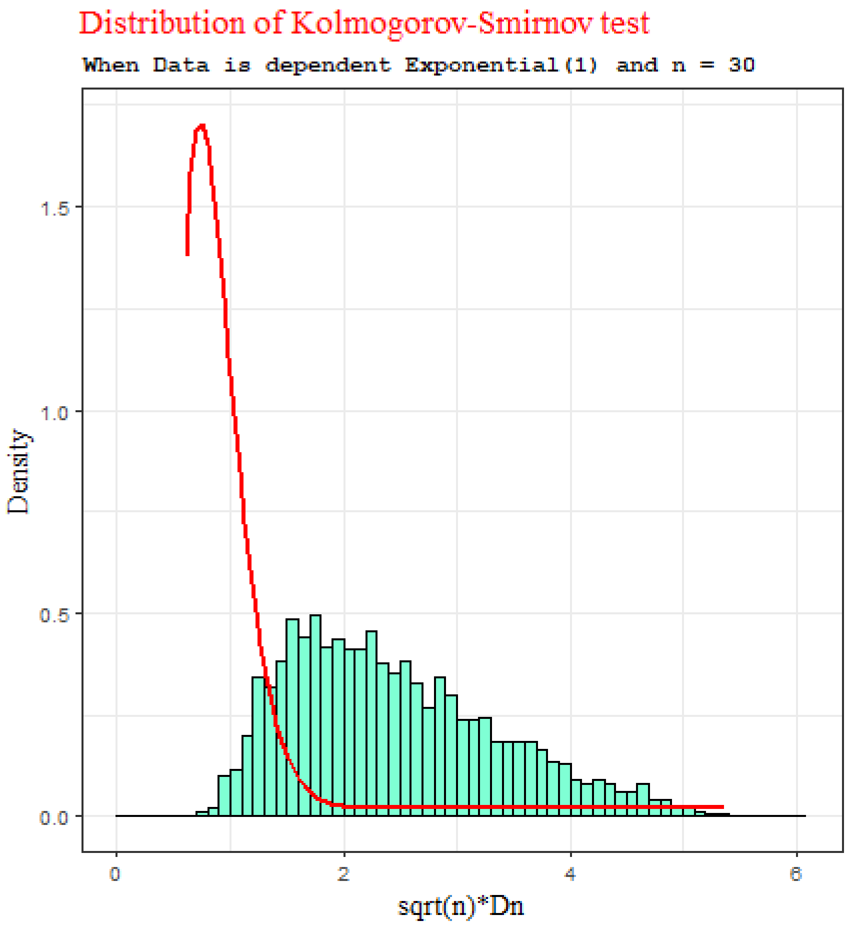
- Consider sequence of independent **Normal(0,1)** random variables $X_1, X_2, \ldots.$. Now, we construct the following sequence of random variables- Y_1 = X_1, Y_2 = $\frac{Y_1+Y_2}{\sqrt{2}}$,..... are dependent **Normal(0,1)** r.v.

> - Dependent Cauchy(0,1) Random Variables :

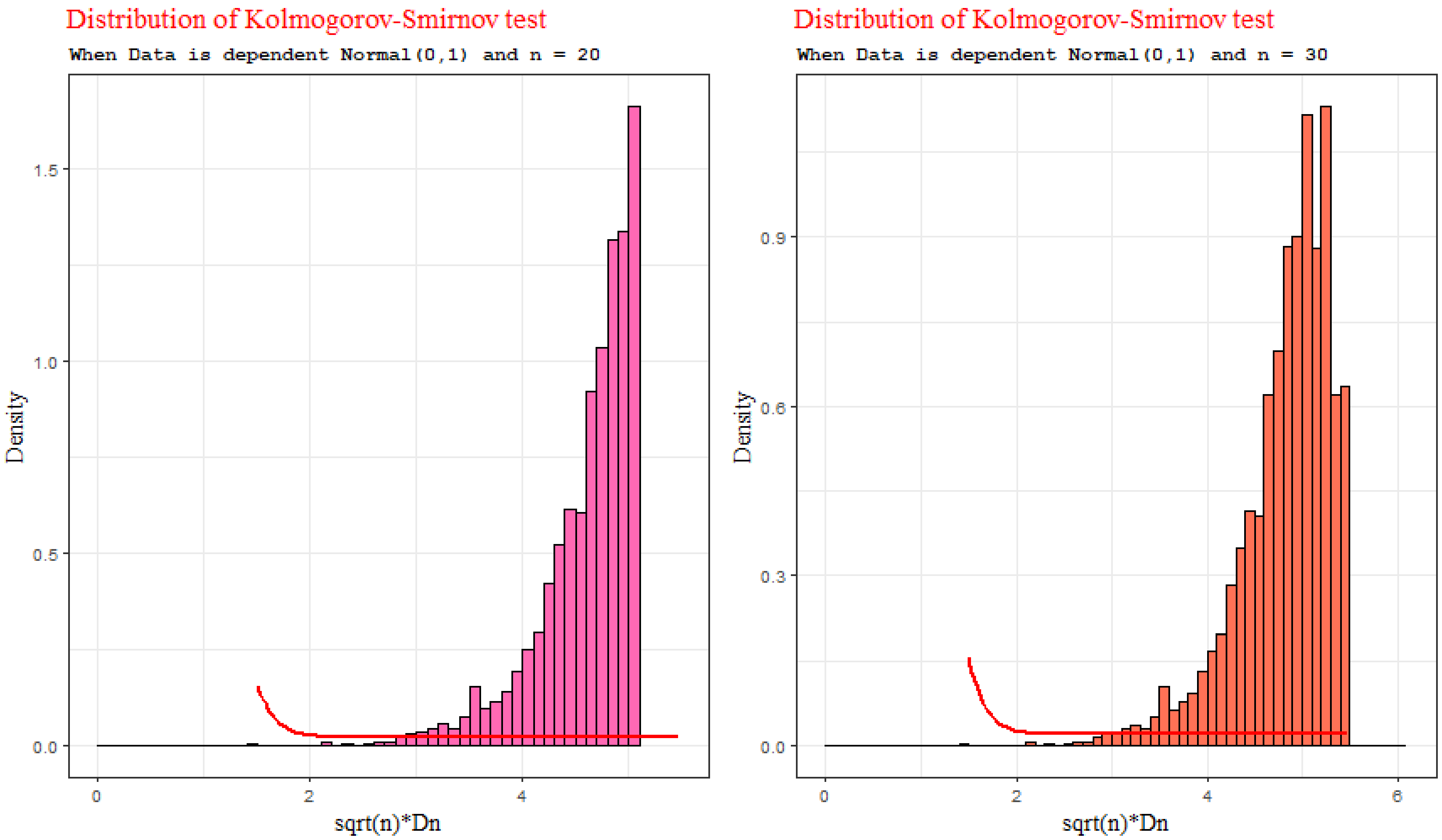
- Consider sequence of independent **Cauchy(0,1)** random variables $X_1, X_2, \ldots.$. Now, we construct the following sequence of random variables- Y_1 = X_1, Y_2 = $\frac{Y_1+Y_2}{2}$,..... are dependent **Cauchy(0,1)** r.v.

Dependent Exponential(1) Samples :

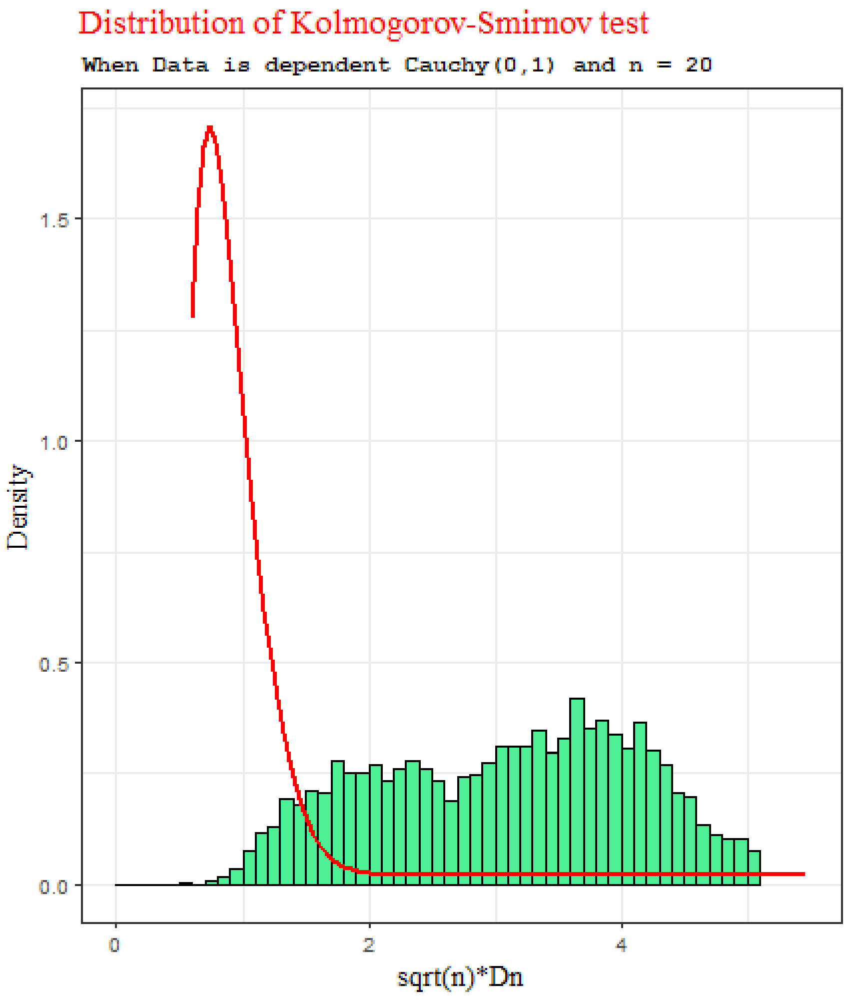


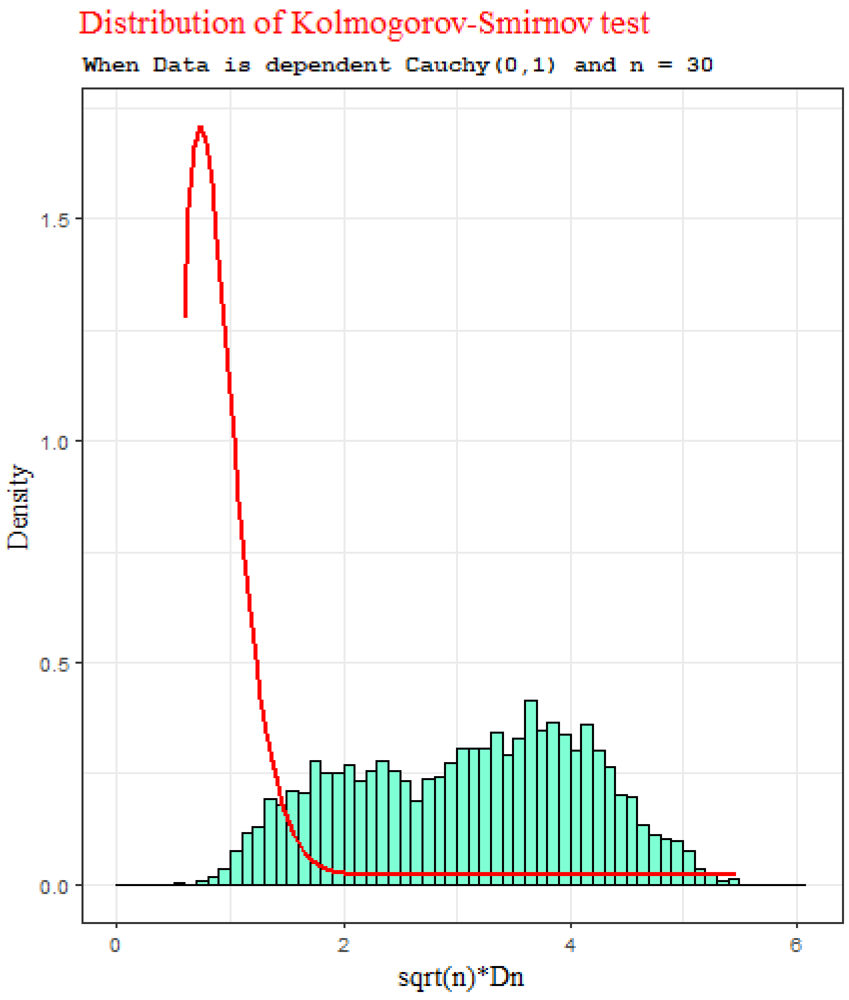


Dependent Normal(0,1) Samples :

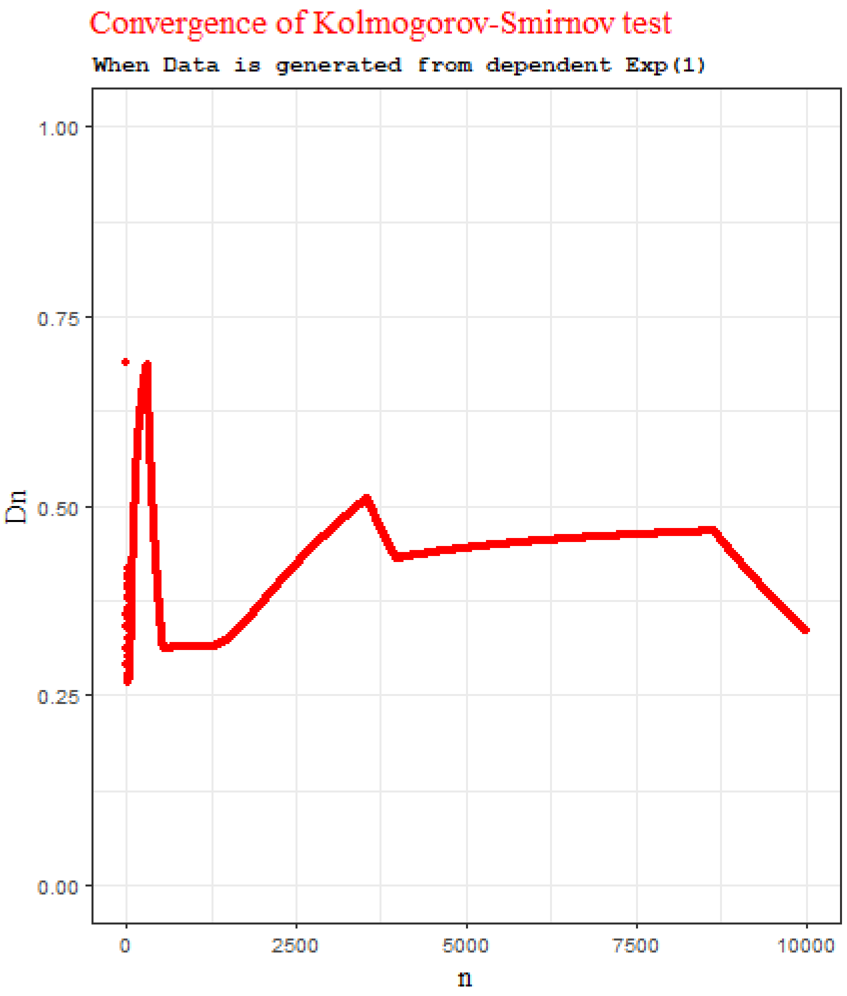


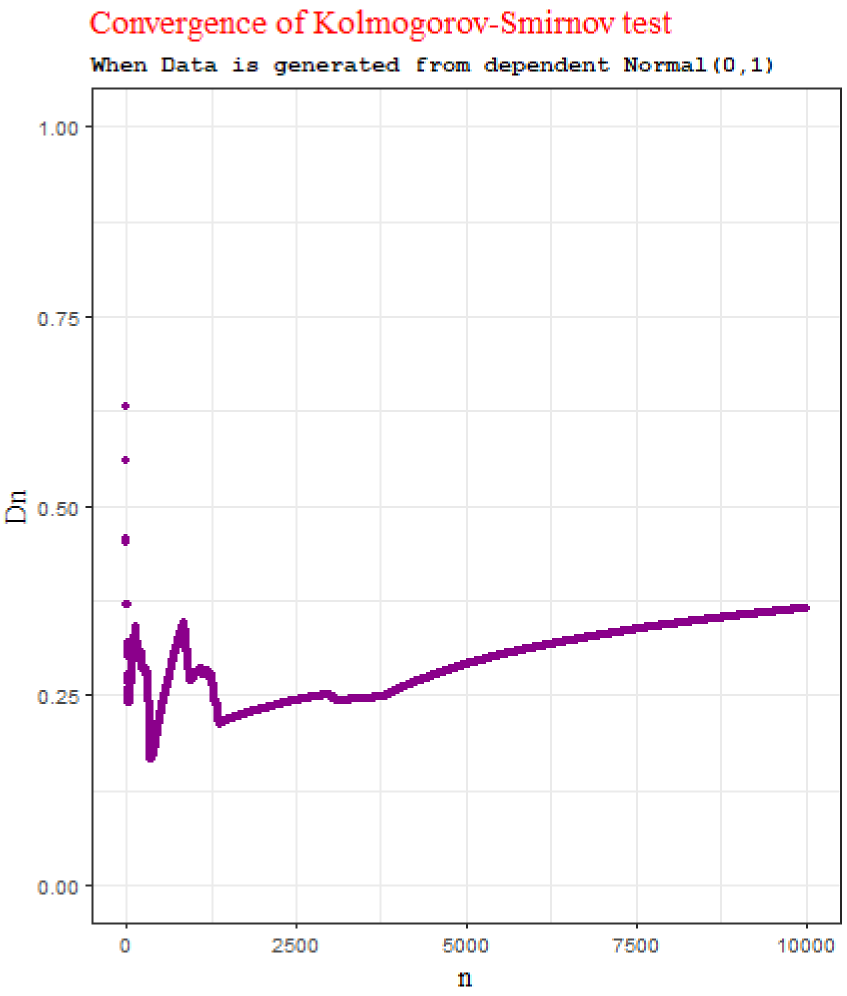
Dependent Cauchy(0,1) Samples :





Does Glivenko-Cantelli Hold here ?





What happens in Multivariate set up ?

- So far, in our discussion, we have talked about one-variate random variables, say $X_1, X_2, \ldots, X_n$.
- Now, one general question that may occur in our mind is that whether all these results are valid if we consider p-variate random vectors $\mathbf{X}_i$,for all $i = 1, 2, \ldots, n$.
- Does Glivenko Cantelli Theorem holds true?
- Does Same set up of Kolmogorov Test can be used in multivariate set up?
- Let us restrict ourselves in 2-variate case only.

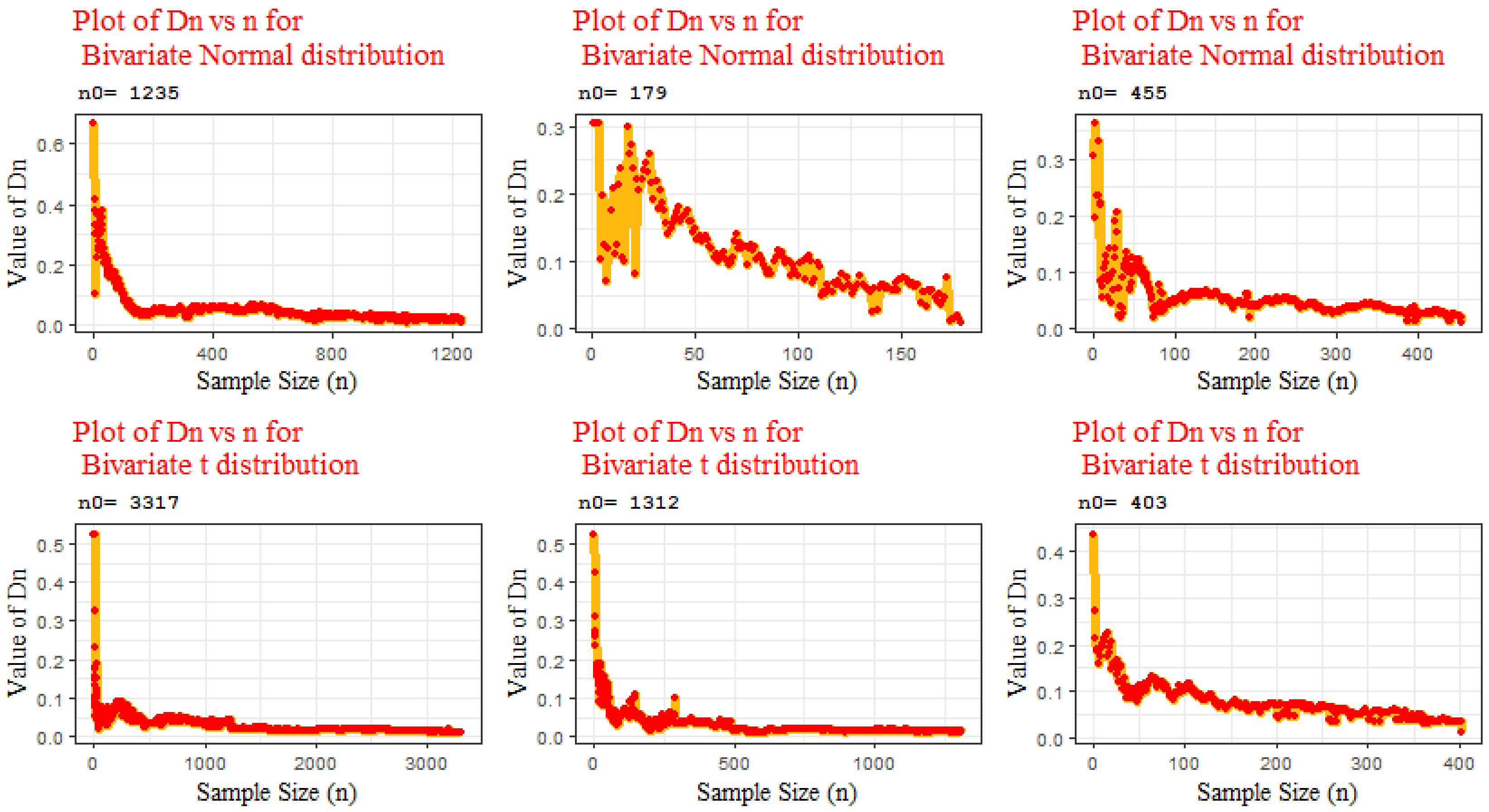
Does Glivenko Cantelli Theorem hold here ?

- Let us consider a sample from a bivariate population with Distribution Function $F(x, y)$ as $(X_1, Y_1), \ldots, (X_n, Y_n)$
- So, let us define the Empirical Distribution function as,

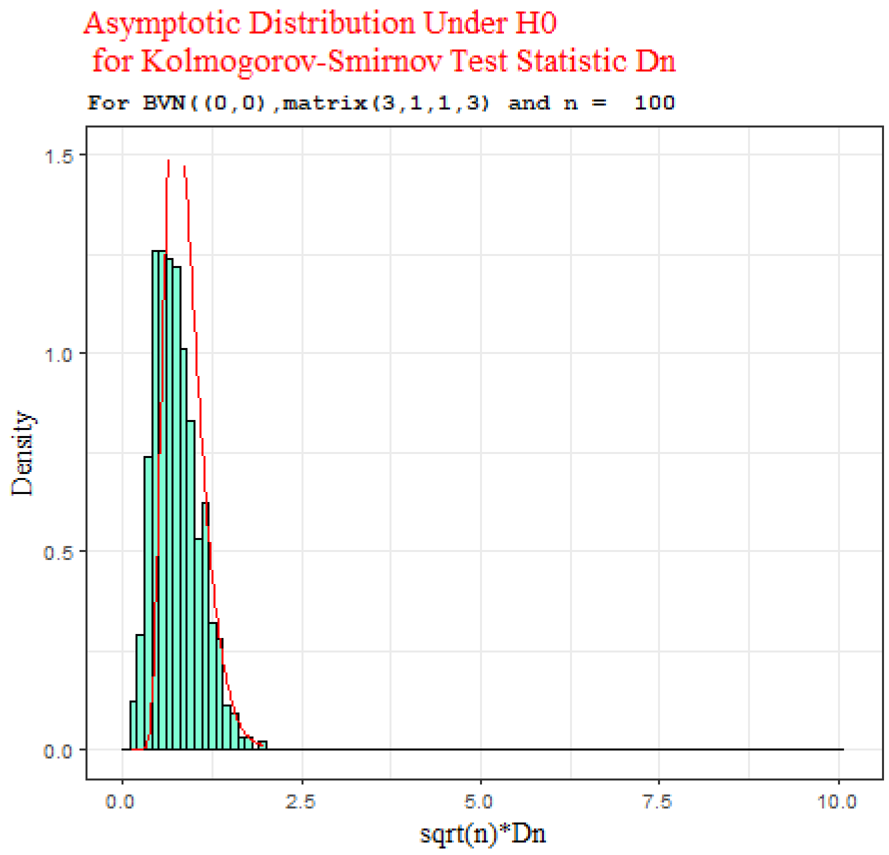
$$\mathbb{F}_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x; Y_i \leq y\}}$$

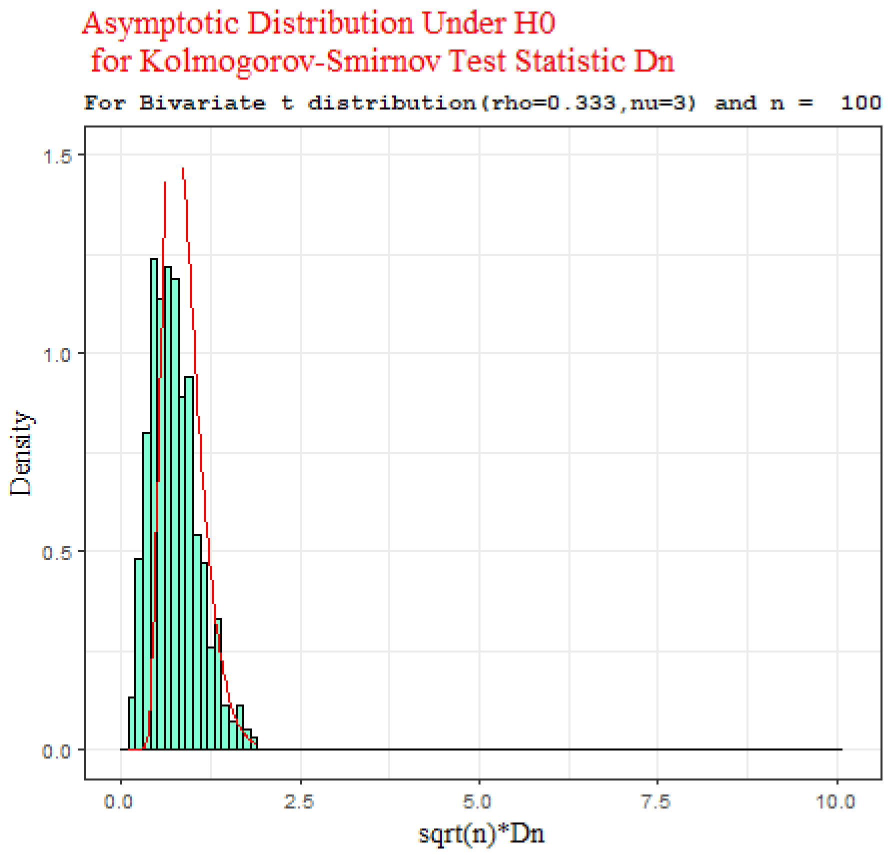
- Hence, the analogous Kolmogorv Statistic in this set up will be, $D_n = \sup_{x \in \mathbb{R}; y \in \mathbb{R}} |\mathbb{F}_n(x, y) - F(x, y)|$
- By Glivenko-Cantelli we have, $\mathbb{P}(\lim_{n \to \infty} D_n = 0) = 1$

Simulation to check Glivenko Cantelli analogous theorem :



Is still Kolmogorov-Smirnov Statistic(analogous) remaining distribution free under H0:





Empirical size for this case :

- H_0: $Bivariate\ t(\rho = 0.333, \nu = 3)$ vs H_1 : Bivariate Normal with same mean and dispersion

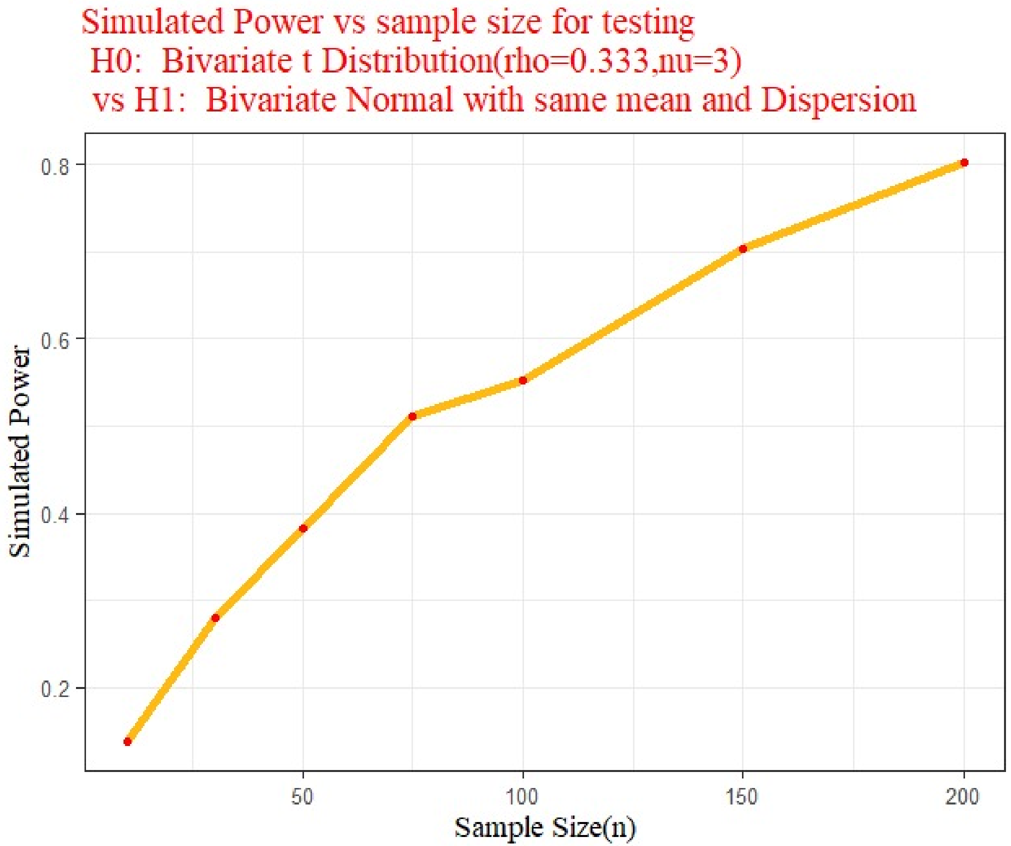
	Empirical Size
n = 25	0.027
n = 30	0.032
n = 40	0.035
n = 60	0.034

- H_0: Bivariate Normal Vs H_1: Bivariate t(parameters same as before)

	Empirical Size
n = 25	0.022
n = 30	0.027
n = 40	0.032
n = 60	0.037

- So, it seems that the test is **Conservative**

Empirical Power Curve for this Case :



- There is a paper on Ana Justel,Daniel Peña,Ruben H. Zamar : Multivariate Kolmogorov-Smirnov test.

Some Other Tests of Goodness of Fit and Comparison with Kolmogorov-Smirnov Test

Smoothed Kernel Type Kolomogorov-Smirnov Statistic :

- We have already an idea about Kernel Density Estimation.
- But here, We are interested in **Kernel CDF Estimation**.
- If Kernel Density estimator is defined as -

$$\hat{f}_n(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

- Then, Kernel CDF estimator is defined as -

$$\hat{F}_n(x,h) = \frac{1}{n}\sum_{i=1}^{n} W(\frac{x - X_i}{h})$$

- Where, $W(x)$ = $\int_{-\infty}^{x} K(y)\ dy$
- In kernel density estimation also, we have some boundary problems. Similarly here also we have boundary problems.

Dealing with Boundary Problems :

- With Little Modifications we will achieve smoothed kernel type estimator :

$$\hat{f}_n(x; h, t) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{t(x) - t(X_i)}{h}) t^{'}(x)$$

- Where $t(.)$ is "good" and "Well-defined" function corresponding to the problem.
- Following the above notation Kernel CDF estimator is defined as -

$$\hat{F}_n(x, h) = \frac{1}{n} \sum_{i=1}^{n} W(\frac{t(x) - t(X_i)}{h})$$

- Where, $W(.)$ is defined before.
- We will use a Nils Lid Hjort, Ingrid K. Glad : Parametrically guided kernel density estimation approach for choosing optimal kernel and bandwidth.
- It's implementation is available in kdensity package in R.

The Test-Statistic and it's distribution :

- The test statistic is given by -

$$\tilde{D_n} = Sup_{x \in \mathbb{R}} |\tilde{\mathbb{F}}_n(x) - F_o(x)|$$

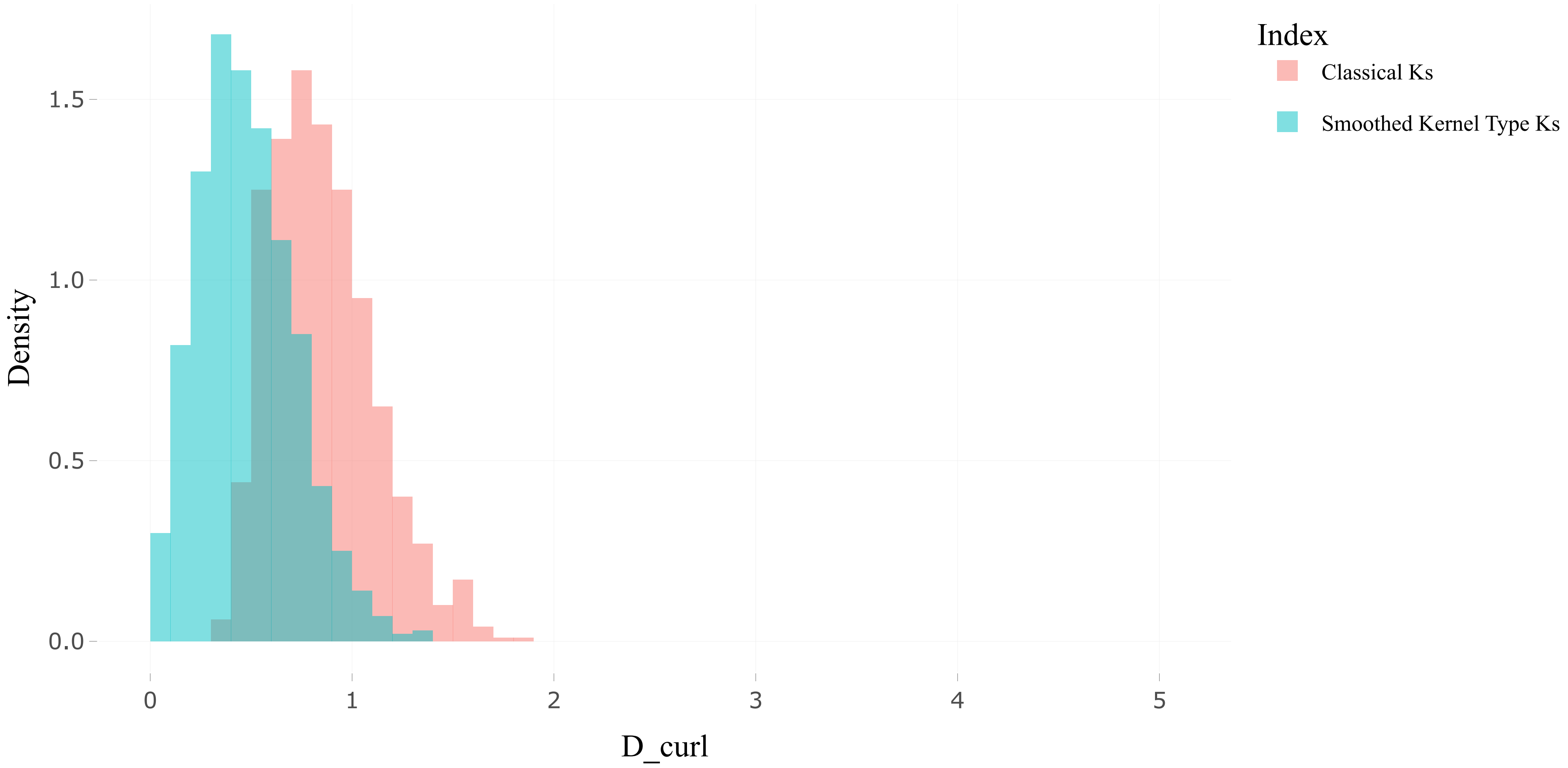
- $\sqrt{n}\tilde{D_n} \xrightarrow{\mathbb{P}} 0$ as, $n \to \infty$
- Also, It's Asymptotic distribution is given by -

$$\lim_{n \to \infty} P(\sqrt{n}\tilde{D_n} \leq x) = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} exp[\frac{-(2i-1)^2\pi^2}{8x^2}]$$

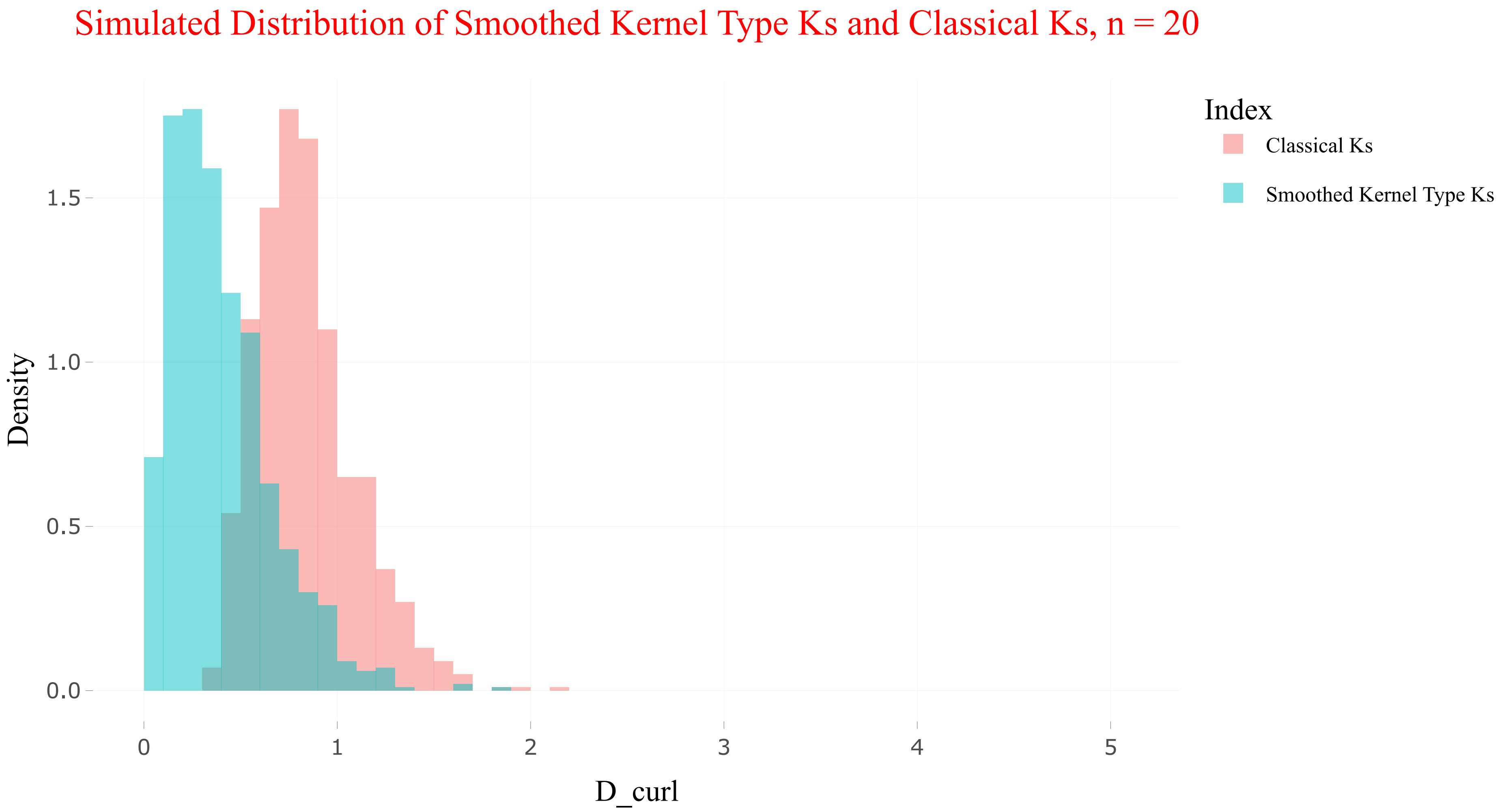
> - The main paper is Rizky Reza Fauzi,Maesono Yoshihiko : Kolmogorov-Smirnov Test Based on Kernel Estimation.

For Normal(0,1) Distribution :

Simulated Distribution of Smoothed Kernel Type Ks and Classical Ks, n = 20 for N(0,1)



For Exponential(1) Distribution :



Some Power Results and Comparison with Classical Kolomogorov-Smirnov Test (n = 50)

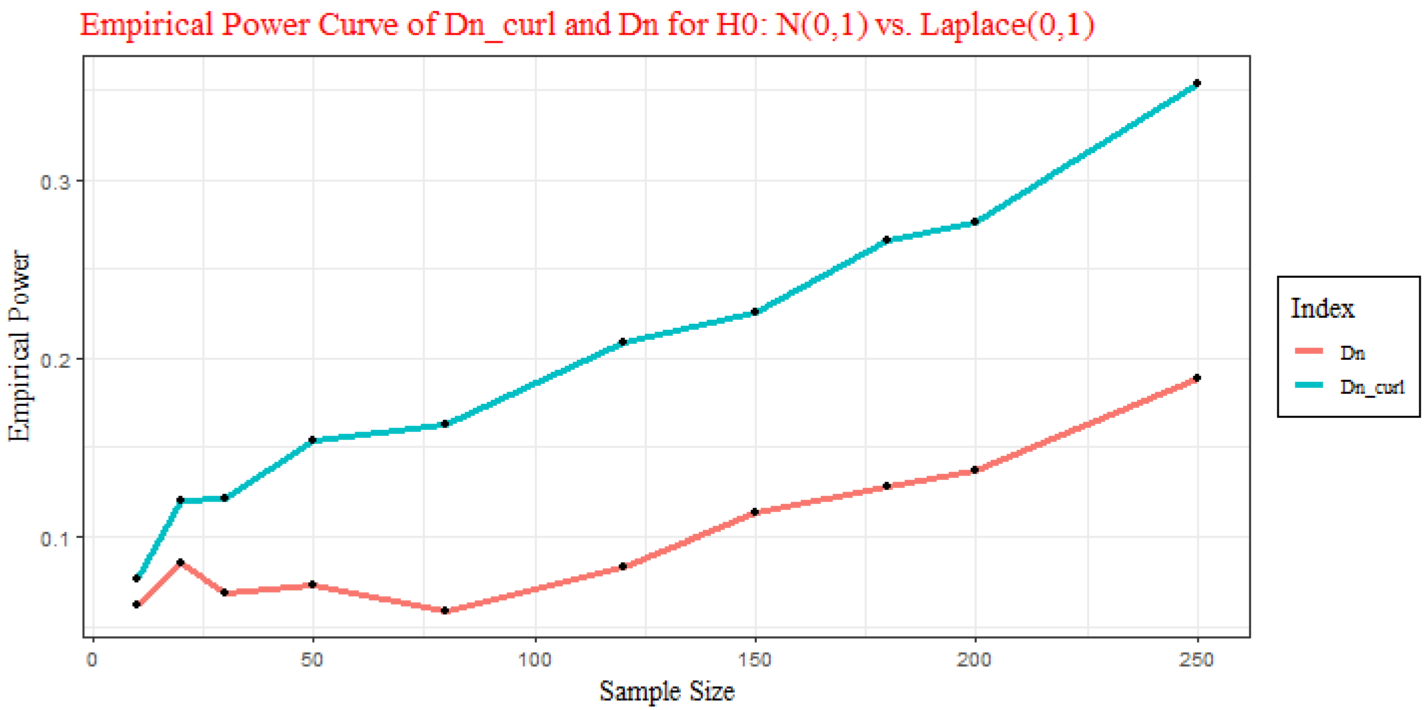
- Simulated Probability of rejecting H_0 for $\tilde{D}_n$

	Exp(1/2)	**Gamma(3,2)**	**Abs N(0,1)**	**Log N(0,1)**
Exp(1/2)	0.05	0.934	0.957	0.976
Gamma(3,2)	0.834	0.051	0.872	0.836
Abs N(0,1)	0.951	0.936	0.05	0.981
Log N(0,1)	0.871	0.829	0.895	0.05

- Simulated Probability of rejecting H_0 for D_n

	Exp(1/2)	**Gamma(3,2)**	**Abs N(0,1)**	**Log N(0,1)**
Exp(1/2)	0.051	0.746	0.855	0.724
Gamma(3,2)	0.887	0.05	0.851	0.834
Abs N(0,1)	0.784	0.748	0.051	0.878
Log N(0,1)	0.862	0.83	0.891	0.052

Empirical Power Curve for H_0 : X ~ N(0,1) vs. H_1 : X ~ Laplace(0,1) :



- Not much good !

Cramér–von Mises test

- For the same hypothesis that we test in two-sided Kolmogorov Test,i.e., Given $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} F$, for testing $H_0 : F = F_0$ vs $H_1 : F \neq F_0$, where F is the distribution function associated with the random variables.
- This test statistic uses **Quadratic Distance** between $F_n(x)$ and $F_0(x)$.

$$W_n^2 := n \int (\mathbb{F}_n(x) - F_0(x))^2 dF_o(x)$$

- If H_0 is true, this statistic tends to be small. So, we need to reject H_0 for large values of W_n^2.
- This statistic can be further simplified as:

$$W_n^2 = \sum_{i=1}^{n} \left\{ U_{(i)} - \frac{2i-1}{2n} \right\}^2 + \frac{1}{12n}$$

where $U_{(j)}$ stands for j^{th} sorted $U_i = F_0(X_i)$

- **Distribution under H_0 :** If H_0 holds and F_0 is continuous, Then W_n^2 has an asymptotic distribution with CDF given by,

$$\lim_{n\to\infty} \mathbb{P}(W_n^2 \leq x) = 1 - \frac{1}{\pi}\sum_{j=1}^{\infty}(-1)^{j-1}W_j(x)$$

where,

$$W_j(x) := \int_{(2j-1)^2\pi^2}^{4j^2\pi^2} \sqrt{\frac{-\sqrt{y}}{sin\sqrt{y}}}\frac{e^{-\frac{xy}{2}}}{y}dy$$

- This test is distribution free if F is continuous and the sample has no ties.

Anderson–Darling test

- For the same hypothesis that we test in two-sided Kolmogorov Test,i.e., Given $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} F$, for testing $H_0 : F = F_0$ vs $H_1 : F \neq F_0$, where F is the distribution function associated with the random variables.

- This statistic uses **Weighted Quadratic distance** between $F_n(x)$ and $F_0(x)$ weighted by $w_0(x)$ = $F_0(x)(1 - F_0(x))^{-1}$.

$$A_n^2 := n \int \frac{(F_n(x) - F_o(x))^2}{F_o(x)(1 - F_o(x))} dF_o(x)$$

- If H_0 is true, this statistic tends to be small. So, we need to reject H_0 for large values of A_n^2.
- It can be noted that, compared to W_n^2, A_n^2 puts more weights on the deviation between $F_n(x)$ and $F_0(x)$ that happens on the tail, i.e. when $F_0(x) \approx 0$ or $F_0(x) \approx 1$

- This statistic can be further simplified as:

$$A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)log(U_{(i)}) + (2n+1-2i)log(1-U_{(i)})\right\}$$

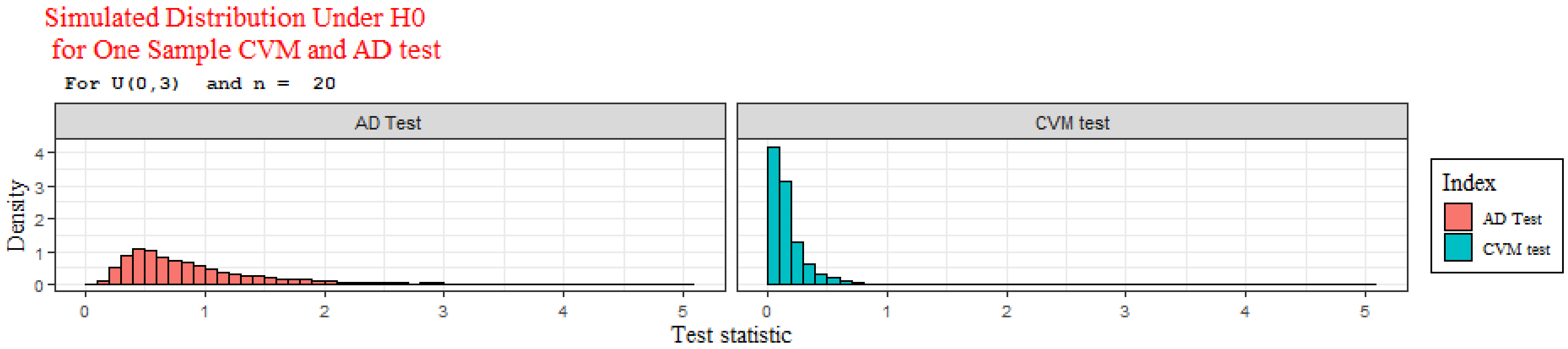
- **Distribution under H_0 :**

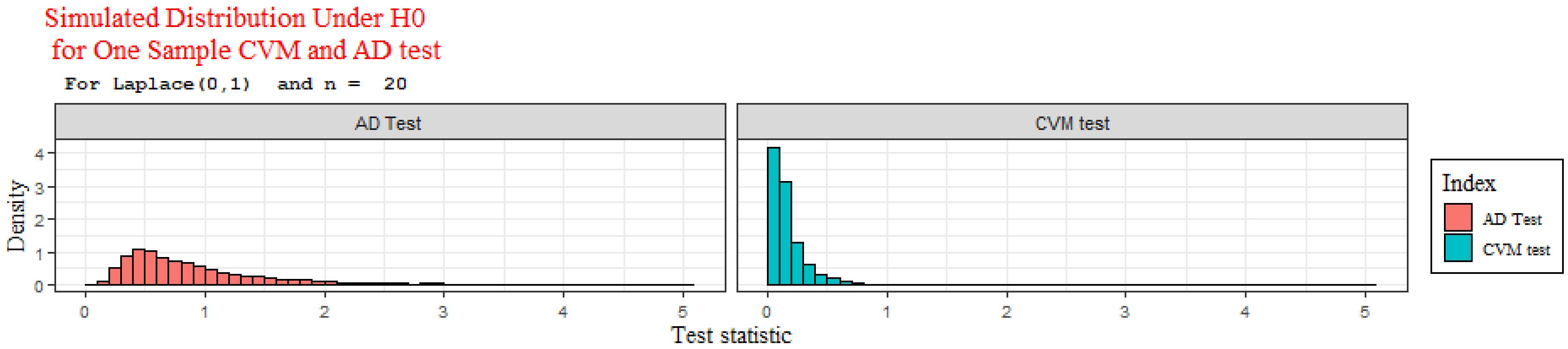
If H_0 holds and F_0 is continuous, Then A_n^2 has an asymptotic distribution given by,

$$\sum_{j=1}^{\infty}\frac{Y_j}{j(j+1)}, \text{where } Y_j \sim \chi_1^2, j \geq 1, \text{are iid}$$

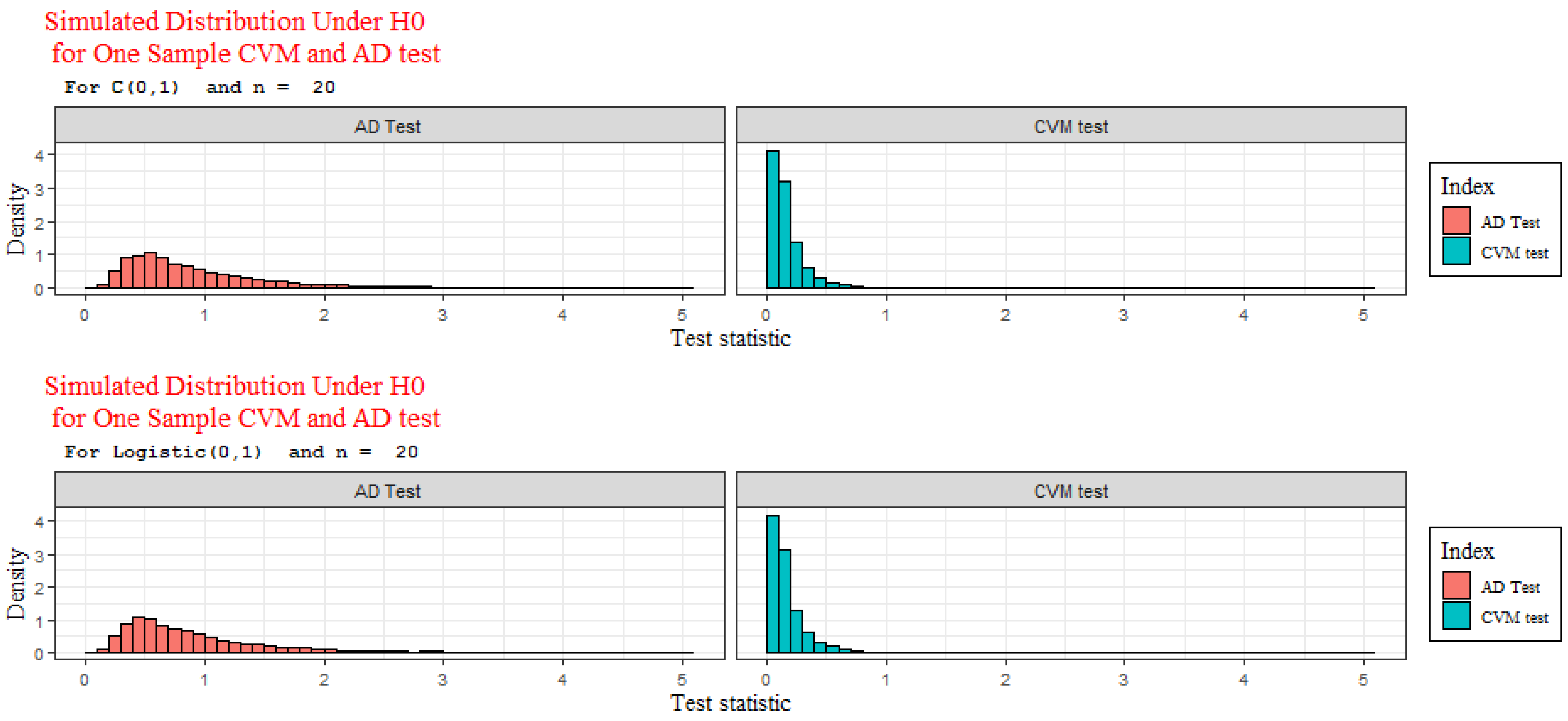
- Reference : Eduardo García-Portugués : Goodness-of-fit tests for distribution models

Simulated Distribution of W_n^2 and A_n^2 (sample size,n=20):

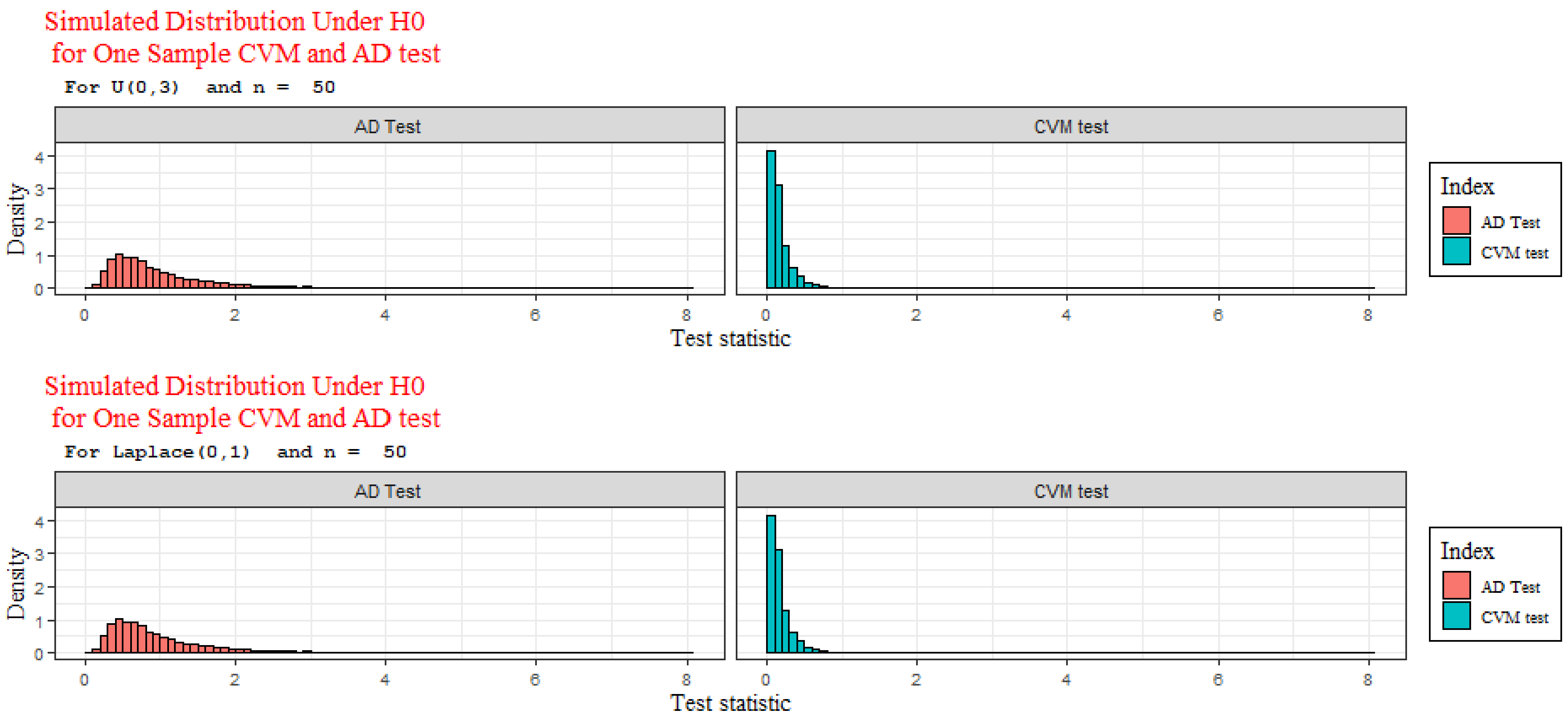




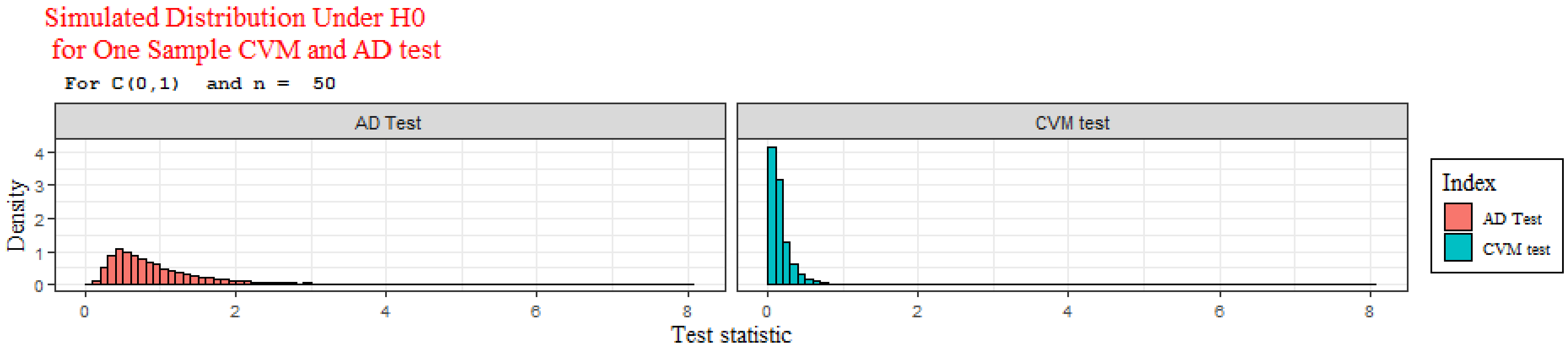
Simulated Distribution of W_n^2 and A_n^2 (sample size,n=20):

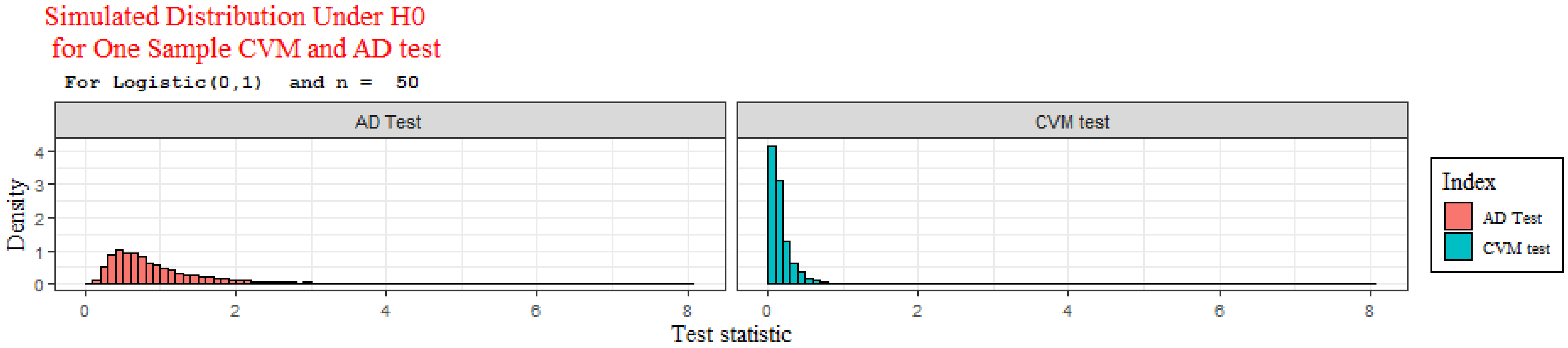


Simulated Distribution of W_n^2 and A_n^2 (sample size,n=50):

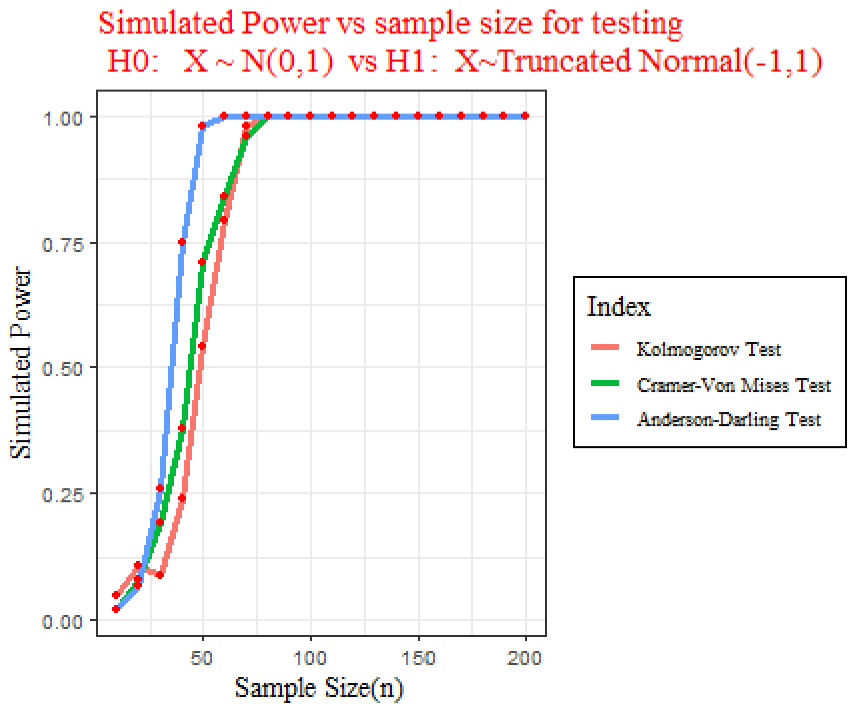


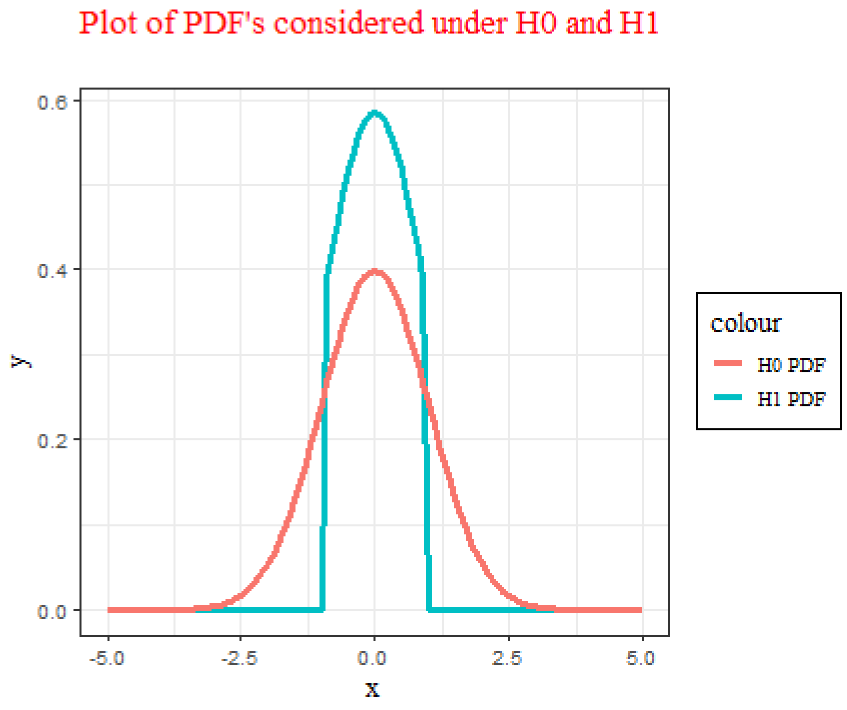
Simulated Distribution of W_n^2 and A_n^2 (sample size,n=50):



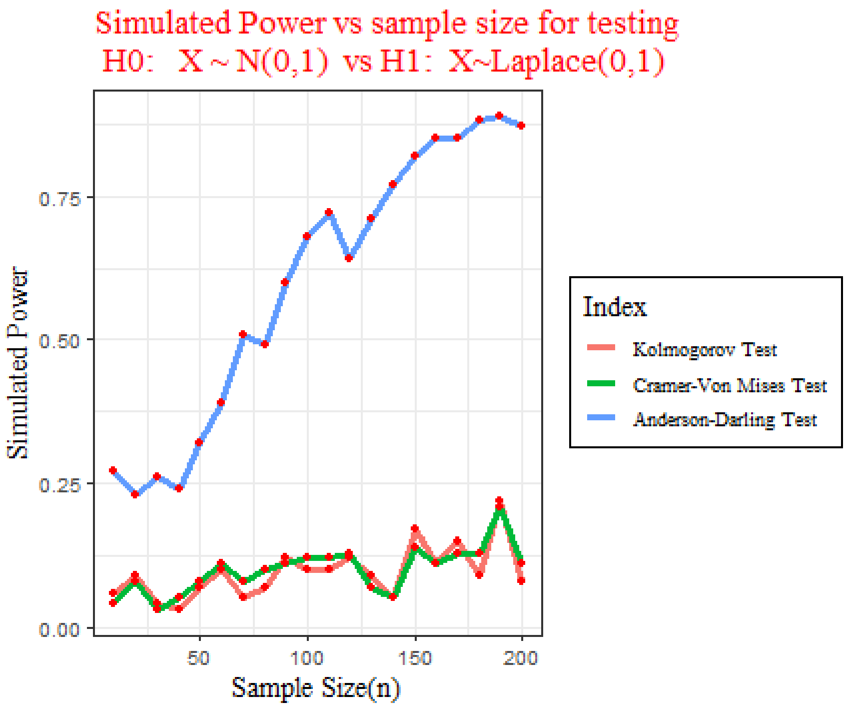


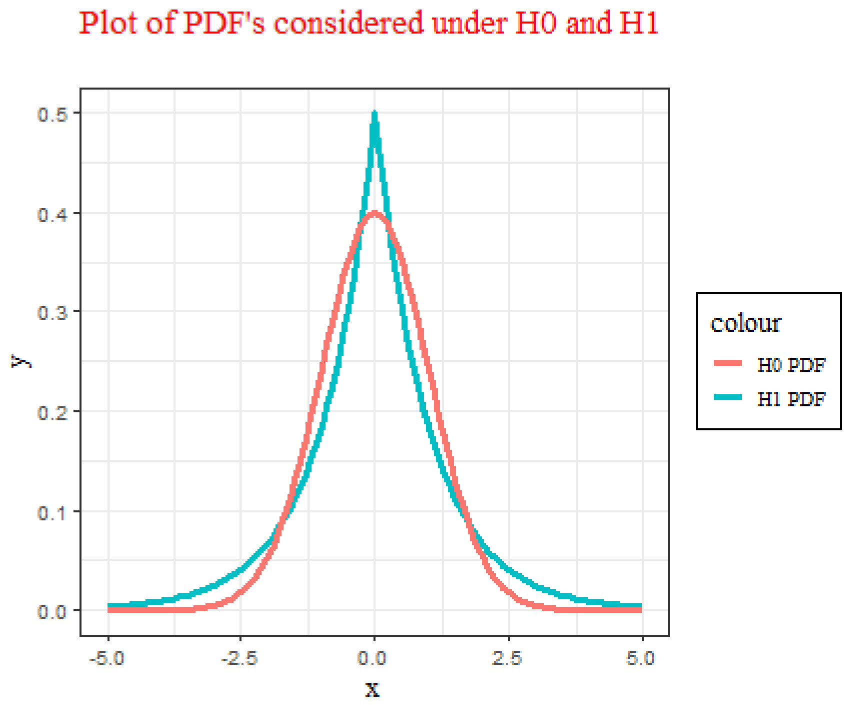
Power Comparison between Kolmogorov-Smirnov,Cramer-Von Mises Test,Anderson-Darling Test :



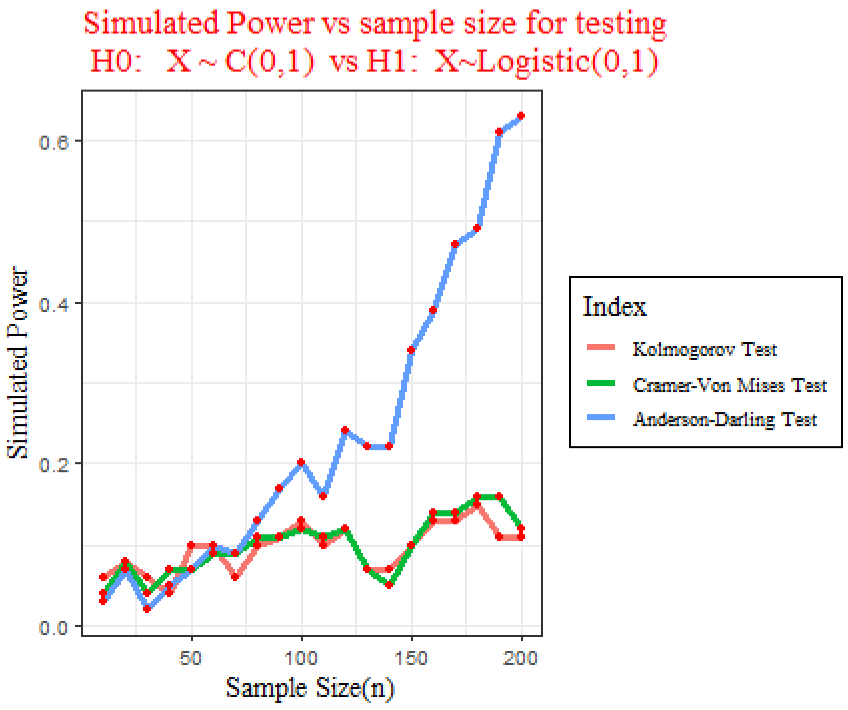


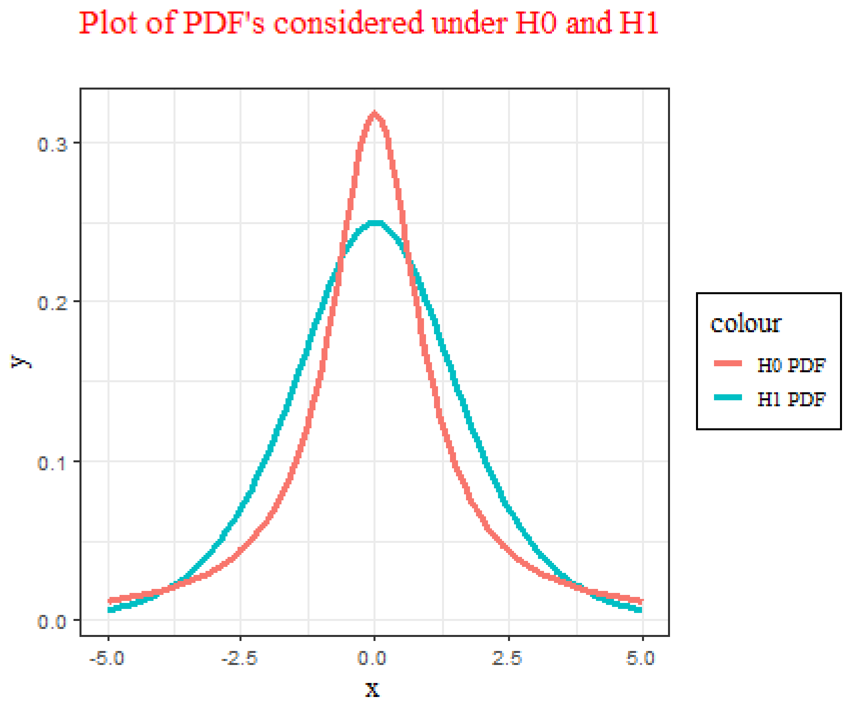
Power Comparison between Kolmogorov-Smirnov,Cramer-Von Mises Test,Anderson-Darling Test :





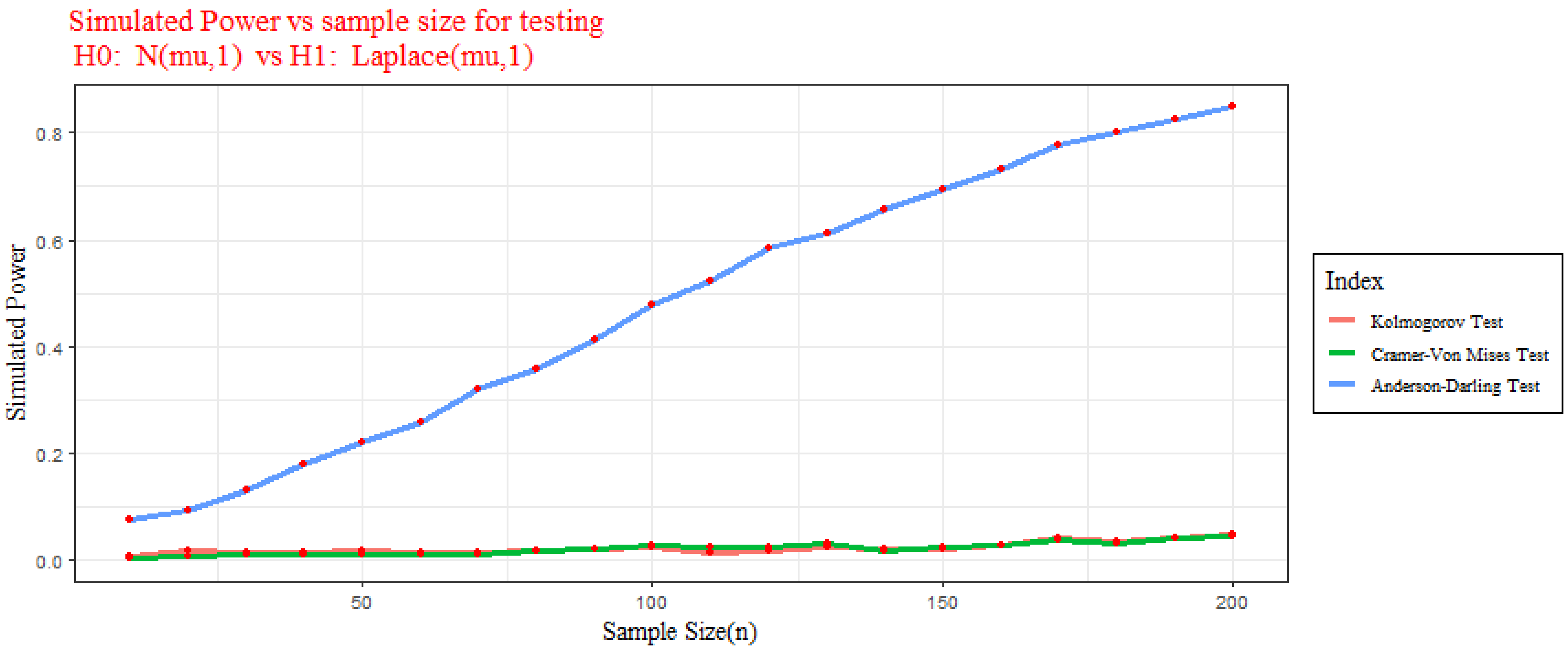
Power Comparison between Kolmogorov-Smirnov,Cramer-Von Mises Test,Anderson-Darling Test :



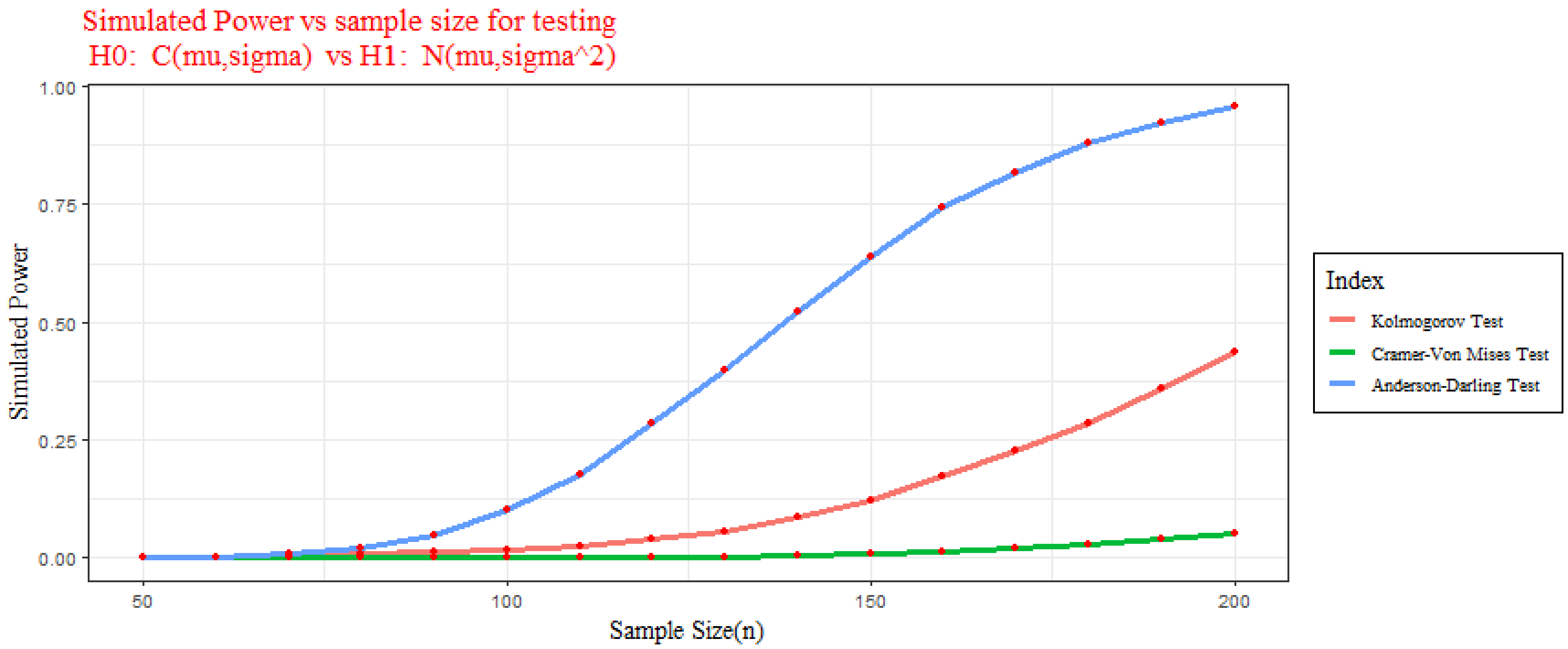


- Clearly, Anderson Darling is performing much better than the other two in both the cases.

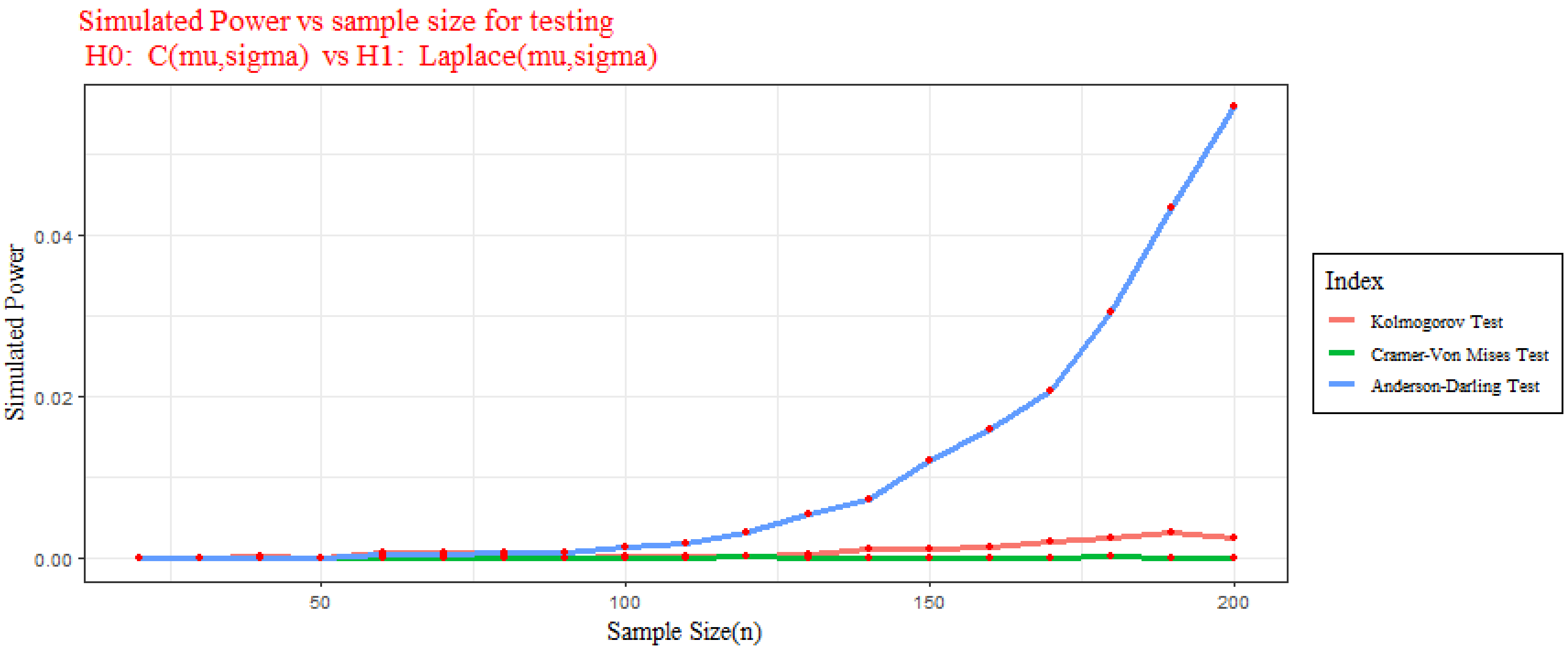
What happens in Partially Specified Case?



What happens in Partially Specified Case?



What happens in Partially Specified Case?



Berk-Jones Test

- In a 1979 paper, Berk and Jones suggested an intuitively appealing method of testing simple goodness-of-fit null hypothesis.
- The Berk-Jones method is just transform the entire goodness-of-fit problem to a Binomial testing problems.
- The key fact that it uses is, if F is the underlying true CDF then for every $x \in \mathbb{R}$, $n\mathbb{F}_n(x) \sim Binomial(n, F(x))$.
- So, for given null hypothesis $\mathcal{H}_o : F = F_o$, for every x what we really want to check is $p(x) = p_o(x)$ where $p(x) = \mathbb{P}(X \leq x)$.
- We can use a likelihood ratio test corresponding to two-sided or one-sided alternative to this hypothesis. Hence we need to maximize the binomial likelihood function with respect to $F(x)$ for every value of $x \in \mathbb{R}$.

Constructing the Likelihood Ratio Test

- The maximum likelihood of $F(x)$ is calculated to be $\mathbb{F}_n(x)$. Hence we have the likelihood ratio statistic ,-

$$\lambda_n(x) = \frac{\mathbb{F}_n(x)^{n\mathbb{F}_n(x)}(1-\mathbb{F}_n(x))^{n-n\mathbb{F}_n(x)}}{\mathbb{F}_o(x)^{n\mathbb{F}_n(x)}(1-\mathbb{F}_o(x))^{n-n\mathbb{F}_n(x)}}$$

$$= \left(\frac{\mathbb{F}_n(x)}{F_o(x)}\right)^{n\mathbb{F}_n(x)} \left(\frac{(1-\mathbb{F}_n(x))}{(1-F_o(x))}\right)^{n-n\mathbb{F}_n(x)}$$

- But since we want to check whether $F(x) = F_o(x)$ for all $x \in \mathbb{R}$. So it makes sense we take the supremum of λ_n over all $x \in \mathbb{R}$.

- Hence the Berk-Jones Statistics is,

$$R_n = n^{-1} \sup_{x\in\mathbb{R}} log(\lambda_n(x)).$$

Berk-Jones Statistic

- The interesting thing about the statistic R_n is it's connection to the Kullback-Leibler Distance between two Binomial populations.
- The Kulback-Liebler Distance between two distributions $Binomial(n, p)$ and $Binomial(n, \theta)$ is defined as,

$$K(p, \theta) = plog\left(\frac{p}{\theta}\right) + (1-p)log\left(\frac{1-p}{1-\theta}\right).$$

- Hence we can write,

$$R_n = \sup_{x \in \mathbb{R}} K(\mathbb{F}_n(x), F(x)).$$

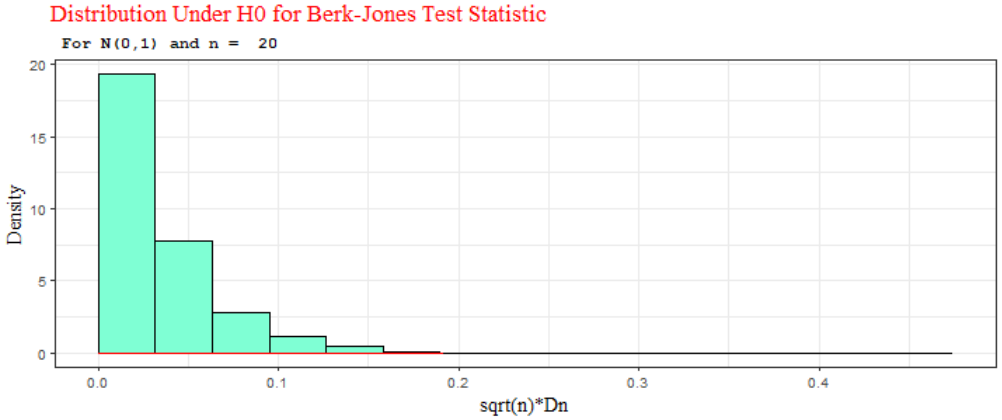
- Hence we reject $\mathcal{H}_o$, for large values of R_n.
- But computing this statistic in R is very difficult because of the $Supremum$.
- Referring back to the original paper of Berk and Jones(1970) and a recent paper Amit Moscovich and Boaz Nadler(2016), we see that instead of working with R_n, we can work with the Likelihood Ratio Statistic $\lambda_n(x)$ itself taking the arguments as a Order statistics of the sample, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

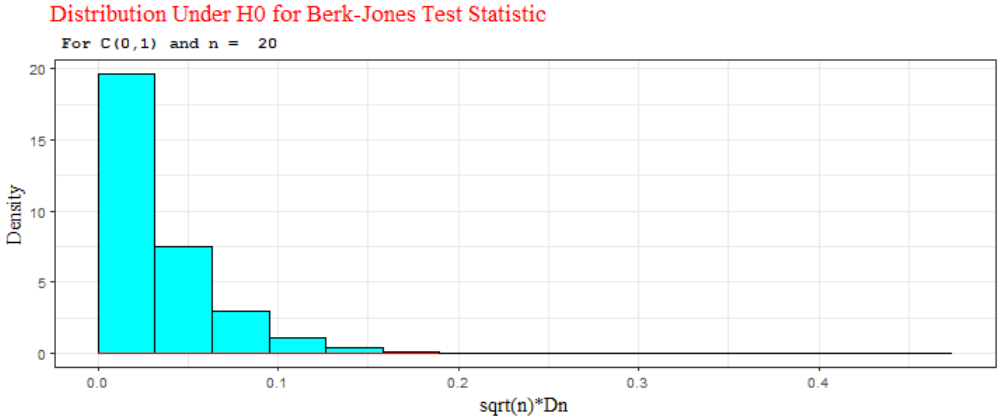
Computing the Exact Berk-Jones Statistic

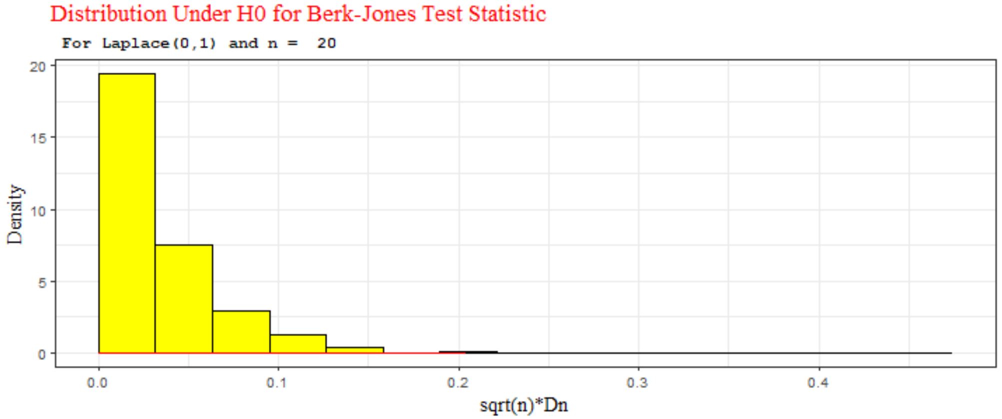
- Using the fact that under $\mathcal{H}_o$, we know $U_i = F_o(X_i) \overset{i.i.d}{\sim} Unif(0,1)$, for F_o continuous.
- Hence, $U_{(1)}, U_{(2)}, \ldots., U_{(n)}$ are order statistics from a sample of size n from $Unif(0,1)$.
- Berk-Jones(1970) showed that R_n and $-n^{-1}log(M_n)$ has same asymptotic properties. Hence,testing with respect to R_n and M_n are equivalent.
- Where $M_n = min(M_n^+, M_n^-)$ where, $M_n^+ := min_{1 \le i \le n} \mathbb{P}(Beta(i, n-i+1) < u_{(i)})$ and $M_n^- := min_{1 \le i \le n}(1 - \mathbb{P}(Beta(i, n-i+1) < u_{(i)}))$.
- Hence, we calculate this M_n which is infact the called the $Exact Berk - Jones Statistic$.

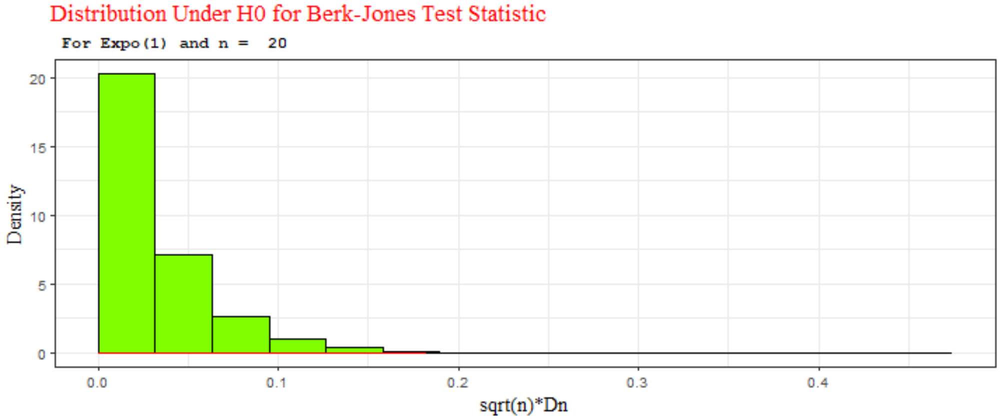
Distribution Free Nature of M_n :

- Here, we present the results of some simulation studies which exhibits the Distriution Free nature of M_n.

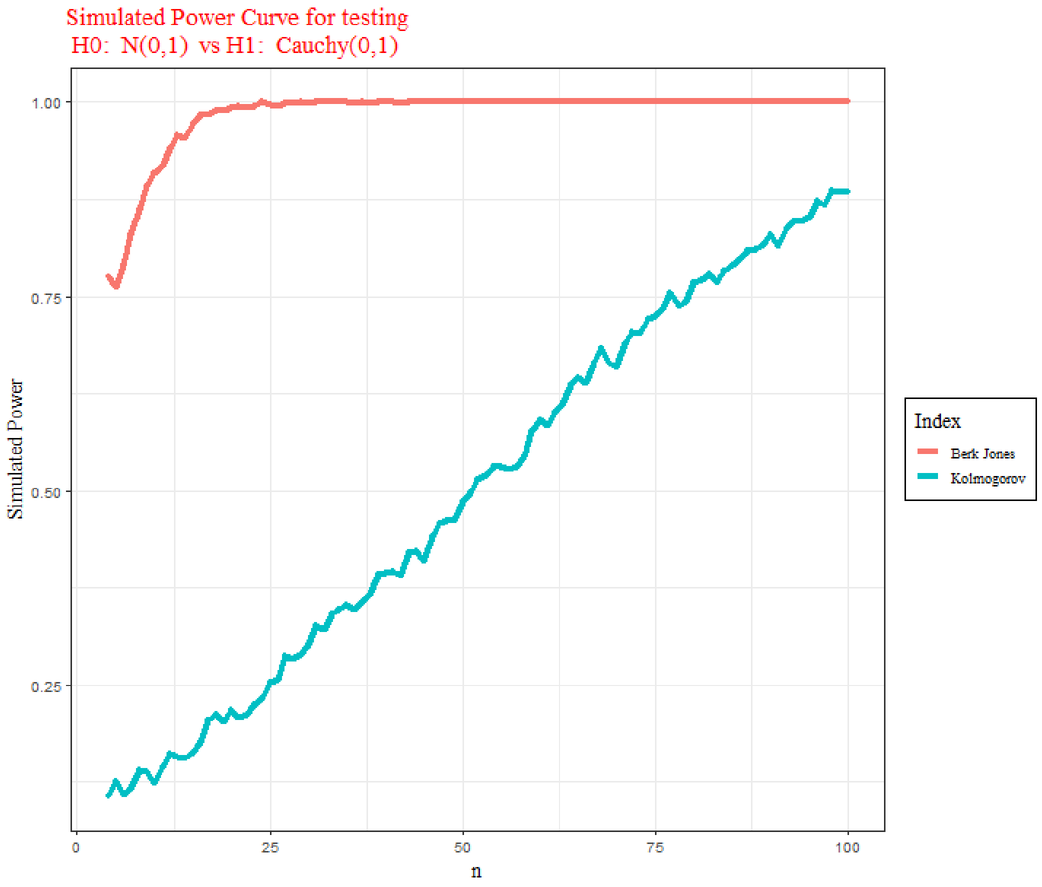


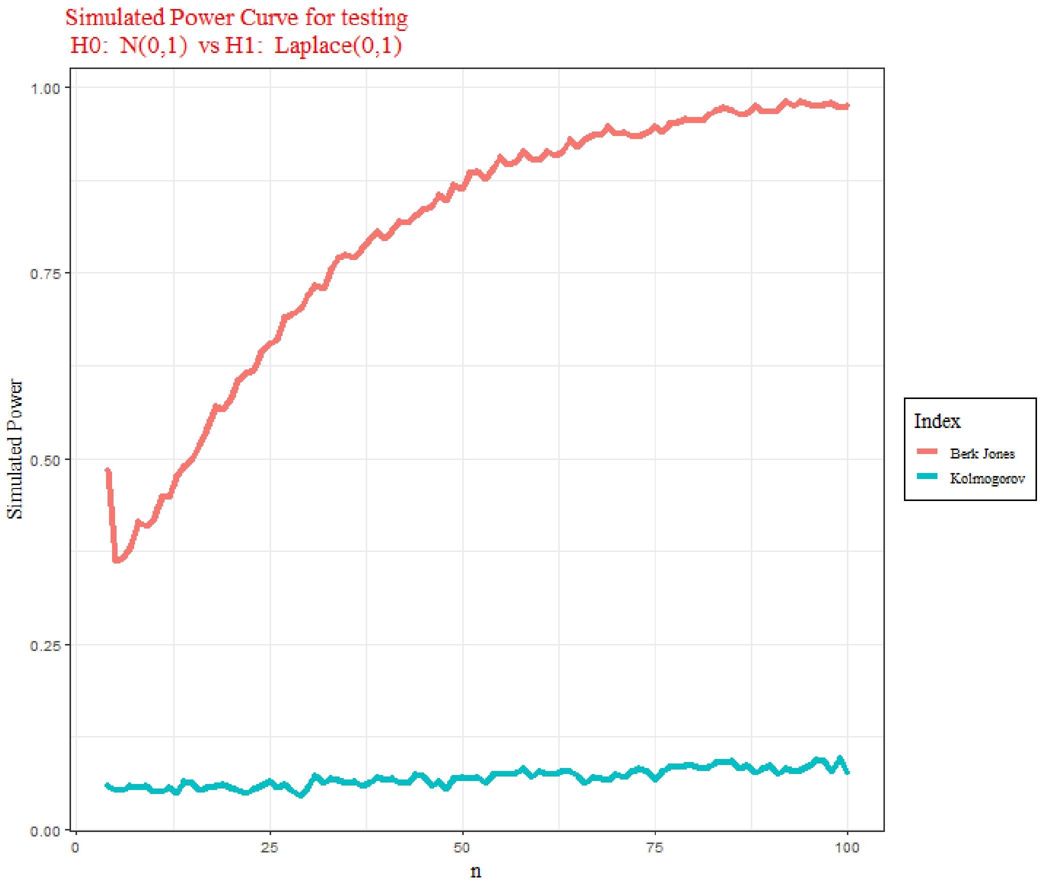






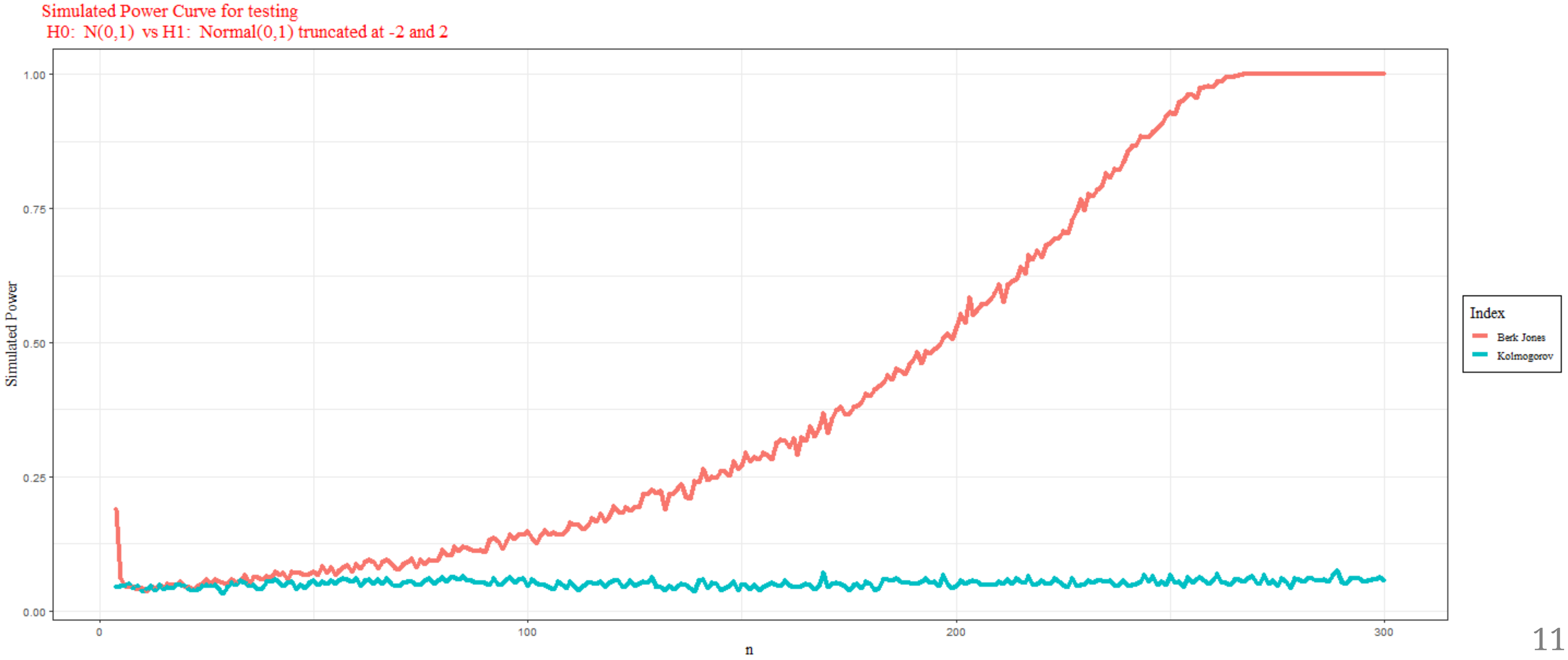
Power Curve Comparison for Berk-Jones :





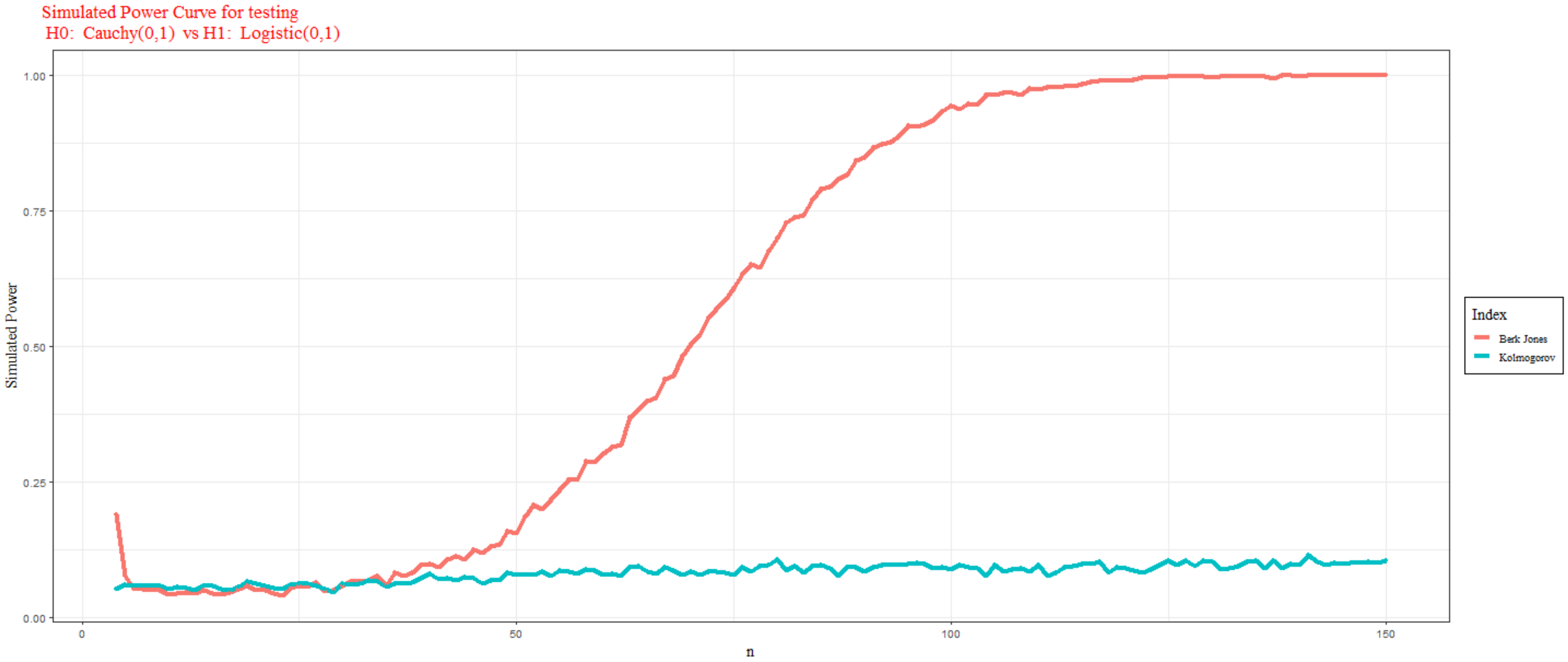
Power Curve Comparison for Berk Jones :

- We saw earlier that Kolmogorov-Smirnov was failing to detect trunctated normal, truncated at -2 and 2.
- This happens mostly because almost the entire probability of Normal is concentrated within -3 and 3.
- Here we try to address this problem using Berk-Jones Test. Can Berk-Jones detect very slight difference in tails ?



Power Curve Comparison for Berk Jones

- Lastly let us also see how Berk-Jones test behave for the test of Cauchy vs Logistic



References :

- Anirban Dasgupta(2008): Asymptotic Theory of Statistics and Probability
- Jean D. Gibbons and Subhabrata Chakraborti(2003) : Nonparametric Statistical Inference
- Peter Gaenssler and Jon A.Weller : A review On Glivenko Cantelli theorems
- Amit Moscovich, Boaz Nadler, Clifford Spiegelman(2016) : On the exact Berk-Jones statistics and their p-value calculation
- Robert H. Berk and Douglas H. Jones(1979) : Goodness-of-Fit Test Statistics that Dominate the Kolmogorov Statistics

Thank You !