Institute of Technology, Kanpur

Department of Mathematics & Statistics

Non-Parametric Inference

Numerical Assignment MTH-516

Submitted by:
Deepak Singh
151091
sdeepak@iitk.ac.in

Submitted to:
Dr. Shubhajit Dutta
Ass. Prof. I.I.T. Kanpur
Department of Mathematics
and Statistics, 208016

April 5, 2017

GitHub Repository:

https://github.com/sdeepak09/MTH_516_Non_parametric_Inference_IITK

Solution 1st:

Results for univariate case with Gaussian Kernel

Dist.	***Difference between estimated density and true density for h=***								
	0.06	**0.12**	**0.25**	**0.5**	**1**	**2**	**4**	**8**	**Min.h**
Nor.	0.0064	0.0029	0.0012	0.0011	0.0044	0.0135	0.0250	0.0350	0.5
Exp.	0.0072	0.0039	0.0012	0.0263	0.0511	0.0841	0.1145	0.1367	0.12
Unif.	0.0128	0.0248	0.0590	0.1597	0.3998	0.6475	0.8114	0.9028	0.06
Cau.	0.0028	0.0014	0.0008	0.0006	0.0013	0.0034	0.0062	0.0086	0.5

Observations from the table:

1. As we can observe that in all the cases Difference between the estimated and true density is firstly decreasing and after that it is increasing w.r.t. increasing values of **h**. Which fits the theory as small h under smooths the data while higher values of h leads to over smoothing of the data.
2. As we can see that in case of Normal and Cauchy distributions Minimising **h** is 0.5 and for exponential distribution it is 0.12 while in case of uniform distribution it is 0.06.
3. As we know from the theory that h should be proportional to (n)^(-1/(d+4)) where d is the dimensionality of the data. From doing the calculations we get that h=0.288. And we can verify that the minimizing h's are close to the calculated h.

<u>Note: we can interpret the below tables as the above one</u>

Results for univariate case with Uniform Kernel:

Dist.	***Difference between estimated density and true density for h=***								
	0.06	**0.12**	**0.25**	**0.5**	**1**	**2**	**4**	**8**	**Min.h**
Nor.	0.0313	0.0160	0.0079	0.0030	0.0049	0.0162	0.0301	0.0338	0.5
Exp.	0.0516	0.0307	0.0272	0.0394	0.0712	0.1124	0.1197	0.1317	0.25
Unif.	0.1854	0.1303	0.1342	0.2027	0.3458	0.5864	0.7715	0.8803	0.12
Cau.	0.0117	0.0064	0.0038	0.0024	0.0029	0.0046	0.0071	0.0091	0.5

Results for d=2 with Multivariate Gaussian Kernel

Dist.	Difference between estimated density and true density for h=								
	0.01	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**Min.h**
Nor.	12.8e-4	6.5e-4	2.5e-4	9.7e-5	6.6e-5	1.4e-4	3.7e-4	7.4e-4	0.25
Exp.	0.0076	0.0061	0.0063	0.0079	0.0108	0.0143	0.0178	0.0213	0.03
Unif.	0.0965	0.1121	0.1528	0.2393	0.3923	0.5809	0.7455	0.8582	0.01
Cau.	2.7e-4	1.3e-4	6.8e-5	3.4e-5	2.1e-5	2.3e-5	3.7e-5	5.9e-5	0.25

Results for d=5 with Multivariate Gaussian Kernel

Dist.	Difference between estimated density and true density for h=								
	0.01	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**Min.h**
Nor.	1.2e-5	2.4e-5	7.7e-6	1.0e-6	1.6e-7	1.0e-7	1.8e-7	2.8e-7	0.5
Exp.	8.0e-5	9.2e-5	4.6e-5	3.2e-5	3.7e-5	4.6e-5	5.4e-5	5.8e-5	0.12
Unif.	0.5978	0.3773	0.4850	0.6651	0.8411	0.9468	0.9468	0.9970	0.03
Cau.	6.6e-8	8.0e-9	4.6e-8	2.5e-8	1.1e-9	2.4e-8	4.8e-8	6.1e-8	0.25

Results for d=10 with Multivariate Gaussian Kernel

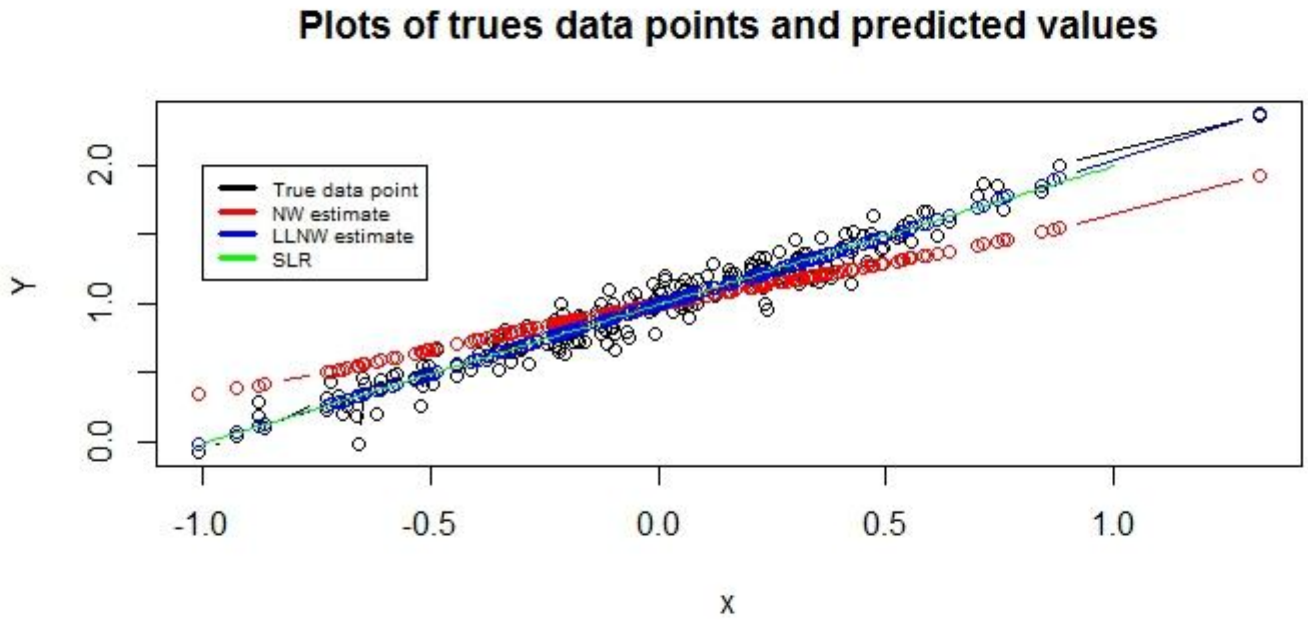
Dist.	Difference between estimated density and true density for h=								
	0.01	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**Min.h**
Nor.	3.6e-15	3.6e-15	3.6e-15	3.6e-15	4.0e-15	4.2e-15	1.4e-15	2.6e-15	1
Exp.	7.2e-9	7.2e-9	7.3e-9	7.3e-9	6.4e-9	6.9e-9	7.1e-9	7.2e-9	0.25
Unif.	6.8485	0.9585	0.8056	0.9267	0.9855	0.9985	0.9999	0.9999	0.06
Cau.	2.1e-24	2.1e-24	2.1e-24	2.1e-24	2.1e-24	2.1e-24	2.0e-24	9.2e-24	1

Results for d=25 with Multivariate Gaussian Kernel

Dist.	Difference between estimated density and true density for h=								
	0.01	**0.03**	**0.06**	**0.12**	**0.25**	**0.5**	**1**	**2**	**Min.h**
Nor.	6.1e-40	6.1e-40	6.1e-40	6.1e-40	6.1e-40	5.7e-40	7.7e-39	7.8e-39	0.5
Exp.	7.1e-25	7.1e-25	7.1e-25	7.1e-25	7.07e-25	7.02e-25	7.1e-25	7.1e-25	0.5
Unif.	0.9999	0.9947	0.9916	0.9994	0.9999	1.0000	1.0000	1.0000	0.06
Cau.	3.0e-58	3.0e-58	3.0e-58	3.0e-58	3.0e-58	3.0e-58	3.0e-58	3.0e-58	1

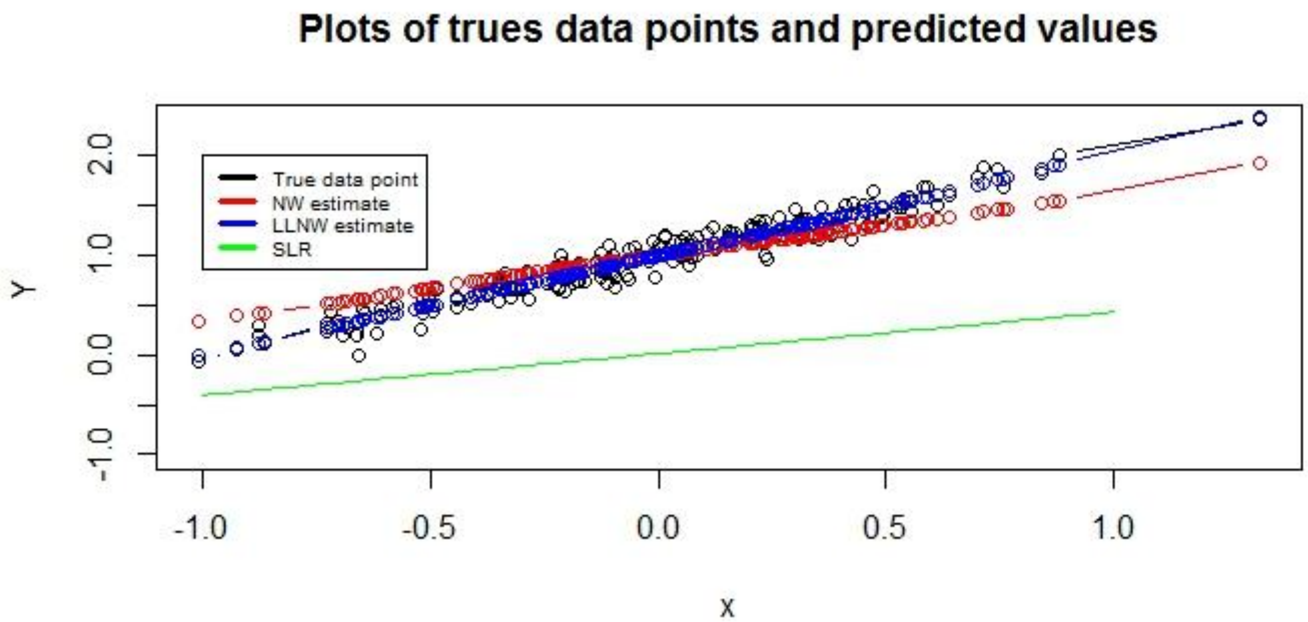
Answer 2nd:

Plot when Y=1+X+epsilon



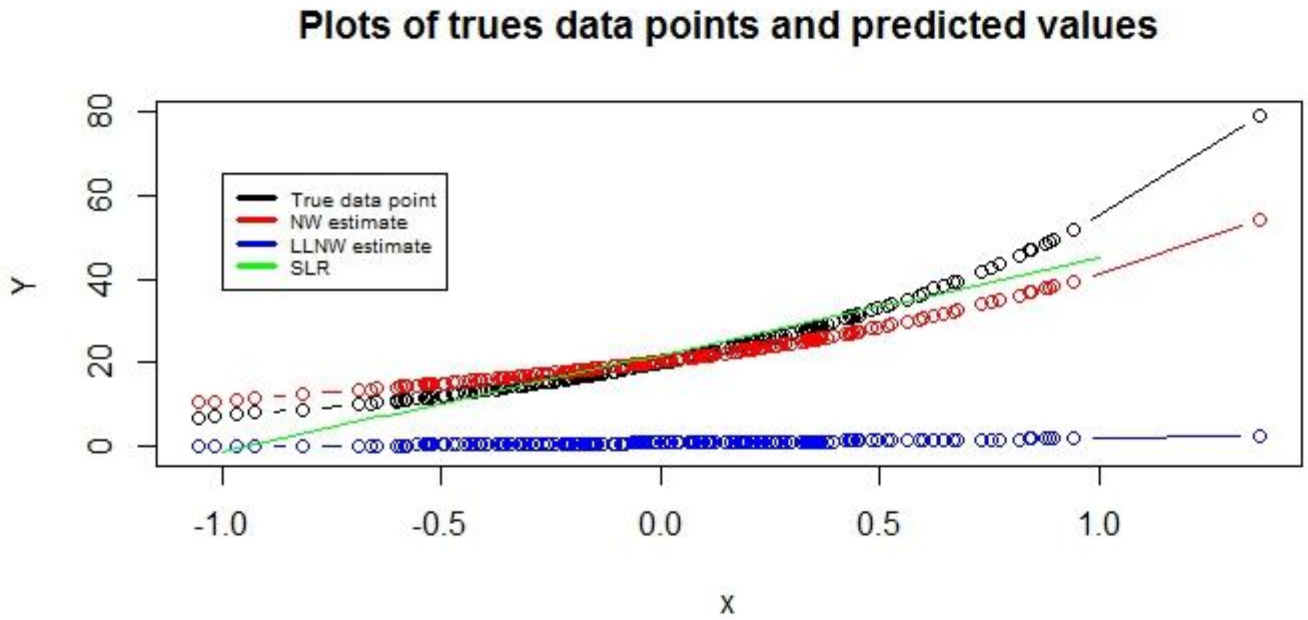
From the above plot we can see that in case of linear relationship SLR and LLNW estimator are fitting well while NW estimate is not fitting so well.

Plot when $Y=0.3*X+X^{23}$+epsilon



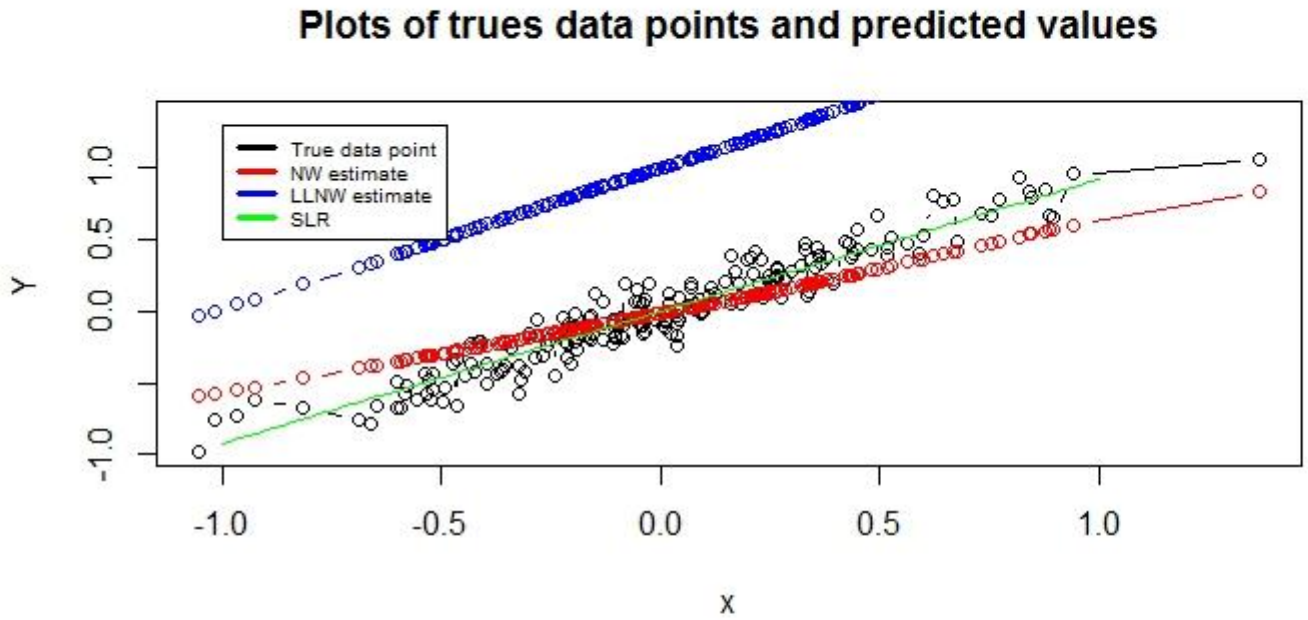
In the case of $\mathbf{Y=0.3*X+X^{23}}$ NW estimate and LLNW estimates are still fitting so well but SLR is not fitting to the data.

Plot when Y=exp(X+3)+epsilon



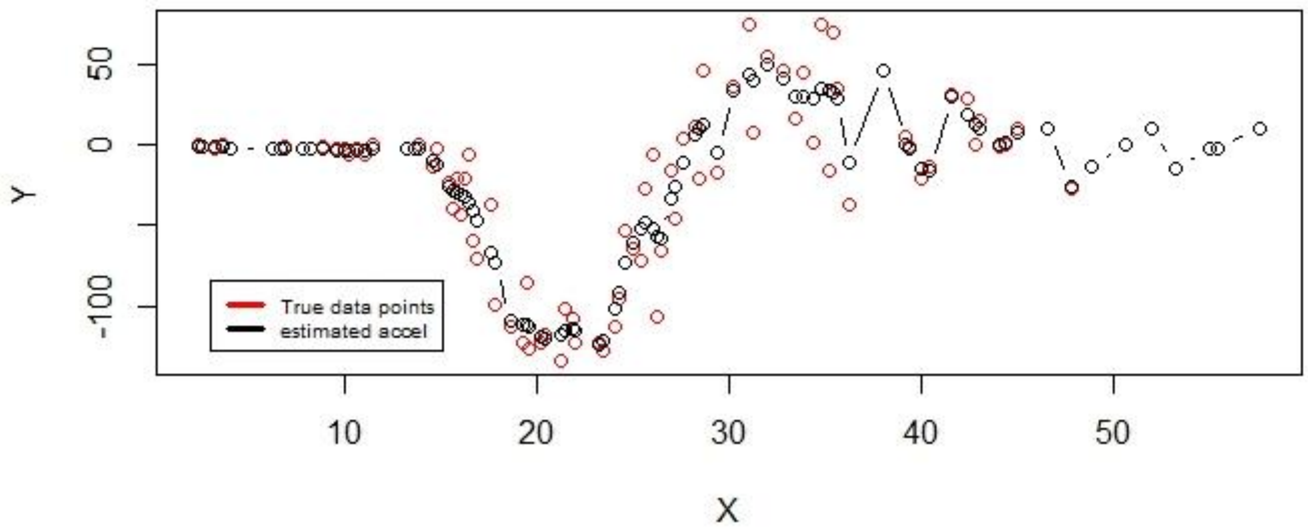
In case of exponential relationship LLNW is fitting very poorly to the data and SLR is fitting better than LLNW but worse than NW.

Plot when Y=sin(X)+epsilon



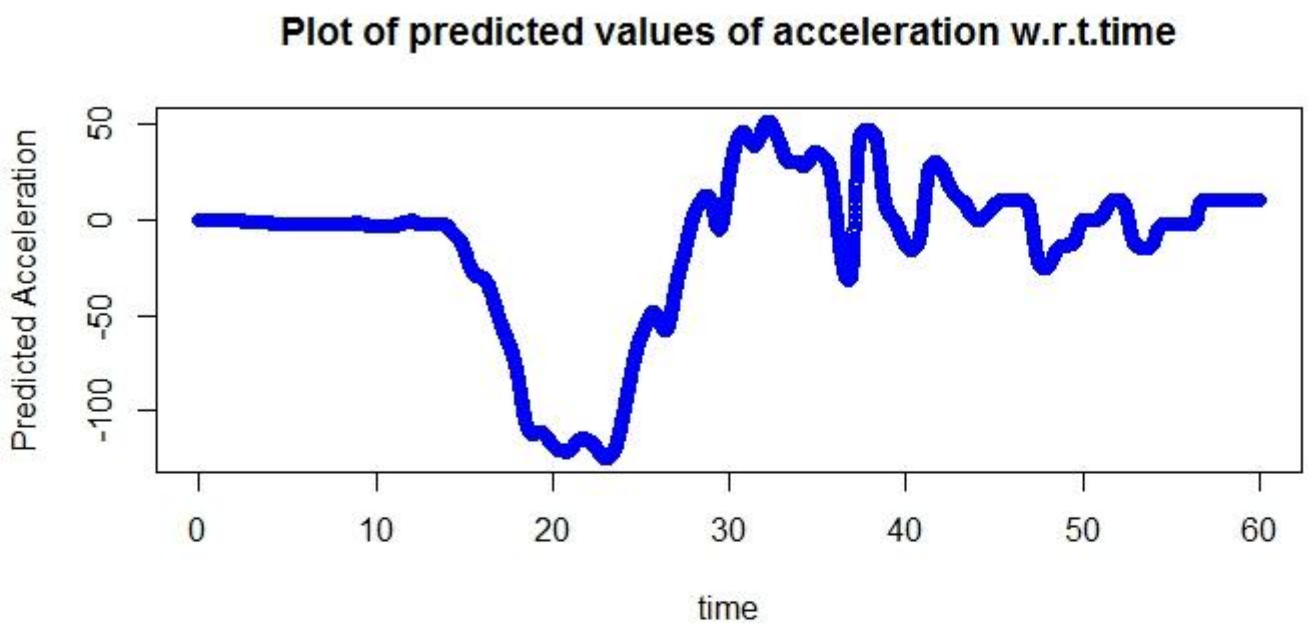
In case of sinusoidal relationship LLNW is again fitting the data very poorly but NW and SLR are fitting well to the data.

Plot of the True and predicted accelaration at given time points in the data:



From the above we can see that in NW regression is fitting so well in case of Non-linear regression. We can also observe that our predicted values are following the same pattern as the data has.

Plot of predicted values of acceleration for 10000 values of time between 0 and 60:



We can verify the same pattern as the data has.

Answer 3rd:

Table below contains the mean sum of squares of the 4 given estimators (Mean, Median, Mode and Hodge-Lehman estimator) about 0 for Normal, Uniform, Laplace and Cauchy case.

Distribution	***Mean Sum of Squares about 0***				
	Mean	**Median**	**Mode**	**H-L Estimator**	**Most Efficient estimator**
Normal	0.00999	0.01490	0.25490	0.01068	Mean
Uniform	0.00350	0.01029	0.40579	0.00381	Mean
Laplace	0.01860	0.01047	0.08316	0.01272	Median
Logistic	0.03585	0.04384	0.53792	0.03278	H-L Est.
Cauchy	591.98280	0.02520	195.00030	0.03760	Median

As we can see that in case of normal distribution Mean has the minimum mean sum of squares so for normal distribution sample mean will be most efficient estimator among the given estimators. Similarly Mean, Median, H-L estimator and Median will be most efficient estimators for Uniform, Laplace, Logistic and Cauchy distribution respectively.

As we can see that the Mean sum of squares of Median and H-L estimator is not changing much with change in the distribution so we can say that 'median' and 'H-L estimator' are robust among the given estimators.

Answer 4th:

I have written a code in R to execute the exact run test. Code with the example can be seen in the R code file. I have executed the exact run test for sweet potato data available in randtests package and got the result that we cannot accept the Null hypothesis of randomness of the sequence.

I have executed the various run test available in randtests package for sweet potato data.

Wald-Wolfowitz Runs Test:

```
        Runs Test

data:  dt
statistic = -4.5751, runs = 17, n1 = 35, n2 = 35, n = 70, p-value
= 4.759e-06
alternative hypothesis: nonrandomness
```

As we can see that p-value is too less so we cannot accept the Null Hypothesis.

Bartels Rank Test:

```
        Bartels Ratio Test

data:  dt
statistic = -4.4925, n = 70, p-value = 7.038e-06
alternative hypothesis: nonrandomness

>
```

As we can see that p-value is too less so we cannot accept the Null Hypothesis.

Mann-Kendall Rank Test:

```
        Mann-Kendall Rank Test

data:  dt
statistic = 3.9594, n = 70, p-value = 7.514e-05
alternative hypothesis: trend
```

As we can see that p-value is too less so we cannot accept the Null Hypothesis.

Turning Point Test:

```
        Turning Point Test

data:  dt
statistic = 0.67514, n = 69, p-value = 0.4996
alternative hypothesis: non randomness
```

As p-value is greater than the 0.025 hence we cannot reject the Null Hypothesis.

Cox-Stuart Trend Test:

```
        Cox Stuart test

data:  dt
statistic = 30, n = 35, p-value = 2.236e-05
alternative hypothesis: non randomness
```

As we can see that p-value is too less so we cannot accept the Null Hypothesis.

Answer 5th:

Table below contains the average power and standard deviation of the powers of Kolmogrov-Smirnov, Cramer von Mises and Chi-squared tests under the H_0: N(0,1) and H_1: N(0.05,1), N(0,2), N(0.05,2), DE(0,1), C(0,1) and Exp(0,1)

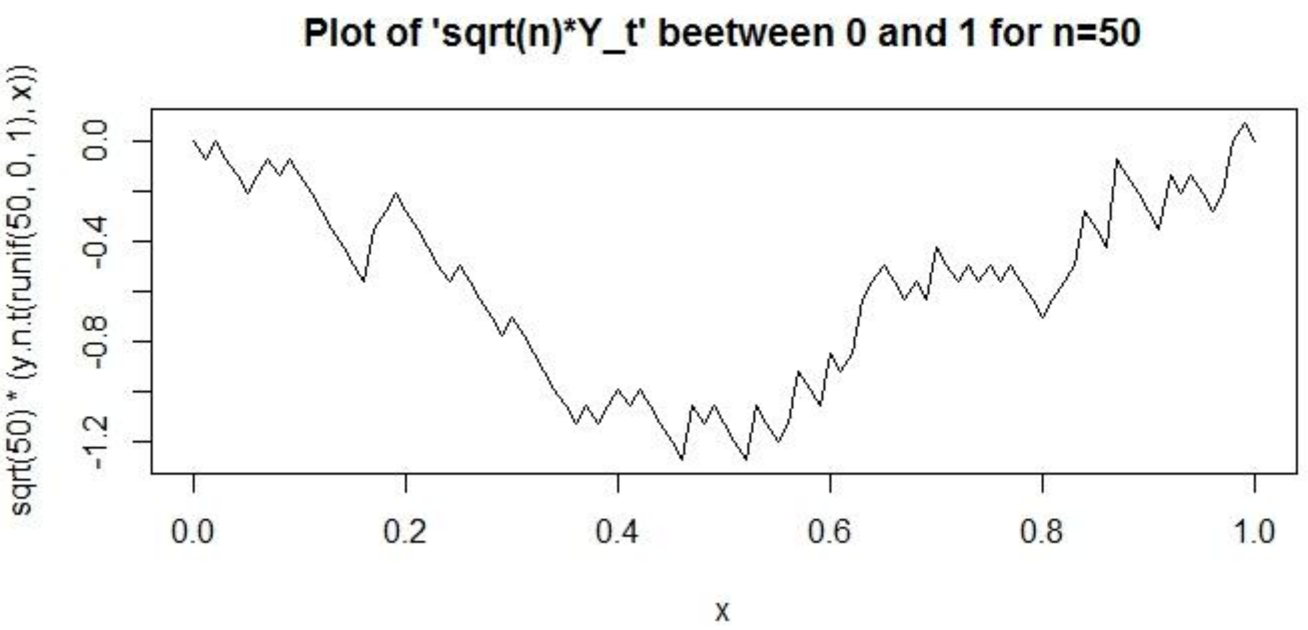
DISTRIBUTION UNDER ALT. HYPOTHESIS	KOLMOGROV-SMIRNOV		CRAMER VON MISSES		CHI-SQUARED	
	Average	Stand. Dev.	Average	Stand. Dev.	Average	Stand. Dev.
N(0.05,1)	0.0417	0.00874	0.0767	0.01301	0.0658	0.01074
N(0,2)	0.4294	0.02109	0.5768	0.02312	0.7026	0.02161
N(0.05,2)	0.4303	0.02178	0.5854	0.02328	0.7093	0.02153
DE(0,1)	0.0604	0.01123	0.0898	0.01375	0.3124	0.02196
C(0,1)	0.8116	0.01776	0.8451	0.01564	0.09258	0.01364
EXP(0,1)	1	0	1	0	1	0

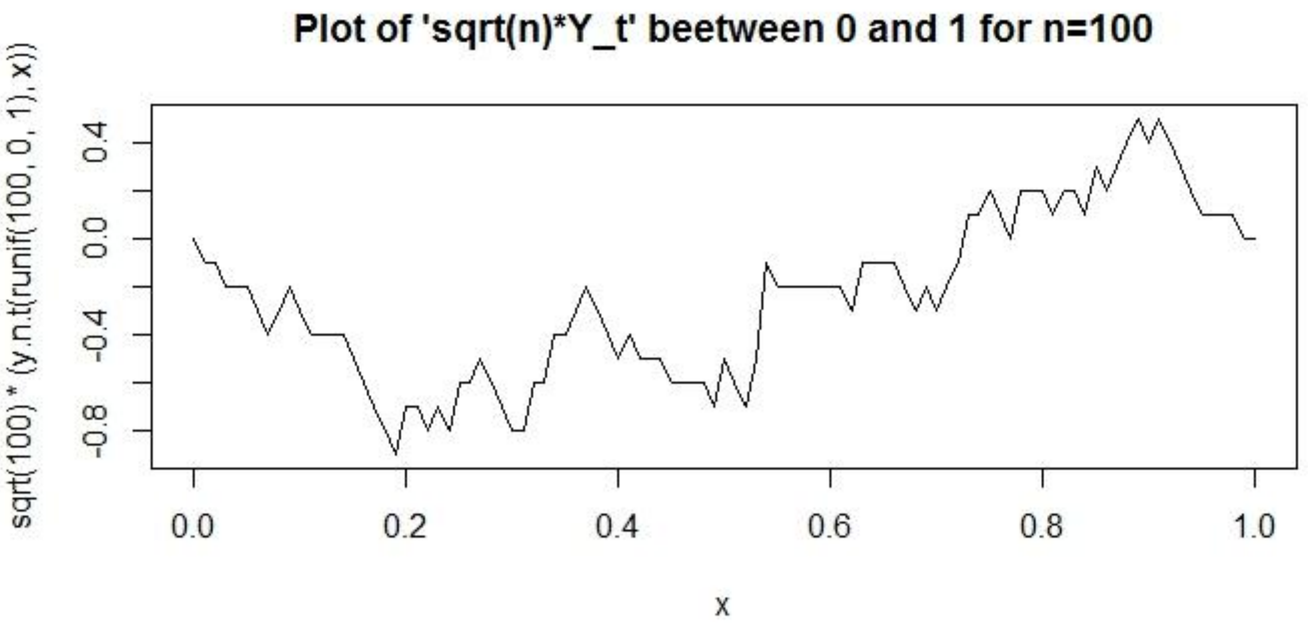
Observations from the table:

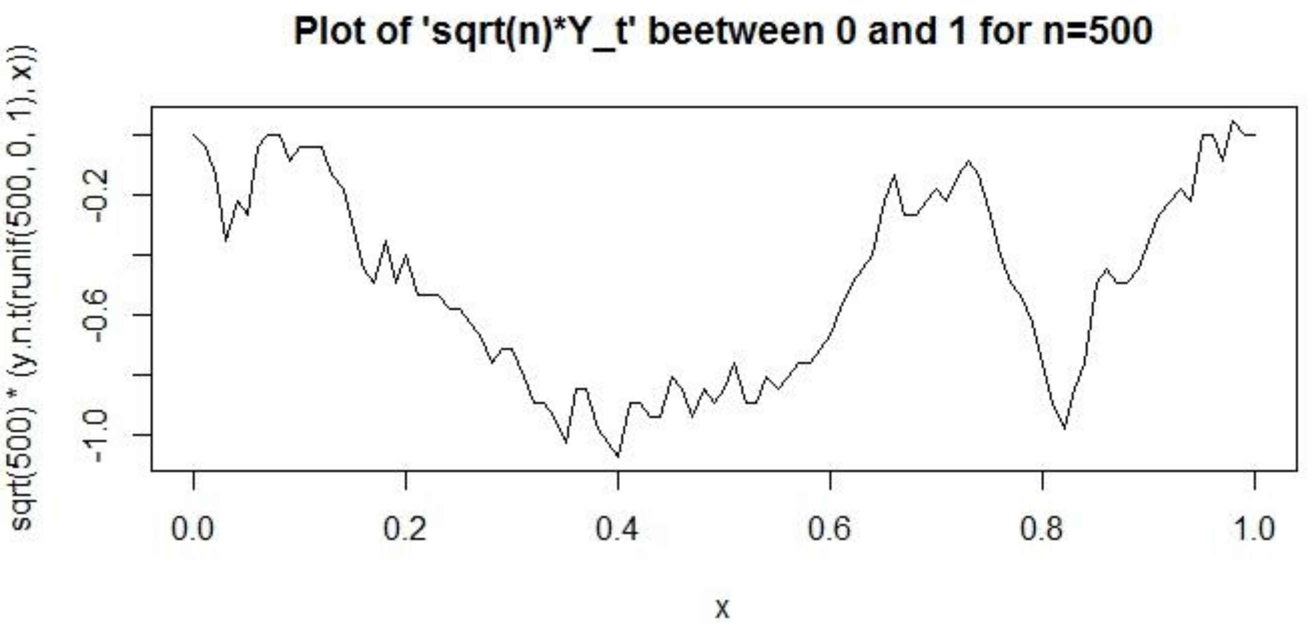
1. In case of H_1: N(0.05,1) all the three tests are not doing well as the powers are aro und 0.04 to 0.07 which is not a good value of power I think.
2. In case of H_1: N(0,2), N(0.05,2) and DE(0,1) chi-squared test is doing better that th e both as its average value of the power is higher than the rest 2 with almost same standard deviation.
3. In case of H_1: C(0,1) K-S and CvM are doing very well but the performance of Chi-S quared test is poor in this case.
4. In case of H_1: Exp(0,1) all the tests are best as the power is 1 with 0 standard devi ation.

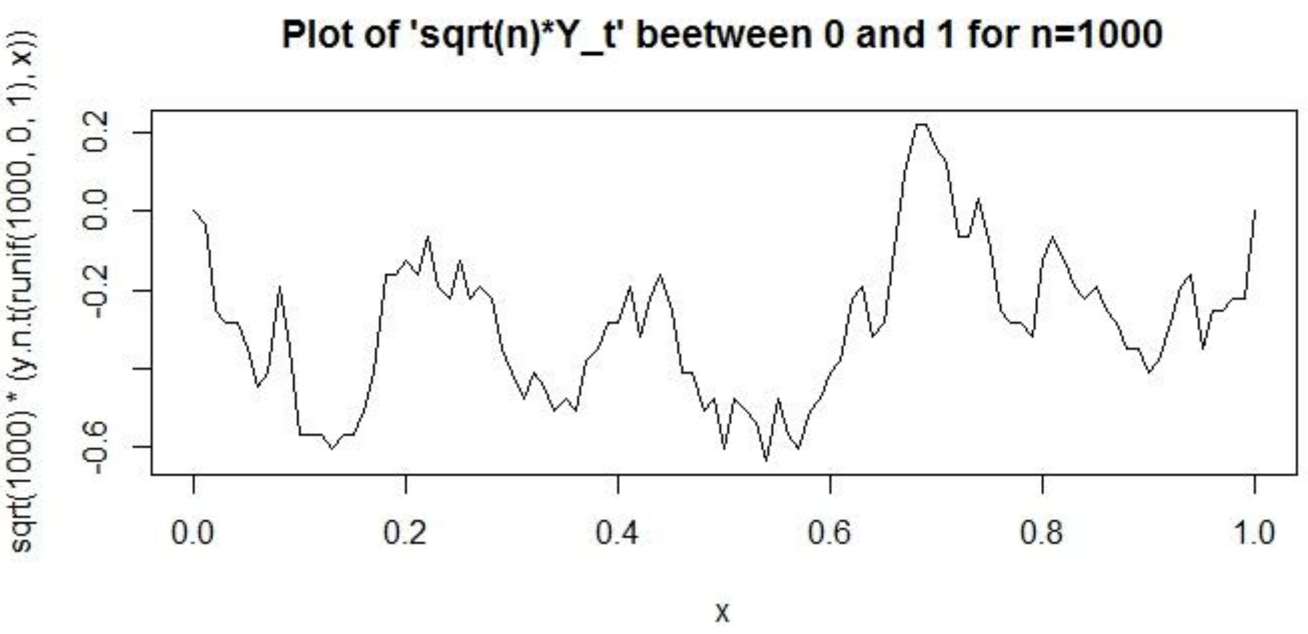
Answer 6^{th}:

Below are the plots of sqrt(n)*Y_n(t) by changing the number of random variables






Plot of 'sqrt(n)*Y_t' beetween 0 and 1 for n=500
sqrt(500) * (y.n.t(runif(500, 0, 1), x))
-1.0
-0.6
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
x


Plot of 'sqrt(n)*Y_t' beetween 0 and 1 for n=1000
sqrt(1000) * (y.n.t(runif(1000, 0, 1), x))
-0.6
-0.2
0.0
0.2
0.0
0.2
0.4
0.6
0.8
1.0
x