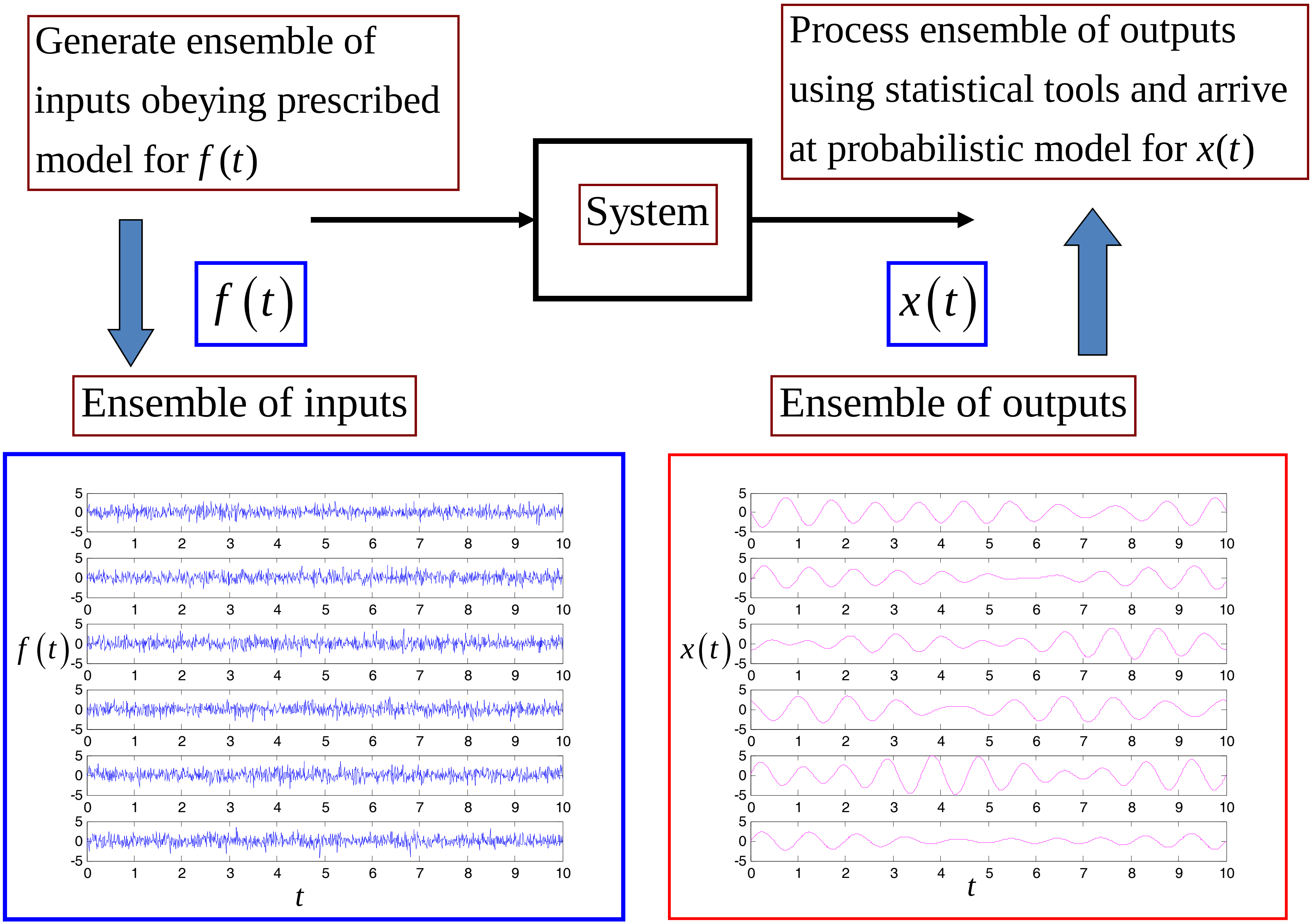
Stochastic Structural Dynamics

Lecture-26

Monte Carlo simulation approach-2

Dr C S Manohar
Department of Civil Engineering
Professor of Structural Engineering
Indian Institute of Science
Bangalore 560 012 India
manohar@civil.iisc.ernet.in


Generate ensemble of inputs obeying prescribed model for f (t)
f (t)
System
x(t)
Process ensemble of outputs using statistical tools and arrive at probabilistic model for x(t)
Ensemble of inputs
Ensemble of outputs
f (t)
t
x(t)
t

Estimation of paramters

- Method of moments

$$\Theta_{(k)} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad [\text{Mean, Variance, Skewness, Kurtosis,...}]$$

- Method of maximum likelihood

Estimate directly the parameters of the pdf

quantities like mode, median, range, ...

can be estimated using this method.

Estimation of mean

Let X be a random variable with PDF $P_X(x)$, pdf $p_X(x)$, mean μ, and standard deviation σ.

Let $\{X_i\}_{i=1}^{n}$ be an iid sequence with common pdf $p_X(x)$.

That is, $X_i \perp X_j \forall i \neq j \in [1, n]$,

$\langle X_i \rangle = \mu, \; Var\,[X_i] = \sigma^2, p_{X_i}(x) = p_X(x) \forall i \in [1, n]$.

$\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator $[\forall n]$ of μ with minimum variance and the lowest variance is $\frac{\sigma^2}{n}$.

Sampling distribution for the estimator of mean

Consider the estimator $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$

Θ is an unbiased estimator of μ with variance $\frac{\sigma^2}{n}$.

Let us consider the case in which σ^2 is known.

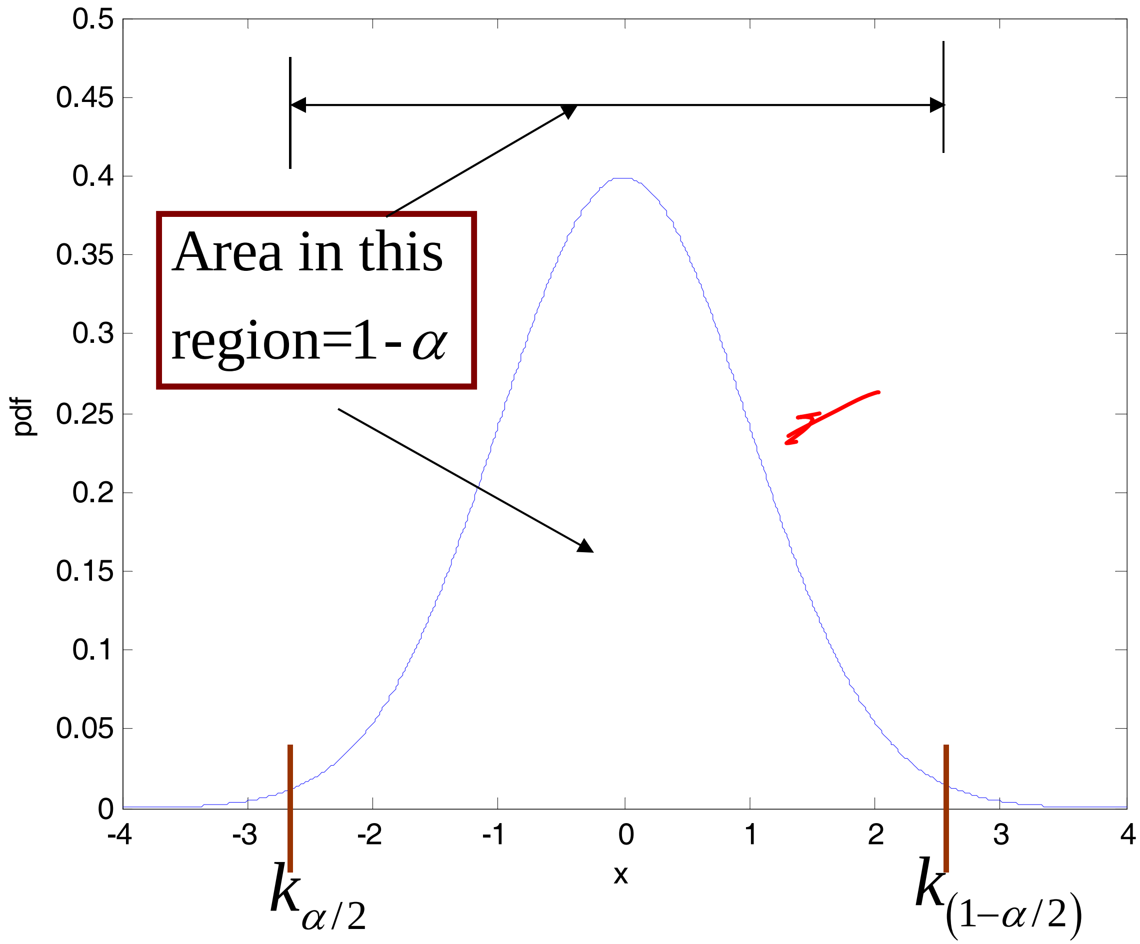
If X is Gaussian, it would mean that $\{X_i\}_{i=1}^{n}$ is an iid sequence of Gaussian random variables and consequently Θ would also be Gaussian distributed.

If X is not Gaussian, by virtue of central limit theorem, for large n, we may still consider Θ to be Gaussian.

It may be inferred that $\Theta \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ or, $\frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$.

Confidence interval estimation

Consider $\dfrac{\Theta-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Consider $(1-\alpha)$ to be a specified probability $\left[\text{e.g., } (1-\alpha)=0.95\right]$.



$$P\left[k_{\alpha/2} < \frac{\Theta-\mu}{\sigma/\sqrt{n}} \le k_{1-\alpha/2}\right] = 1-\alpha$$

$$k_{\alpha/2} = -\Phi^{-1}(1-\alpha/2)$$
$$k_{(1-\alpha/2)} = \Phi^{-1}(1-\alpha/2)$$

$$P\left[k_{\alpha/2} < \frac{\Theta-\mu}{\sigma/\sqrt{n}} \le k_{1-\alpha/2}\right] = 1-\alpha$$

$$\Rightarrow$$

$$P\left[\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu \le \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

$\theta =$ observed value of Θ from the sample.

$\left[\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right] = (1-\alpha)$ confidence interval for the population mean μ. Suppose, 1-$\alpha = 0.95$.

We say with 95% confidence that the true mean is contained in the interval $\left[\theta + k_{0.025}\frac{\sigma}{\sqrt{n}}, \theta + k_{0.075}\frac{\sigma}{\sqrt{n}}\right]$

$$P\left[\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu \le \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Remark

• This should be interpreted as the probability that the random interval $\left(\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right)$ contains the population mean μ is $1-\alpha$.

• Remember μ is a deterministic quantity.

• $\theta = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a point estimate &

$\left(\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right)$ is a confidence interval estimate.

Example

$x =$

-0.4326
-1.6656
0.1253
0.2877
-1.1465
1.1909
1.1892
-0.0376
0.3273
0.1746

$$\theta = \frac{1}{10}\sum_{i=1}^{10} x_i = 0.0013$$

$$\alpha = 0.05$$

$$k_{\alpha/2} = -1.96 \;\&\; k_{1-\alpha/2} = 1.96$$

95% confidence interval

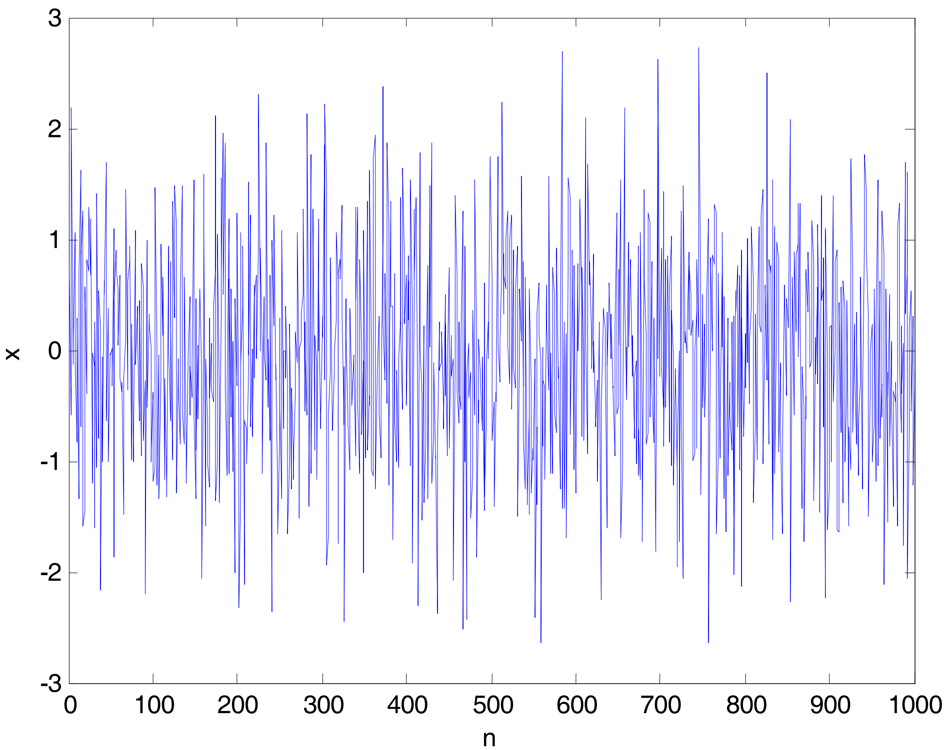
$$= \left(0.013 - \frac{1.96}{\sqrt{10}}, 0.013 + \frac{1.96}{\sqrt{10}}\right)$$

$$= (-0.6068, 0.6328)$$

The point estimate of mean is 0.0013.

With 95% confidence we say that the population mean is contained in the interval $= (-0.6068, 0.6328)$.

n=1000



$$\theta = \frac{1}{1000}\sum_{i=1}^{1000} x_i = \text{-}0.0446$$

$$\alpha = 0.05$$

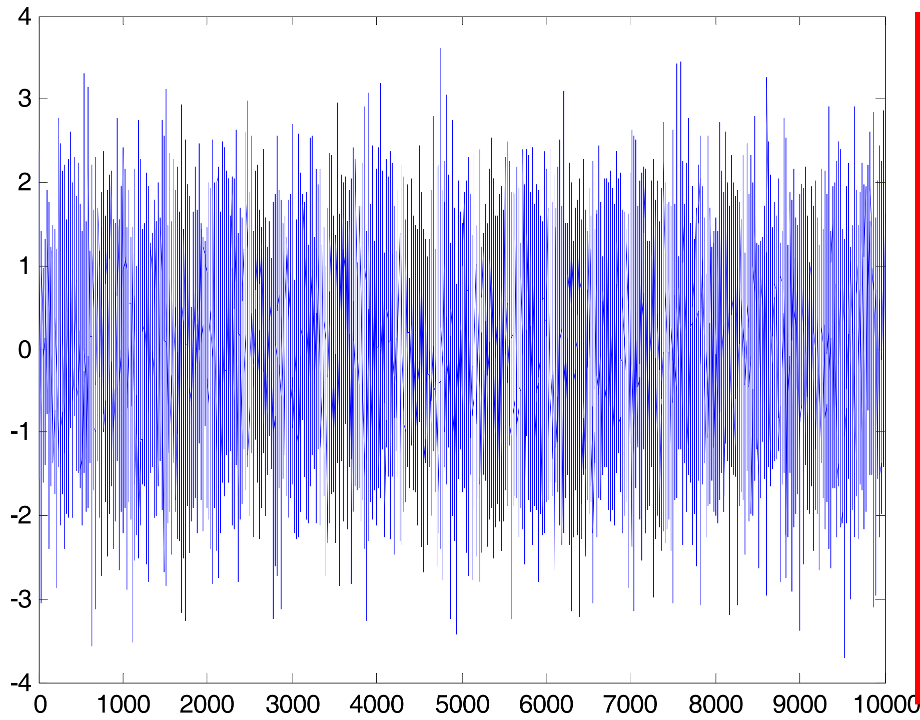
$$k_{\alpha/2} = -1.96 \,\&\, k_{1-\alpha/2} = 1.96$$

95% confidence interval

$$= \left(-0.0446 - \frac{1.96}{\sqrt{1000}}, -0.0446 + \frac{1.96}{\sqrt{10}} \right)$$

$$= \left(-0.1065, 0.0174 \right)$$

n=10000



$$\theta = \frac{1}{10000}\sum_{i=1}^{10000} x_i = 0.0033$$

$$\alpha = 0.05$$

$$k_{\alpha/2} = -1.96 \;\&\; k_{1-\alpha/2} = 1.96$$

95% confidence interval

$$= \left(-0.0446 - \frac{1.96}{\sqrt{10000}}, -0.0446 + \frac{1.96}{\sqrt{10000}} \right)$$

$$= \left(-0.0163, 0.0229 \right)$$

Sampling distribution for variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \text{ with } \bar{\mathrm{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$\left\langle S^2 \right\rangle = \frac{1}{n-1}\sum_{i=1}^{n}\left\langle \left(X_i - \bar{X}\right)^2 \right\rangle$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left\langle \left[\left(X_i - \mu\right) - \left(\bar{X} - \mu\right)\right]^2 \right\rangle$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left\langle \left(X_i - \mu\right)^2 + \left(\bar{X} - \mu\right)^2 - 2\left(X_i - \mu\right)\left(\bar{X} - \mu\right) \right\rangle$$

Note: $\left\langle \left(X_i - \mu\right)^2 \right\rangle = \sigma^2 \ \& \left\langle \left(\bar{X} - \mu\right)^2 \right\rangle = \frac{\sigma^2}{n}$

$$\Rightarrow \left\langle S^2 \right\rangle = \sigma^2$$

Exercise

Similarly we can show that $\mathrm{Var}\left[S^2\right] = \frac{\sigma^4}{n}\left(\frac{\mu^4}{\sigma^4} - \frac{n-3}{n-1}\right)$

$\Rightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is an unbiased and consistent estimator of σ^2.

$$\Rightarrow \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}\right)^2$$

If population is Gaussian, X_i and $\bar{X}$ are Gaussian.

RHS: sum of squares of Gaussian random variables

such sums have $\chi^2 -$ distributions.

$\Rightarrow \frac{(n-1)S^2}{\sigma^2}$ is $\chi^2 -$ distributed with $(n\text{-}1)$ dof.

The pdf of such a random variable is given by

$$p(u) = \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} u^{\left(\frac{n-1}{2}-1\right)} \exp\left(-\frac{u}{2}\right); 0 < u < \infty$$

Student's t - distribution

$$T = \frac{X}{\sqrt{Y/n}}$$

$$X \sim N(0,1)$$

$Y \sim \chi^2$ – distribution with n-dofs

$X \perp Y$

$$\Rightarrow p_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}; -\infty < t < \infty$$

Sampling distribution for the estimator of mean with variance not known

Consider the estimator $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$

Θ is an unbiased estimator of μ with variance $\frac{\sigma^2}{n}$.

$\frac{\Theta - \mu}{s/\sqrt{n}}$ ~ Student t-distribution with n-1 dofs.

s = estimate of standard deviation from the sample.

$$p_\Theta(\theta) = \frac{\Gamma\left[(f+1)/2\right]}{\sqrt{\pi f}\,\Gamma(f/2)}\left(1+\frac{\theta^2}{f}\right)^{-\frac{1}{2}(f+1)} ;-\infty < \theta < \infty$$

f = dof

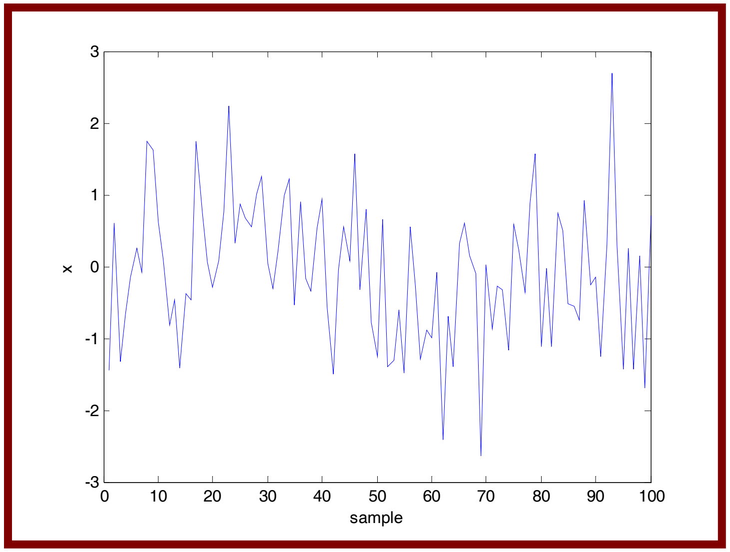
$$p_\Theta(\theta) = \frac{\Gamma\left[(f+1)/2\right]}{\sqrt{\pi f}\,\Gamma(f/2)}\left(1+\frac{\theta^2}{f}\right)^{-\frac{1}{2}(f+1)} ;-\infty < \theta < \infty$$

Consider the statement

$$P\left(\theta_{\frac{\alpha}{2},n-1} < \frac{\Theta-\mu}{s/\sqrt{n}} \le \theta_{\left(\frac{1-\alpha}{2}\right),n-1}\right) = 1-\alpha$$

From this one can obtain the confidence interval for the population mean.

Example



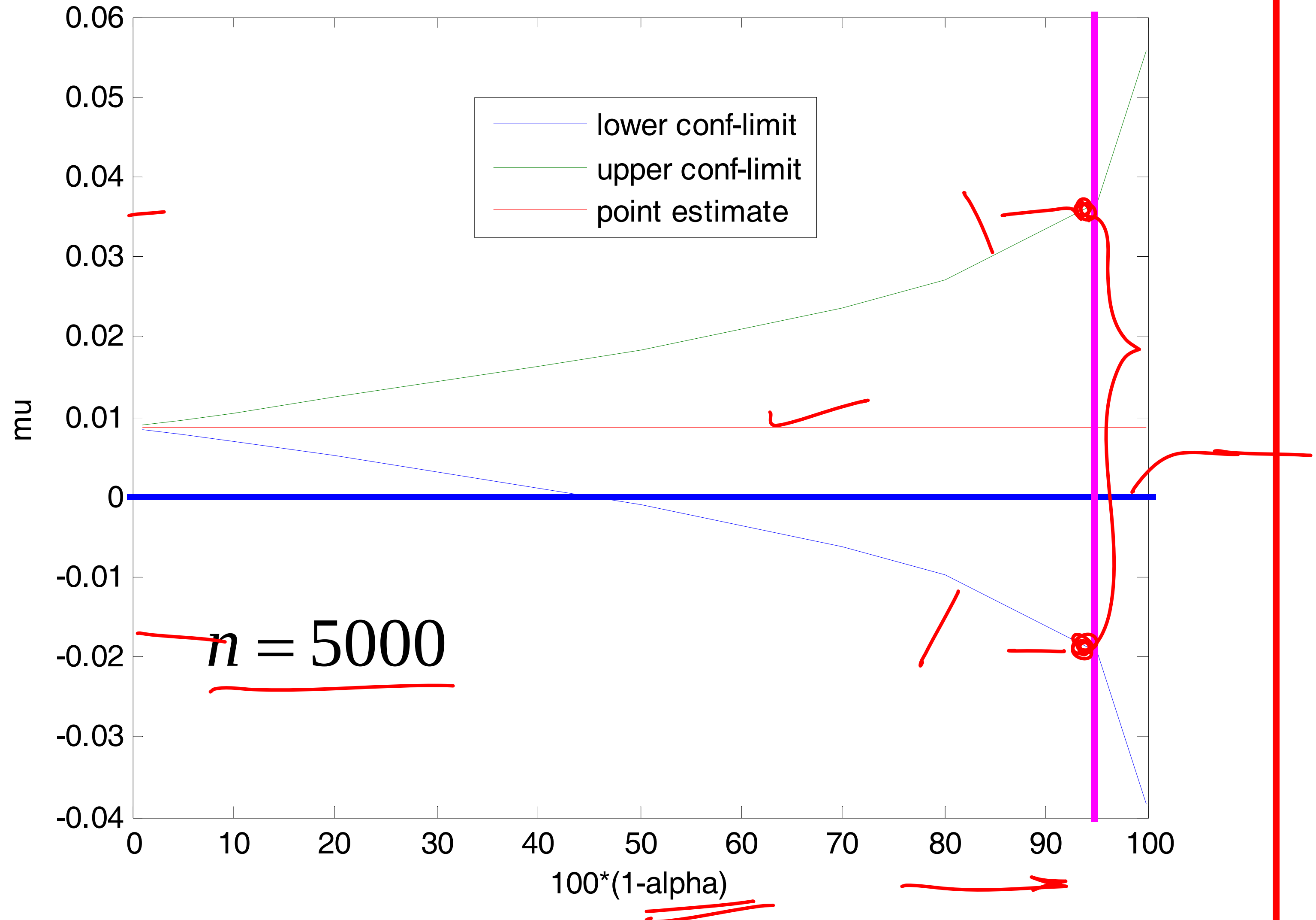
Point estimates

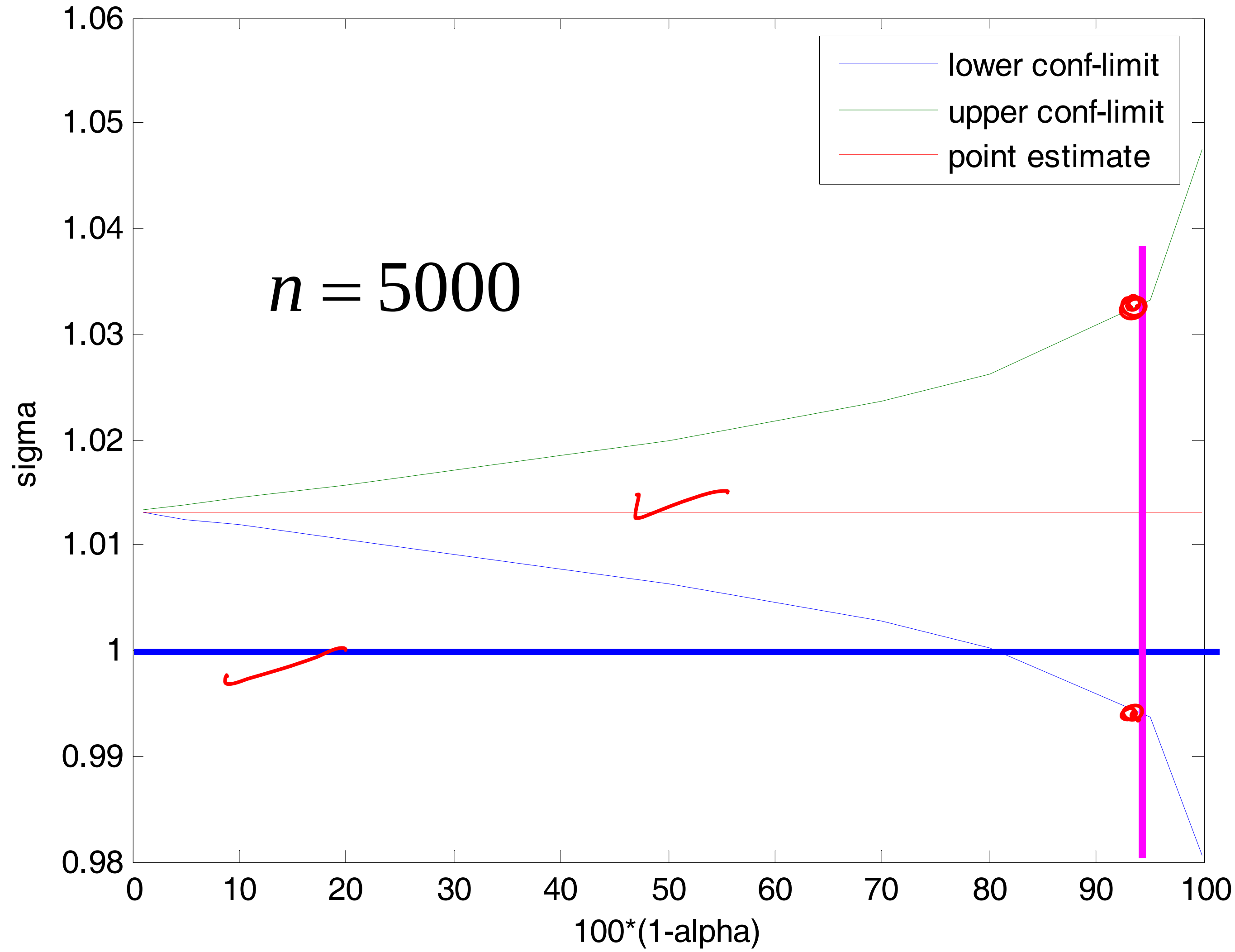
$\hat{\mu}$ =-0.0677; $\hat{\sigma}$=0.9777

$100\times(1-\alpha)$	$(\underset{\sim}{\mu},\tilde{\mu})$		$(\underset{\sim}{\sigma},\tilde{\sigma})$	
99.0000	-0.0689	-0.0665	0.9802	0.9819
95.0000	-0.0738	-0.0615	0.9767	0.9855
90.0000	-0.0800	-0.0554	0.9723	0.9899
80.0000	-0.0925	-0.0429	0.9636	0.9990
50.0000	-0.1339	-0.0015	0.9357	1.0302
5.0000	-0.2617	0.1263	0.8585	1.1358
0.1000	-0.3993	0.2639	0.7893	1.2652

Data used in the example n=100

-1.4440	0.6123	-1.3235	-0.6616	-0.1461	0.2481	-0.0766	1.7382	1.6220
0.6264	0.0918	-0.8076	-0.4613	-1.4060	-0.3745	-0.4709	1.7513	0.7532
0.0650	-0.2928	0.0828	0.7662	2.2368	0.3269	0.8633	0.6794	0.5548
1.0016	1.2594	0.0442	-0.3141	0.2267	0.9967	1.2159	-0.5427	0.9122
-0.1721	-0.3360	0.5415	0.9321	-0.5703	-1.4986	-0.0503	0.5530	0.0835
1.5775	-0.3308	0.7952	-0.7848	-1.2631	0.6667	-1.3926	-1.3006	-0.6050
-1.4886	0.5585	-0.2774	-1.2937	-0.8884	-0.9865	-0.0716	-2.4146	-0.6943
-1.3914	0.3296	0.5985	0.1472	-0.1014	-2.6350	0.0281	-0.8763	-0.2655
-0.3276	-1.1582	0.5801	0.2398	-0.3509	0.8921	1.5783	-1.1082	-0.0259
-1.1106	0.7508	0.5002	-0.5173	-0.5592	-0.7534	0.9258	-0.2485	-0.1498
-1.2584	0.3126	2.6903	0.2897	-1.4228	0.2468	-1.4358	0.1486	-1.6931
				0.7192				


lower conf-limit
upper conf-limit
point estimate
mu
100*(1-alpha)
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
0
10
20
30
40
50
60
70
80
90
100
$n = 5000$



lower conf-limit
upper conf-limit
point estimate
$n = 5000$
sigma
100*(1-alpha)
1.06
1.05
1.04
1.03
1.02
1.01
1
0.99
0.98
0
10
20
30
40
50
60
70
80
90
100

Factors influencing confidence interval

- The statistic used as the estimate
- The observations made
- Confidence level
- Sampling distribution
- Sample size

Number of samples needed for a given width of confidence interval

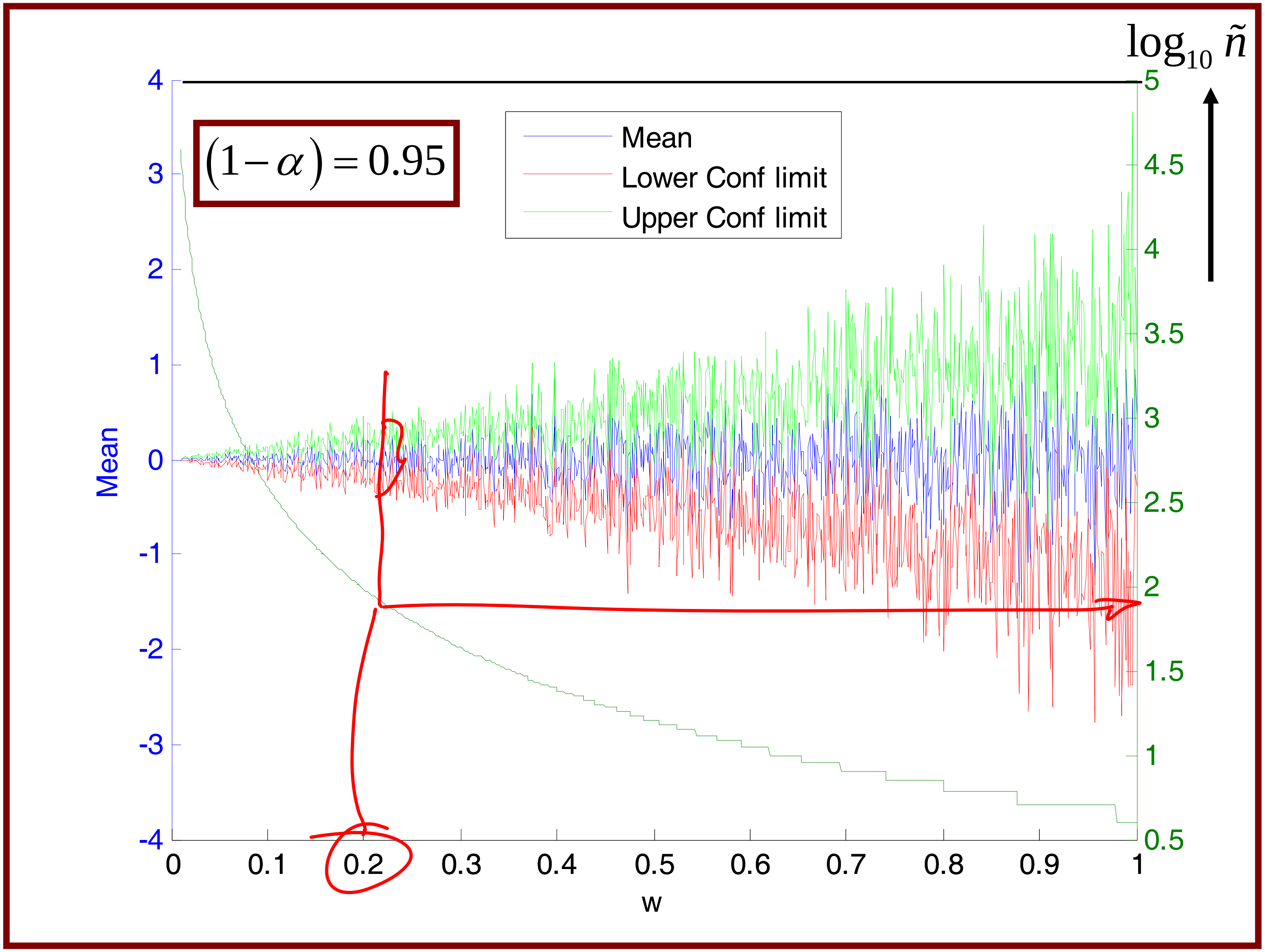
Consider the estimator for population mean with known variance.

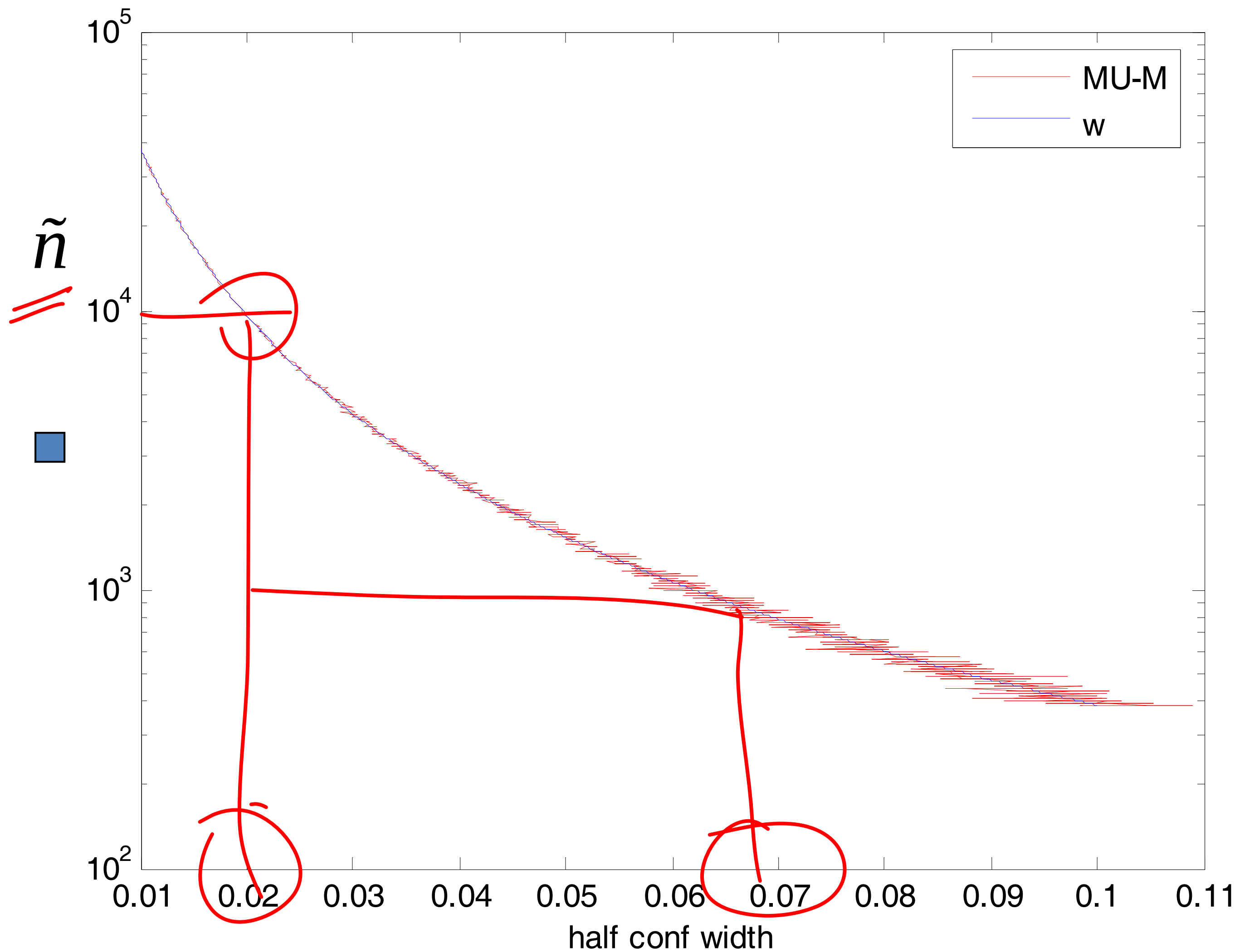
$$\bullet \Theta = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ or, } \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1).$$

$$\bullet P\left[\theta + k_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu \le \theta + k_{(1-\alpha/2)}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Let $w = k_{(1-\alpha/2)}\dfrac{\sigma}{\sqrt{n}}$ = half width of confidence interval be specified. Minimum number of samples required

$$\tilde{n} = \frac{1}{w^2}\left[\sigma k_{(1-\alpha/2)}\right]$$


$\log_{10} \tilde{n}$
$(1-\alpha) = 0.95$
Mean
Lower Conf limit
Upper Conf limit
Mean
w
4
3
2
1
0
-1
-2
-3
-4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5



ñ
10^5
10^4
10^3
10^2
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.11
half conf width
MU-M
w

Hypothesis testing

- A method for making decisions on properties of a population based on observed samples.

Two competing hypotheses:

- Null hypothesis $[H_0]$
- Alternative hypothesis $[H_A]$

Sample (n=10)

3.7023

0.0032

5.3760

5.8630

1.5606

8.5727

8.5675

4.8871

5.9819

5.5239

Sample mean

=5.0038 //

Null Hypothesis:

$H_0 : \mu = 5.0$

That is, the sample is drawn from a population whose mean is 5.0.

"Null" hypothesis: hypothesis of no difference.

Alternative hypothesis:

$H_A : \mu \neq 5.0$

Is the observed difference between estimate and the population mean is due to sampling fluctuations (that is, random causes) or due to systematic (non-random) causes?

OR

Is the oberved variation $|\theta - \mu|$ arising due to some assignable causes or due to non-assignable causes?

OR

Is the observed variation $|\theta - \mu|$ significant?

Significant : variation due to assignable causes.

If the difference is due to random causes, no action is needed. Otherwise, action is necessary.

Decisions: accept or reject the hypothesis

Errors

- Reject the hypothesis when it should have been accepted. [Type I error: error of commision]
- Accept the hypothesis when it should have been rejected. [Type II error: error of omission]

- Type I error: action when no action was needed.
- Type II error: inaction when action was needed.

Which error is more dangerous?

Type II [∵ inaction when action was needed]

Note :

Type I error: error of commision

Based on the action taken, one would come to know that a wrong decision was taken.

Can the errors be eliminated?

No. ∵ As long as decisions are based on samples, sampling errors cannot be avoided.

Null Hypothesis: $H_0 : \mu = 5.0$

Alternative hypothesis: $H_A : \mu \neq 5.0$

Decision ↓	H_0 is true	H_0 is not true
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

$$P[\text{Committing Type I error}] = \alpha$$

$$P[\text{Committing Type II error}] = \beta$$

$$P[\text{Accepting } H_0 \text{ when } H_0 \text{ is true}] = 1 - \alpha$$

$$P[\text{Accepting } H_A \text{ when } H_0 \text{ is true}] = 1 - \beta$$

$\text{P}\left[\text{Committing Type I error}\right]=\alpha$ is fixed.

It is difficult to fix $\text{P}\left[\text{Committing Type II error}\right]=\beta$ since it is difficult to assess the consequence of action not taken.

Steps in hypothesis testing

Step 1

$H_0 : \mu = \mu_0$

$H_A : \mu \neq \mu_0$

Step 2 : Choose α [level of significance;0.01,0.05,0.1]

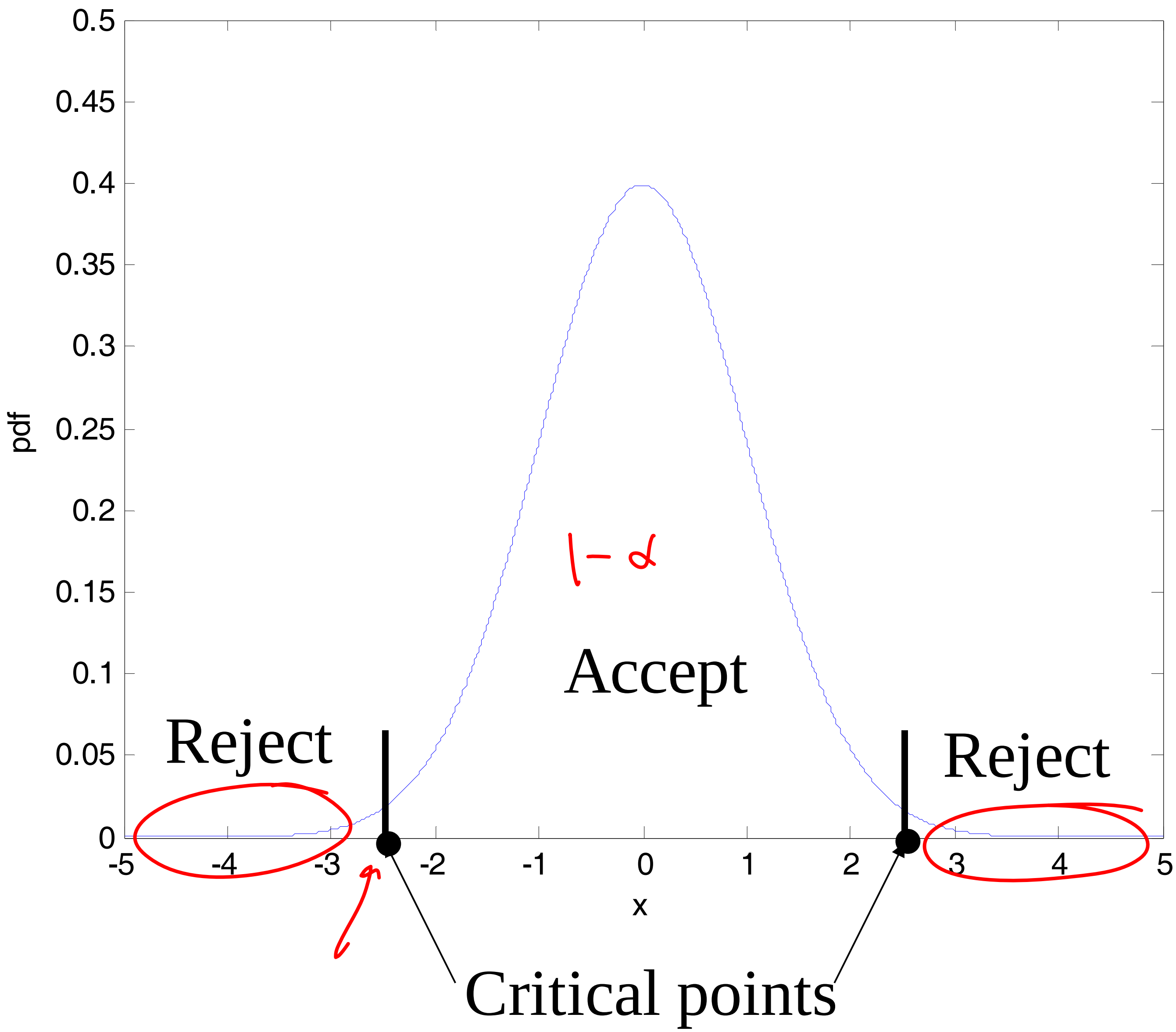
$1-\alpha =$ Level of confidence.

Step 3 : Identify the test statistic and its distribution.

Here $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$. $Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4 : Based on the sample obtain an estimate of the test statistic.

Step 5 : Define the region of rejection of the null hypothesis


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
pdf
-5
-4
-3
-2
-1
0
1
2
3
4
5
x
1 − α
Accept
Reject
Reject
Critical points

x =

-0.1867
0.7258
-0.5883
2.1832
-0.1364
0.1139
1.0668
0.0593
-0.0956
-0.8323
0.2944
-1.3362
0.7143
1.6236
-0.6918

$n = 15; \sigma = 1$

Step 1

$$H_0 : \mu = 0.0$$

$$H_A : \mu \neq 0.0$$

Step 2: $\alpha = 0.05$

Step 3: $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i; \ Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4: $\theta = 0.1943; \ z = 0.7524$

Step 5:

$$\Phi^{-1}(0.025) = -1.96$$

$$\Phi^{-1}(1 - 0.025) = 1.96$$

$$-1.96 < z = 0.7524 < 1.96$$

Accept the null hypothesis at 5% significance level

x =

0.8617
0.1555
0.7128
0.2034
0.7749
0.7823
0.7970
0.8862
0.1036
5.1879
1.2127
3.0126
0.3665
0.4306
0.0172

$n = 15; \sigma = 1$

Step 1

$H_0 : \mu = 0.0$

$H_A : \mu \neq 0.0$

Step 2: $\alpha = 0.05$

Step 3: $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$; $Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4: $\theta = 1.0336$; $z = 4.0033$

Step 5:

$\Phi^{-1}(0.025) = -1.96$

$\Phi^{-1}(1 - 0.025) = 1.96$

z does not lie in the acceptance region

$-1.96 < z \leq 1.96$

Reject the null hypothesis at 5% significance level

x =

0.8617
0.1555
0.7128
0.2034
0.7749
0.7823
0.7970
0.8862
0.1036
5.1879
1.2127
3.0126
0.3665
0.4306
0.0172

$n = 15; \sigma = 1$

Step 1

$H_0 : \mu = 1.0$

$H_A : \mu \neq 0.0$

Step 2 : $\alpha = 0.05$

Step 3 : $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i;\ Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4 : $\theta = 1.0336;\ z = 0.1301$

Step 5 :

$\Phi^{-1}(0.025) = -1.96$

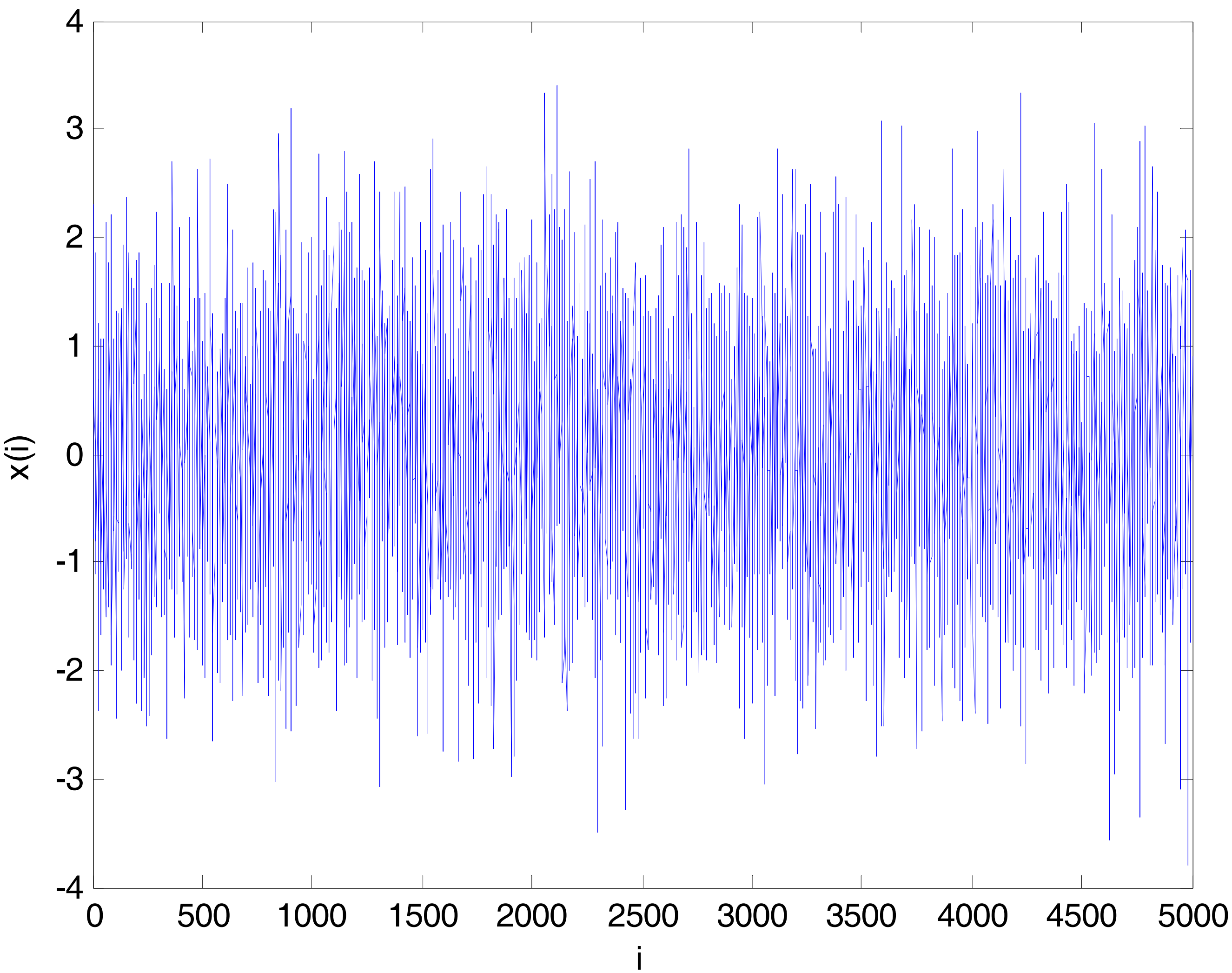
$\Phi^{-1}(1 - 0.025) = 1.96$

z does lie in the acceptance region

$-1.96 < z \leq 1.96$

Accept the null hypothesis at 5% significance level

$$n = 5000; \sigma = 1$$



Step 1

$H_0 : \mu = 0.0$

$H_A : \mu \neq 0.0$

Step 2 : α **=** 0.05

Step 3 : $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$; $Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4 : θ **=** 0.0123; $z = 0.8732$

Step 5 :

$\Phi^{-1}(0.025) = -1.96$

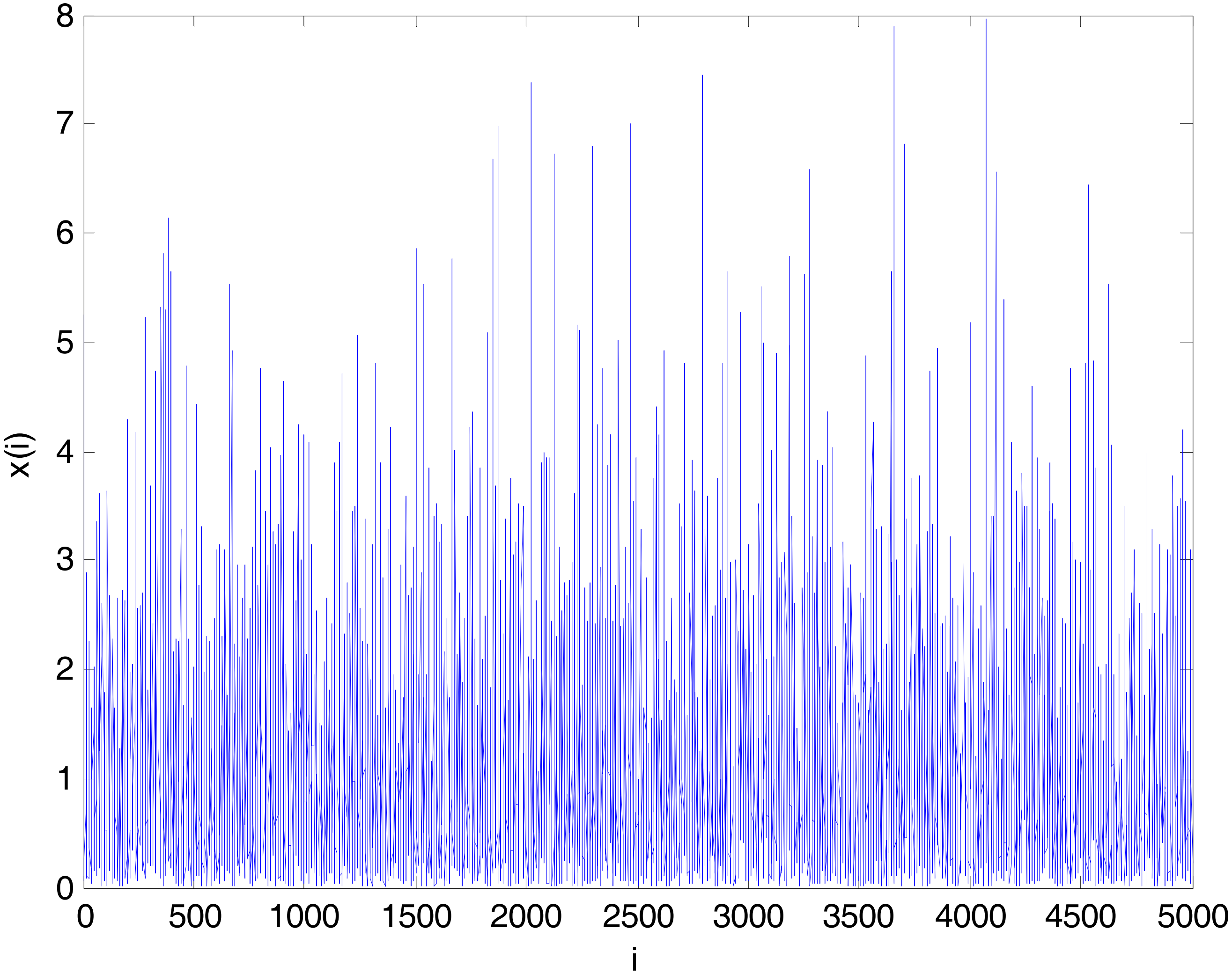
$\Phi^{-1}(1 - 0.025) = 1.96$

z does lie in the acceptance region

$-1.96 < z \leq 1.96$

Accept the null hypothesis at 5% significance level

$$n = 5000; \sigma = 1$$



Step 1

$\mathrm{H_0} : \mu = 1.0$

$\mathrm{H_A} : \mu \neq 0.0$

Step 2 : α **=** 0.05

Step 3 : $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i$; $Z = \frac{\Theta - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

Step 4 : θ **=** 0.990; z=-0.7040

Step 5 :

$\Phi^{-1}(0.025) = -1.96$

$\Phi^{-1}(1 - 0.025) = 1.96$

z does lie in the acceptance region

$-1.96 < z \leq 1.96$

Accept the null hypothesis at 5% significance level

Population standard deviation not known

Step 1

$H_0 : \mu = \mu_0$

$H_A : \mu \neq \mu_0$

Step 2 : Choose α [level of significance;0.01,0.05,0.1]

$1 - \alpha =$ Level of confidence.

Step 3 : Identify the test statistic and its distribution.

Here $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \Theta\right)^2.$

$T = \frac{\Theta - \mu}{s/\sqrt{n}} \sim$ Student's t-distribution with n dof.

Step 4 : Based on the sample obtain an estimate of the test statistic.

Step 5 : Define the region of rejection of the null hypothesis

Step 1

$H_0 : \sigma^2 = 100$

$H_A : \sigma^2 > 100$

Step 2 : Choose α [level of significance;0.01,0.05,0.1]

$1-\alpha =$ Level of confidence.

Step 3 : Identify the test statistic and its distribution.

Here $\Theta = \frac{1}{n}\sum_{i=1}^{n} X_i ; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \Theta\right)^2.$

$C = \frac{(n-1)^2 S^2 \Theta - \mu}{\sigma^2} \sim \chi^2$ with $(n\text{-}1)$ dof.

Step 4 : Based on the sample obtain an estimate of the test statistic.

Step 5 : Define the region of rejection of the null hypothesis

Probability papers

Let X be a random variable with PDF $P_X(x)$.

Let $\{x_i\}_{i=1}^{n}$ be a sample of X.

Probability paper is a special plotting device in which y-axis is scaled in such a way that the PDF function appears as a straight line.

Example

$$P_X(x) = 1 - \exp(-\lambda x)$$ $x \geq 0$

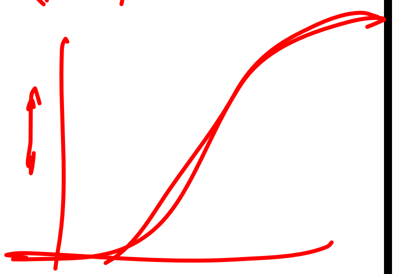
$$1 - P_X(x) = G_X(x) = \exp(-\lambda x)$$

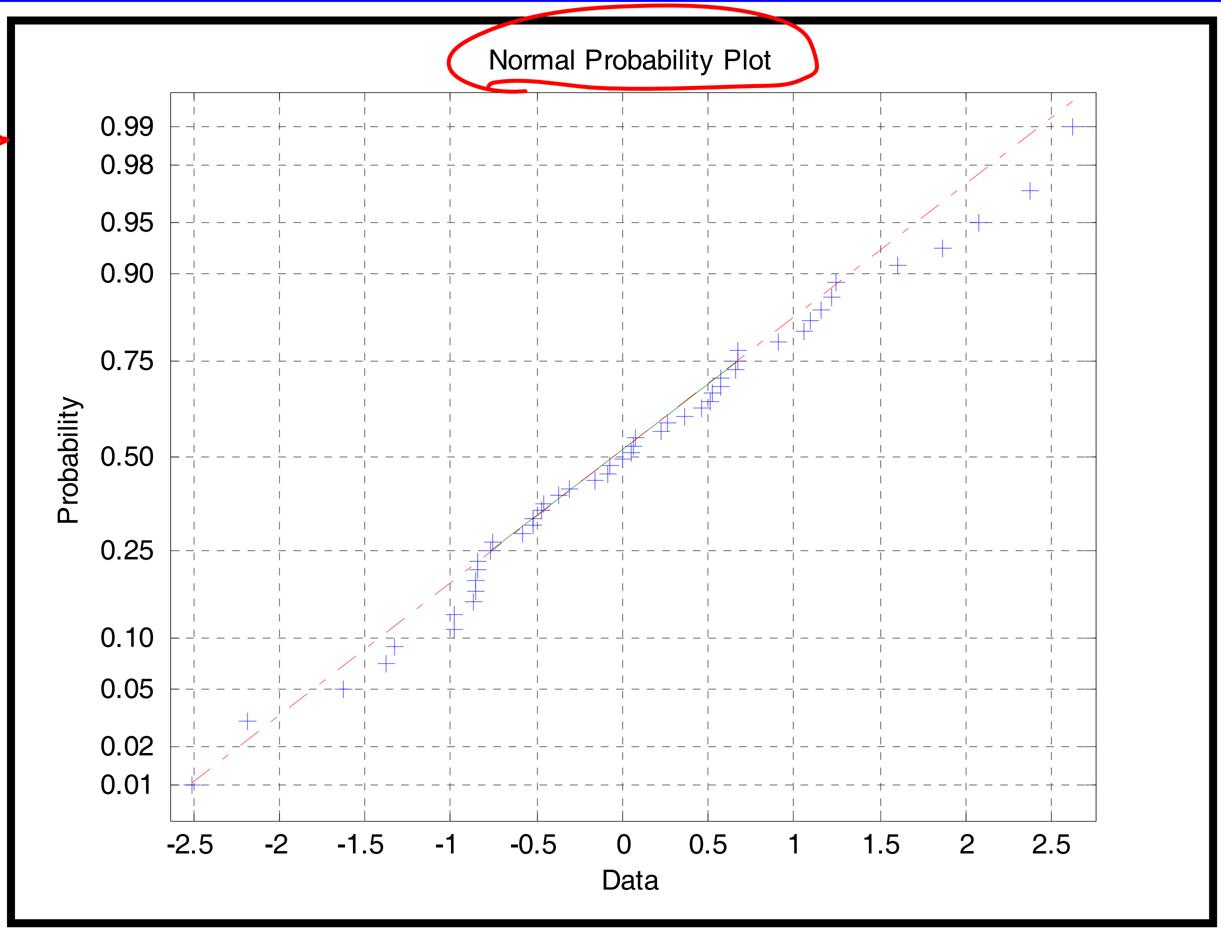
$$\log G_X(x) = -\lambda x$$

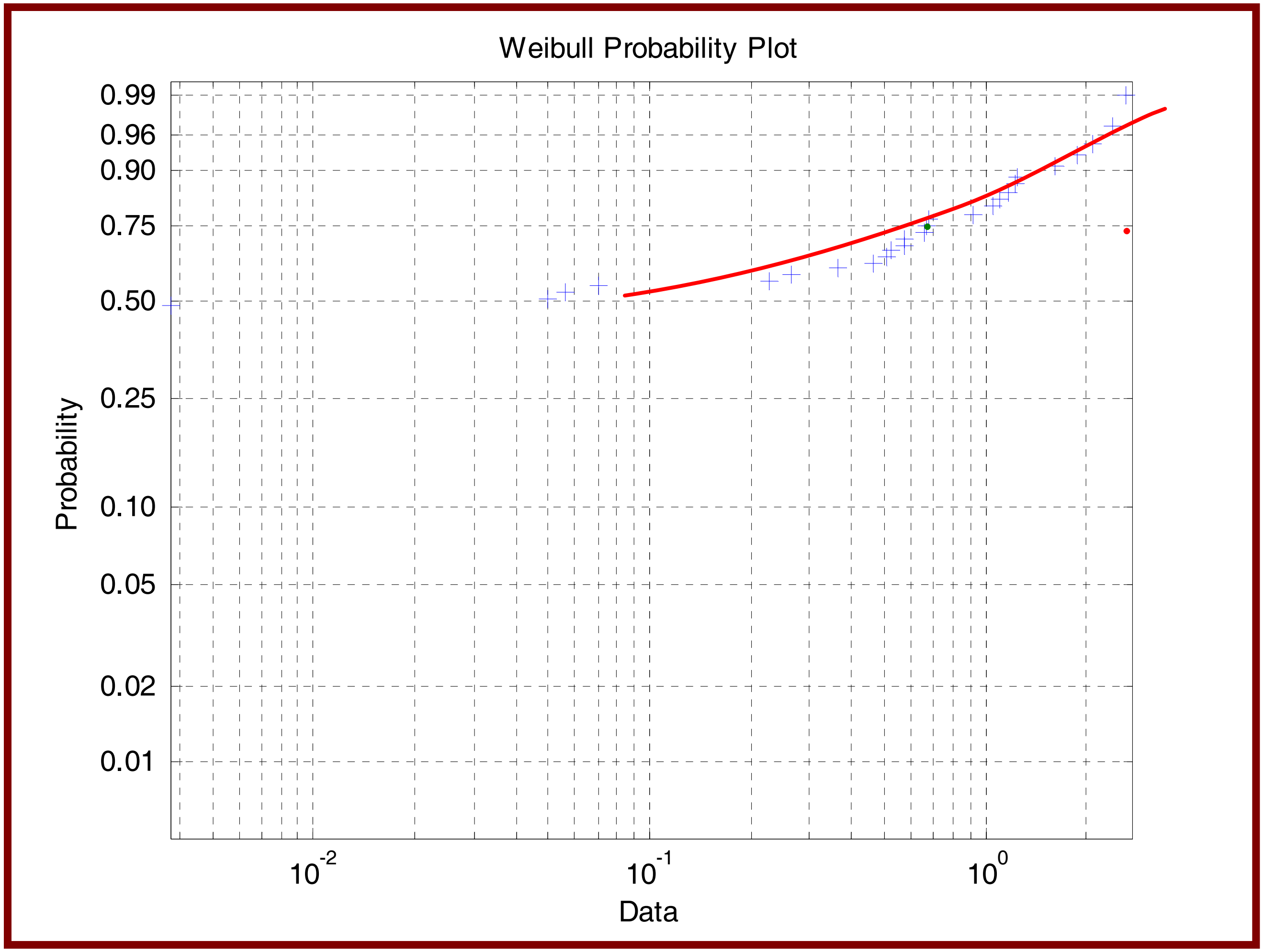
The complement of the cumulative PDF appears as a straight line.

2.6229	**2.3756**	**0.0564**	**-0.5879**	**-0.9830**	**-1.3790**	**-0.0870**	**-0.7607**	**-0.8745**
-0.4685	**-0.3767**	**-1.6317**	**-0.8433**	**2.0772**	**1.8603**	**-0.0755**	**-2.5102**	**-2.1873**
0.5252	**-0.8524**	**-0.3110**	**1.6036**	**0.6590**	**0.6732**	**-0.7697**	**1.0570**	**0.2261**
0.5711	**-0.5269**	**0.5755**	**0.9128**	**-0.1585**	**0.6686**	**-0.8599**	**1.1602**	**0.0702**
-0.5166	**-0.9860**	**0.2651**	**0.0494**	**0.4624**	**-0.4575**	**0.3642**	**1.2451**	**1.0969**
		-1.3231	**0.5071**	**1.2127**	**-0.8402**	**0.0038**		

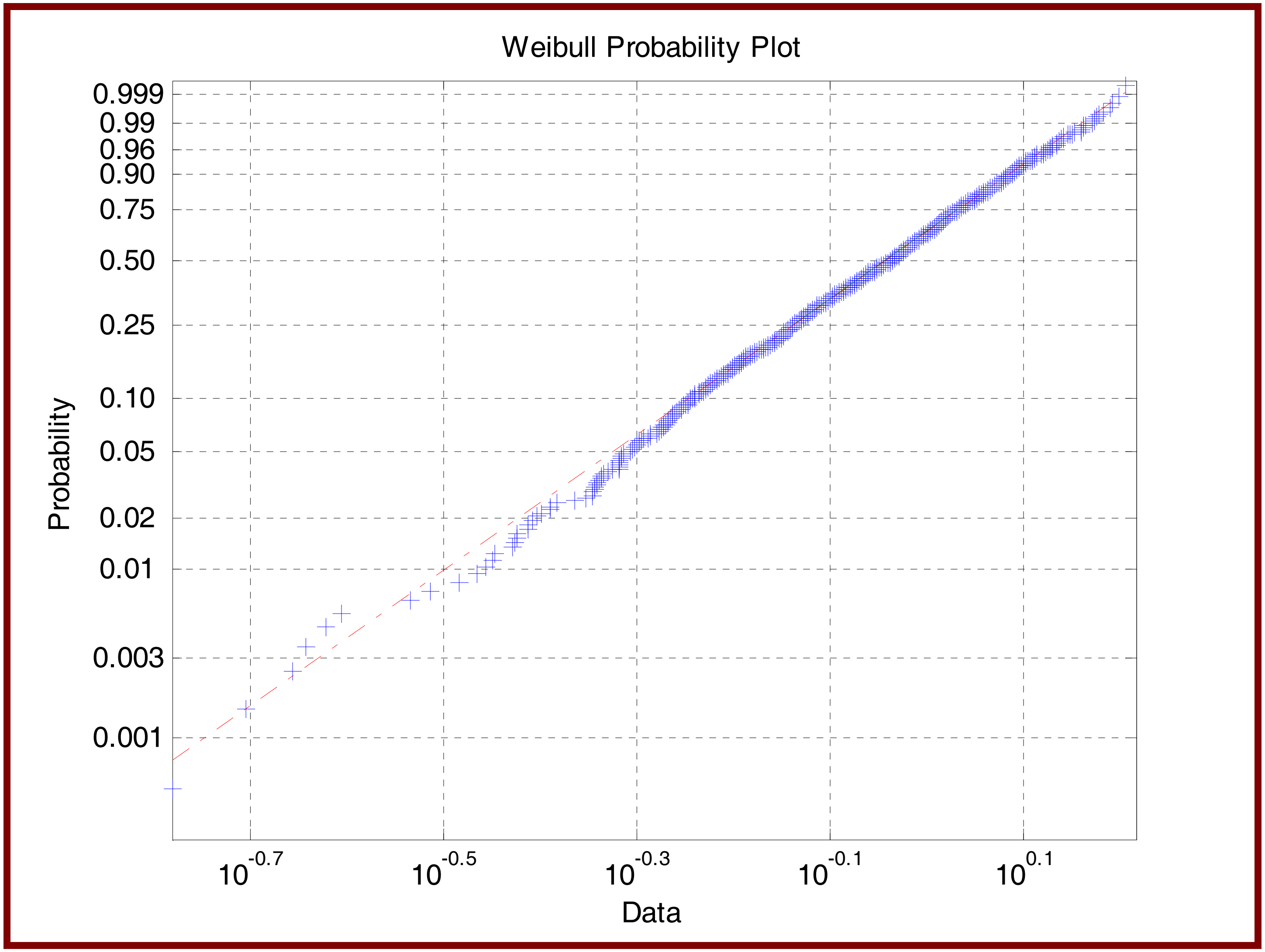
$F_n(x)$

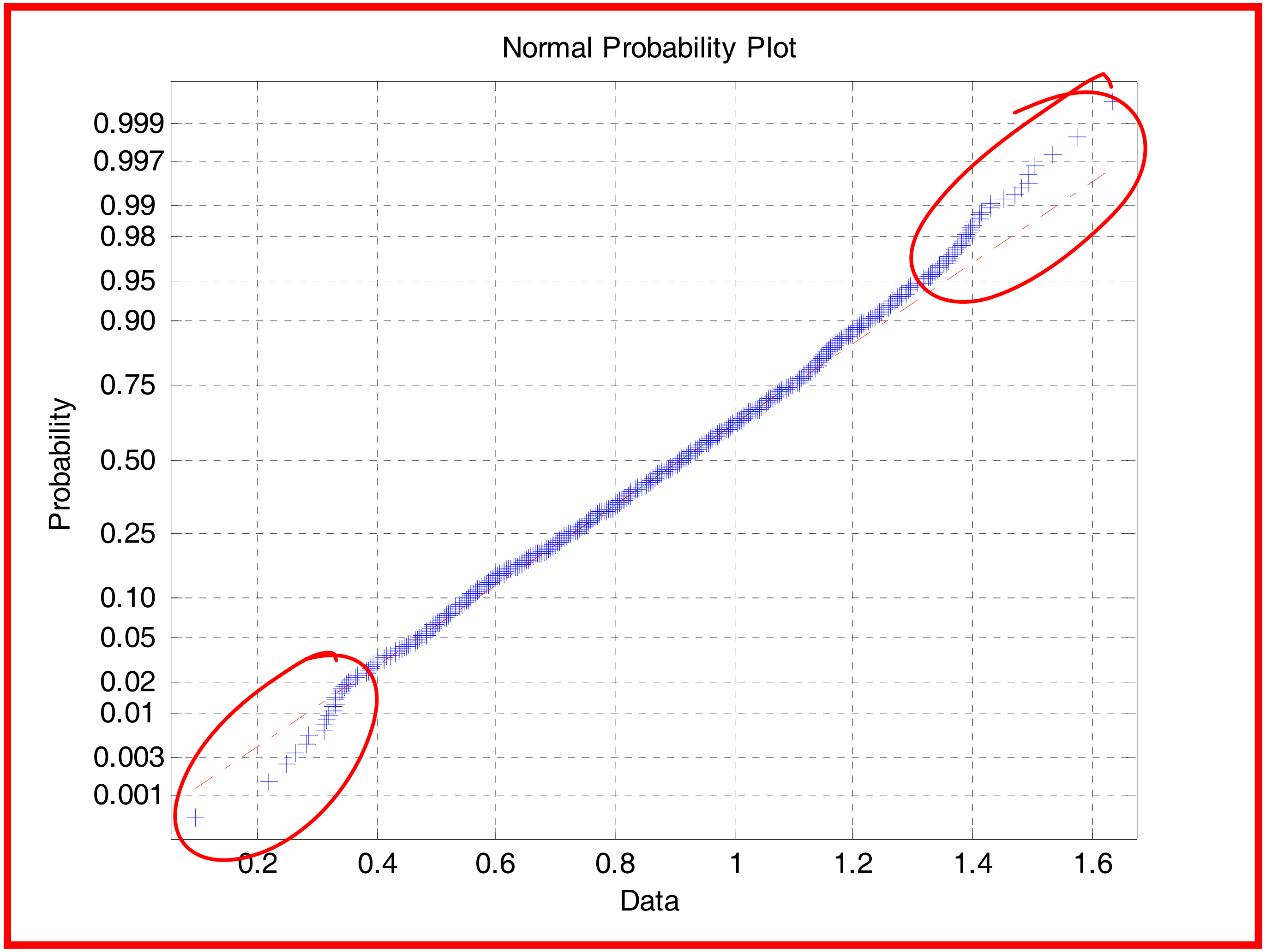






Weibull Probability Plot
0.99
0.96
0.90
0.75
0.50
0.25
0.10
0.05
0.02
0.01
Probability
10^{-2}
10^{-1}
10^{0}
Data


Weibull Probability Plot
0.999
0.99
0.96
0.90
0.75
0.50
0.25
0.10
0.05
0.02
0.01
0.003
0.001
Probability
$10^{-0.7}$
$10^{-0.5}$
$10^{-0.3}$
$10^{-0.1}$
$10^{0.1}$
Data


Normal Probability Plot
Probability
0.999
0.997
0.99
0.98
0.95
0.90
0.75
0.50
0.25
0.10
0.05
0.02
0.01
0.003
0.001
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Data

Kolmogorov - Smirnov test

Step 1

H_0 : X has a specified distribution $P_X(x)$.

H_A : The PDF of X is other than what is specified.

Step 2 : Choose α [level of significance;0.01,0.05,0.1]

$1-\alpha=$ Level of confidence.

Step 3 : Define $X^{(i)}=i^{\text{th}}$ largest value in observed sample.

$$P^*\left(X^{(i)}\right)=\frac{i}{n}$$

$$D_2=\max_{i=1}^{n}\left[\left|P^*\left(X^{(i)}\right)-P_X(x)\right|\right]$$

$x = X^{(i)}$

Step 4 : **B**ased on the sample obtain an estimate of the test statistic.

Step 5 : Define the region of rejection of the null hypothesis

Accept H_0 if $D_2 \leq c$.

Kolmogorov - Smirnov test : Example

Step 1

H_0 : X has a specified distribution $N(0,1)$.

H_A : The PDF of X is other than what is specified.

Step 2 : Choose $\alpha = 0.05$

Step 3 : Define $X^{(i)} = i^{\text{th}}$ largest value in observed sample.

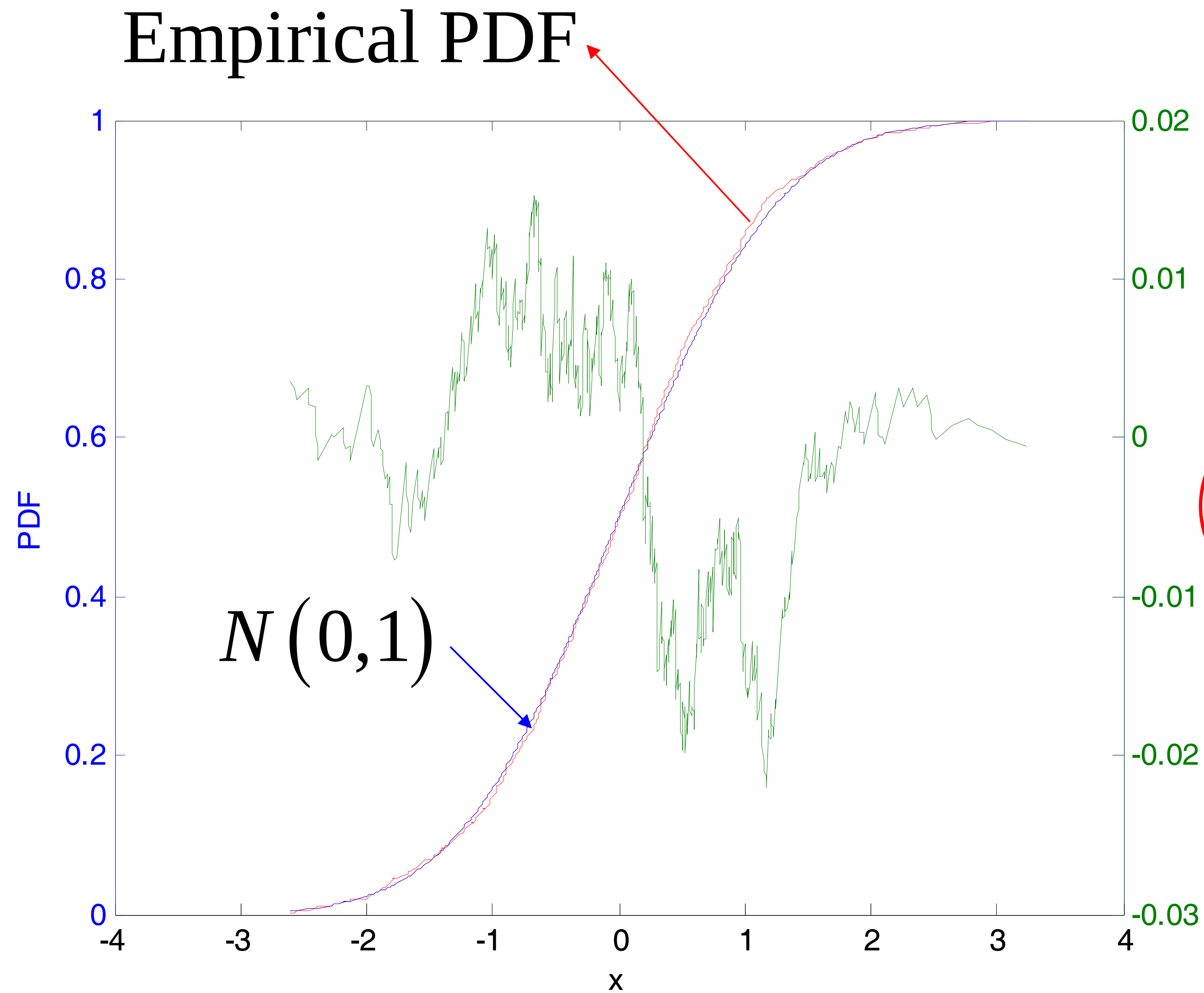
$$P^*\left(X^{(i)}\right) = \frac{i}{n}$$

$$D_2 = \max_{i=1}^{n}\left[\left|P^*\left(X^{(i)}\right) - P_X(x)\right|\right]$$

Step 4 : Based on the sample obtain an estimate of the test statistic.

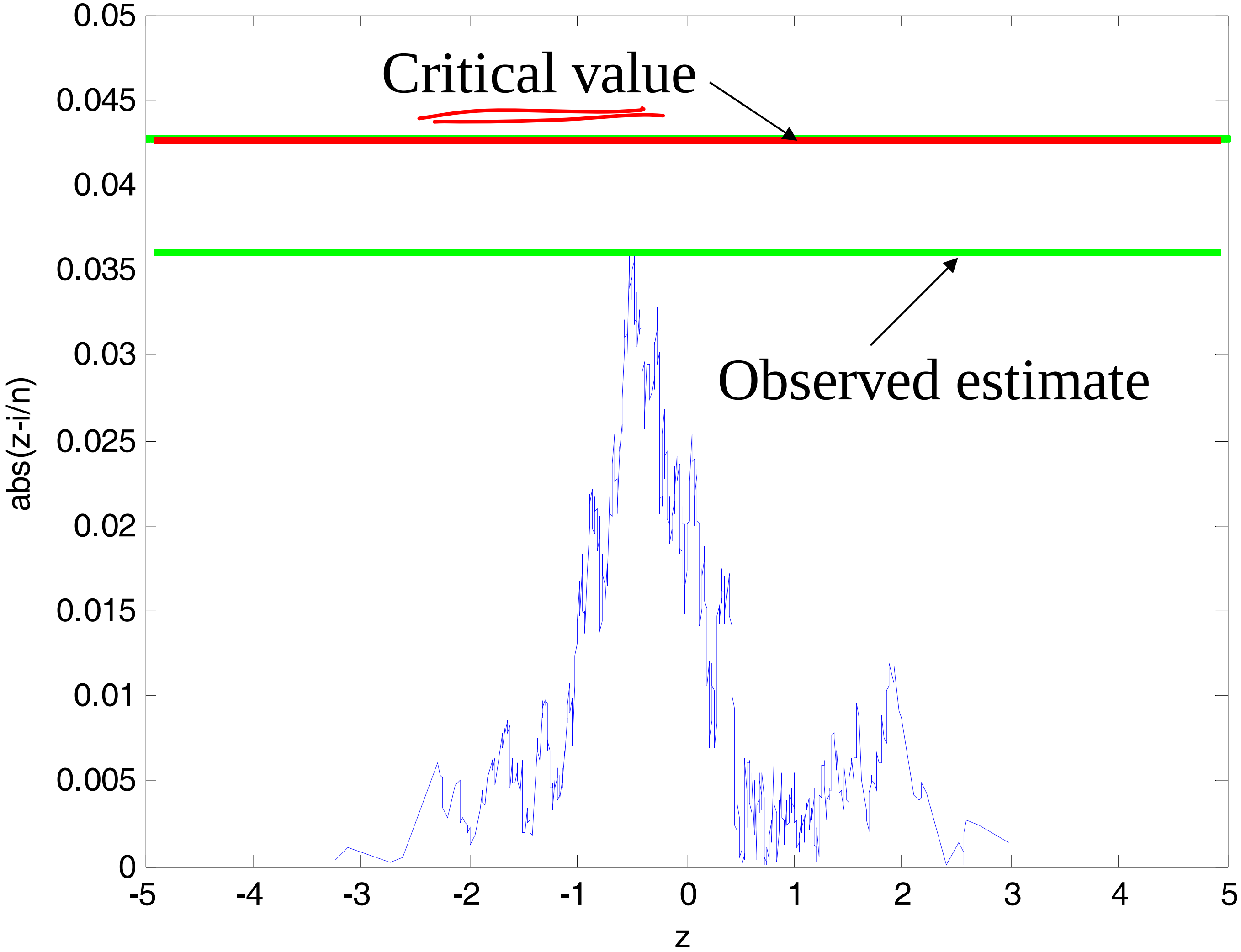
Step 5 : Define the region of rejection of the null hypothesis

Accept H_0 if $D_2 \le c$.

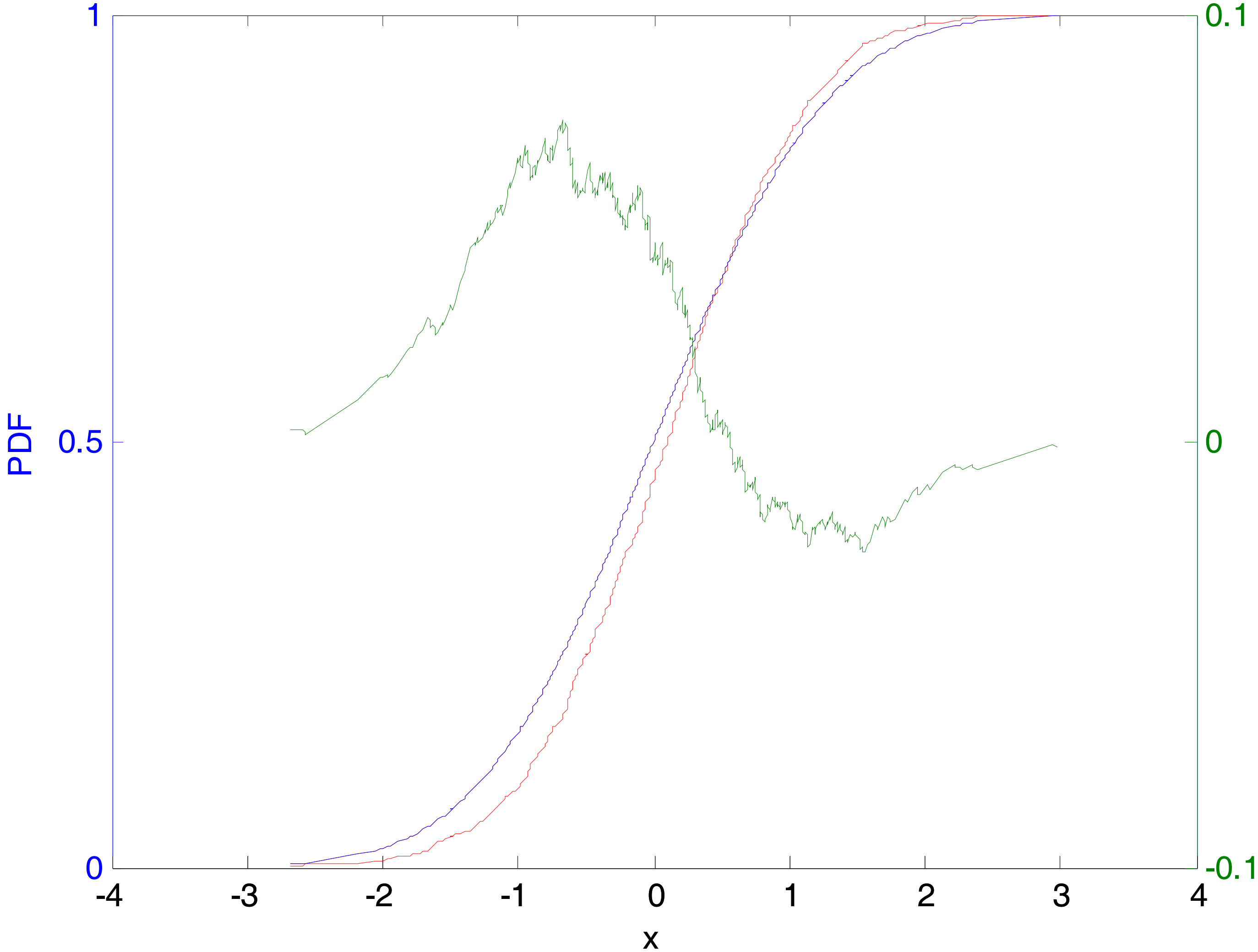

Empirical PDF
N(0,1)
PDF
x
1
0.8
0.6
0.4
0.2
0
-4
-3
-2
-1
0
1
2
3
4
0.02
0.01
0
-0.01
-0.02
-0.03

$$z - \frac{i}{n}$$

Accept H_0

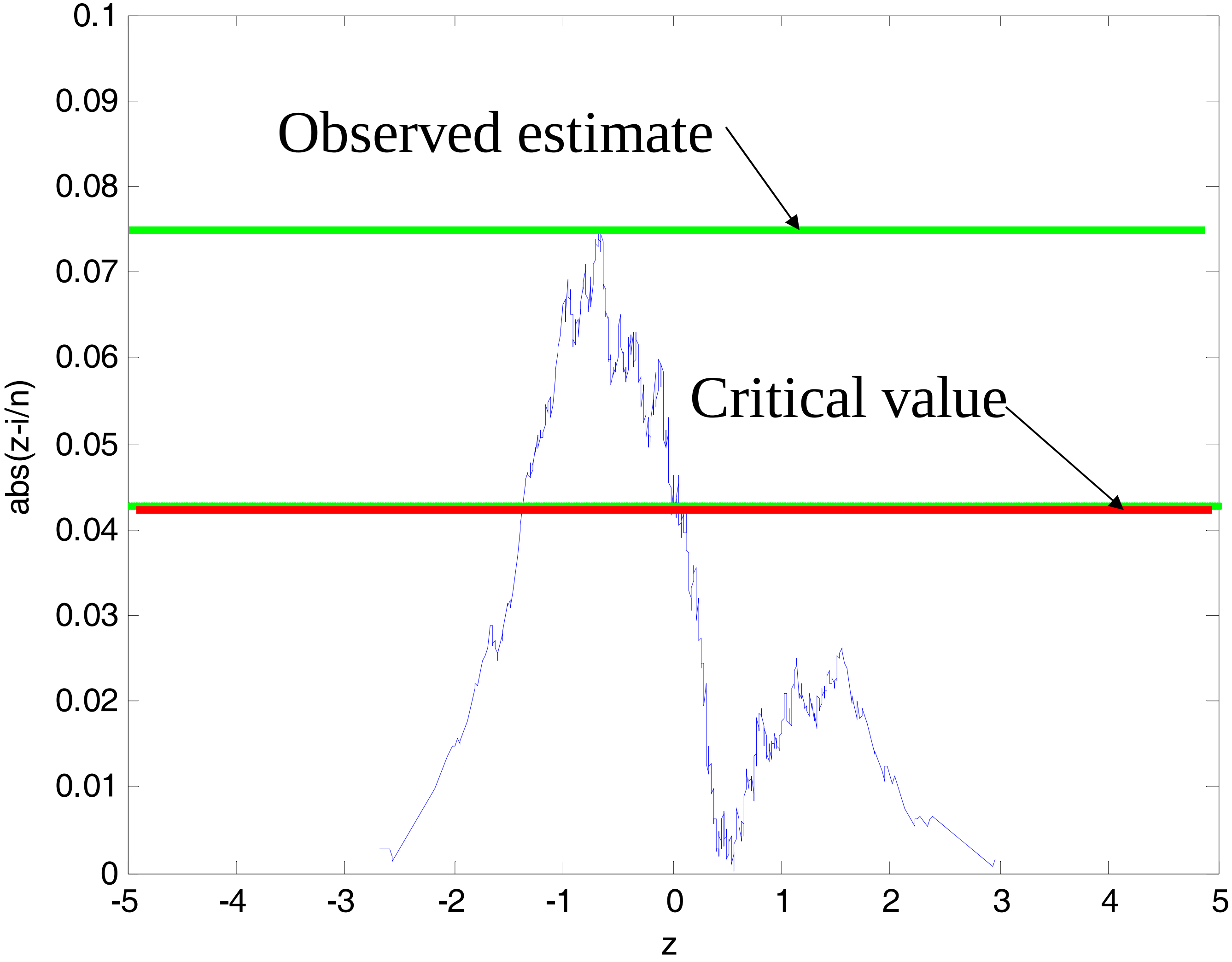


Critical value
Observed estimate
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
abs(z-i/n)
-5
-4
-3
-2
-1
0
1
2
3
4
5
z



$$z - \frac{i}{n}$$

Reject H_0


Observed estimate
Critical value
abs(z-i/n)
z
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-5
-4
-3
-2
-1
0
1
2
3
4
5

χ^2 – goodness of fit test

Statistic: Histogram's deviation from the predicted value

$$D_1 = \sum_{i=1}^{k} \frac{\left(N_i - np_i\right)^2}{np_i} \sim \chi^2 \text{ distributed with } k\text{-1 dof}$$

k = number of intervals

N_i = observed frequencies

np_i = frequencies calculated from the assumed theoretical model for the PDF

Digital simulation of samples of random variables

Let X be a random variable with PDF $P_X(x)$.

How to generate samples $\{x_i\}_{i=1}^{n}$ of X on a computer so that the estimated model for PDF of X from the data $\{x_i\}_{i=1}^{n}$ matches with the target PDF $P_X(x)$?

Pseudo-random number generators