Engineering Maths EMAT 30007, Parameter Estimation(1)

1 Estimation of population mean - confidence intervals

The confidence intervals for the parameter μ that we described in this section are all based on point estimates that could be assumed to be approximately normally distributed.

(a) To work out the coefficient in front of the standard error in the margin of error (μ estimate$\pm$margin of error) you can use the commands `norminv and tinv` in the following code lines:

In case of normal distribution:

```
alpha=0.05; % for 95% confidence level
zv=norminv(1-alpha/2) % z-value for two-sided CI
zv=norminv(1-alpha) % z-value for one-sided CI
```

In case of t-distribution:

```
n=25; % sample size
tv=tinv(1-alpha/2, n-1) % t-value for two-sided CI
tv=tinv(1-alpha, n-1) % t-value for one-sided CI
```

(b) The following measurement were recorded for the drying time (in hours) of a certain paint: 3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8. Assume that measurements represent a random sample from a normal distribution. Construct a two-sided 95% confidence interval for working out the mean drying time.

```
>> clear all
>> x=[3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8] % random samp
>> n=length(x);
>> mx=mean(x);
>> s=std(x);
>> alpha=0.05 ; % for 95% confidence
>> tcr=tinv(1-alpha/2, n-1); % critical t value
>> me=tcr*s/sqrt(n); % margin of error
>> CI1= mx-me; % lower CI bound
>> CI2= mx+me; % upper CI bound
```

(c) Why was t-distribution used?

(d) Explain what the phrase *95% confidence* implies. Can we say that the mean drying time lies in the interval with probability 95%?

(e) Now construct a 90% confidence interval for the mean drying time by changing one line of the code.

(f) Many cardiac patients were implanted pacemakers to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 and an approximate normal distribution, find a 95% confidence interval for the mean of all connector modules made by a certain company. A random sample of 75 modules has an average of 0.310 inch.

```
>> clear all
>> mx=0.310;
>> s=0.0015;
>> n=75
>> alpha=0.05; % for 95% confidence level
>> zv=norminv(1-alpha/2) % z-value for two-sided CI
>> me=zv*s/sqrt(n); % margin of error
>> CI1= mx-me; % lower CI bound
>> CI2= mx+me; % upper CI bound
```

(g) Download data from WORKHRS.xls from Blackboard. Find a 95% confidence interval for estimating the mean using both z-score and t-score. Compare the results.

(h) Download data from ROBOTS.xls from Blackboard. Estimate μ of wheels on all social robots with 99% confidence. In repeated sampling, what proportion of confidence intervals will contain the true mean?

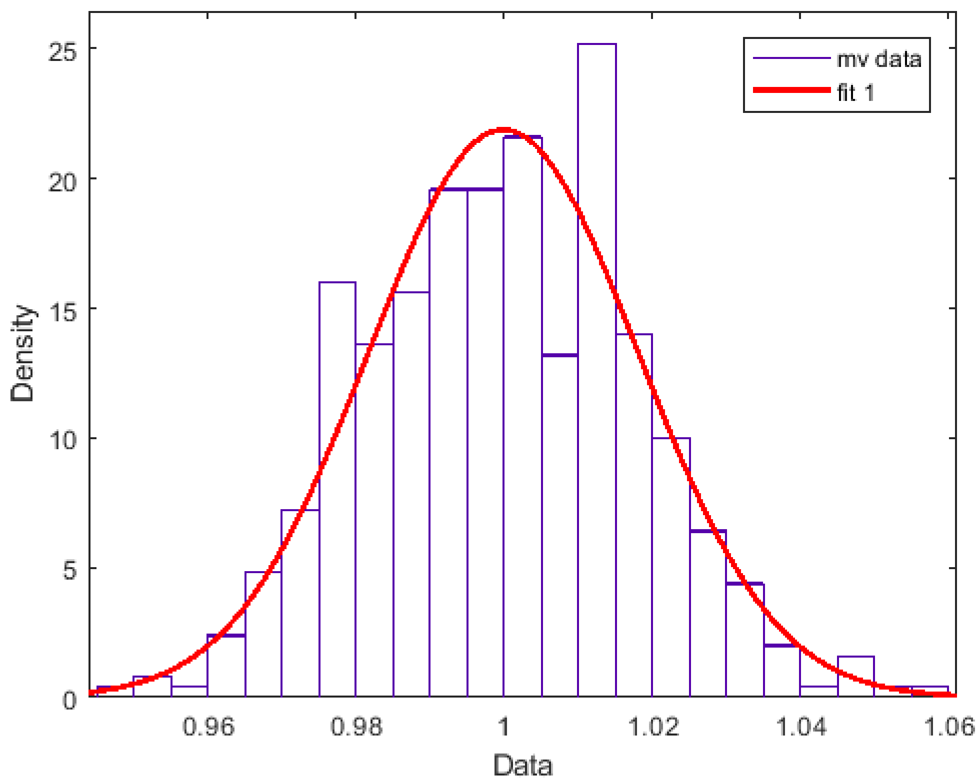
2 Random sampling from distributions (Monte Carlo simulations): Use the commands `***rnd` (e.g., `normrnd, weibullrnd, ...`). For visual representation of data use `Distribution fitter` in `Apps`.

(a) Below is the example of the code that generates normal sample of 100 elements with $\mu = 1$ and $\sigma = 0.2$. The experiment is repeated 500 times. The graph shows the distribution (approximated to normal) of the means from simulation of a random sampling.

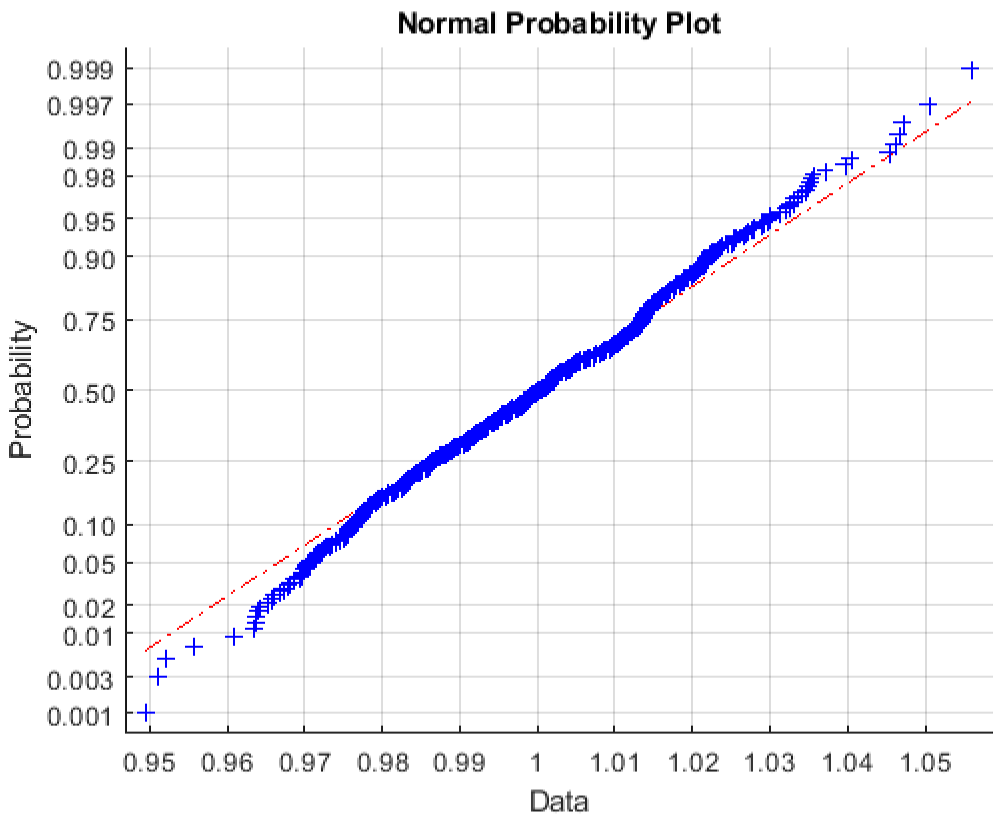
```
>> clear all
>> rng(31) % choose your seed
>> for i = 1 : 500 % take 500 random samples
>> x = normrnd(1,0.2, [1 100]) % each sample contains 100 elements
>> m = mean(x)
>> mv(i) = m; % save means of samples
>> end

Mean:             0.999886
Variance:         0.000332487

Parameter  Estimate   Std. Err.
mu          0.999886  0.000815459
sigma      0.0182342  0.000577484
```



Use `normplot(mv)` to assess visually whether the sample data comes from a population with a normal distribution (see graph below).



(b) There is a normal population of rainfalls with mean 900 mm and standard deviation 200 mm. Show using simulations that the more data that is collected, the more accurate the μ estimate and hence the smaller its standard error.

Do you expect any 'unlucky samples' which mean $\bar(X)$ may not be included into 95% confidence interval?

(c) Show that the sample mean is less variable than individual values from the population so the mean's standard error is less than the standard deviation.

Note that standard deviation describes the variability of sample values; the standard error of an estimator describes the variability of that estimator.

(d) Compare mean estimator with a median estimator. Which one has a less spread of errors? Are they all unbiased?

3 Central Limit Theorem:

Even if population is not normally distributed, the Central Limit Theorem states that the sampling distribution of the sample mean becomes normal for $n > 30$.

(a) You can run the code below to verify that the distribution of means is approximately normal in Poisson distribution for large n.

```
clear all
rng(31) % choose your seed
for i = 1 :500
x = poissrnd(2,[1 100])
pm = mean(x)
pmv(i) = pm;
end
```

(b) Change the code for $n < 30$.

(c) **When sample size is 30 or more, we consider the sample size to be large and there is no need to check whether the sample comes from a Normal Distribution. We can use the t-distribution.**