Consequences :-

- Extremely useful & very necessary to appreciate.

- For practical applications, it is an important tool in inference.

Confidence Interval :-

- Already we have seen the 68, 98, 99. - rule for identifying with normal.

- $P(|Z| \le 1.96) = 0.95$, $Z \sim N(0,1)$

Suppose $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \text{ "}\simeq\text{" } Z$ for large enough n

then $P\left(\left|\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)\right| \le 1.96\right) \simeq 0.95$

↑ very close for n-large

$\Longleftrightarrow P\left(-\frac{\sigma}{\sqrt{n}} 1.96 \le \bar{X}_n - \mu \le \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95$

$$\Leftrightarrow \quad \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}}1.96 + \overline{X}_n \le \mu \le \overline{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95$$

true mean

$$P\left(\mu \in \left(-\frac{\sigma}{\sqrt{n}}1.96 + \overline{X}_n,\ \overline{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right)\right) = 0.95$$

So instead of giving a - point estimator - $\overline{X}_n$ we are able to provide a Confidence - interval for μ

Questions :-

1. How large should n be for Central limit theorem to take shape?

   i.e. when is $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \equiv N(0,1)$

2. Is n the same no matter what the distribution of X is ?

3. What does Confidence interval exactly mean?

   Does it mean : Given a realised interval

   $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}_n , \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right)$ there is

   95% chance μ is in it?

4. What if σ is not known?

5. Can we say something for n small?

Confidence Intervals

Using the Central Limit Theorem for large n we have

$$P\left(\left|\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right| \leq 1.96\right) \approx 0.95$$

which is the same as saying

$$P\left(\mu \in \left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)\right) \approx 0.95$$

The interval $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$ is called the 95% confidence interval for μ.

Confidence Intervals

95% confidence interval for μ is $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$

Meaning: for n large if we did m (large) repeated trials and computed the above interval for each trial then true mean would belong to approximately 95% of m intervals calculated.

Confidence Intervals

The below is code for finding the confidence interval for a data x.

```
> cifn = function(x, alpha=0.95){
+ z = qnorm( (1-alpha)/2, lower.tail=FALSE)
+ sdx = sqrt(1/length(x))
+ c(mean(x) - z*sdx, mean(x) + z*sdx)
+ }
```

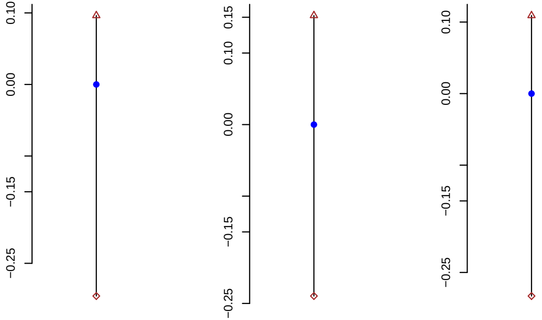
Three Confidence Intervals for Normal(0,1)

```
> x1 = rnorm(100,0,1);y = cifn(x1)
> y
[1] -0.29570433  0.09628847
> x2 = rnorm(100,0,1);z = cifn(x2)
> z
[1] -0.2396115  0.1523813
> x3 = rnorm(100,0,1);w = cifn(x3)
> w
[1] -0.2829300  0.1090628
```

Does 0 belong to all the three confidence intervals ?

Confidence Intervals Plots

The below is a plot of the three confidence intervals computed in the previous slide.



Confidence Intervals : 10 Trials

We generate 10 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

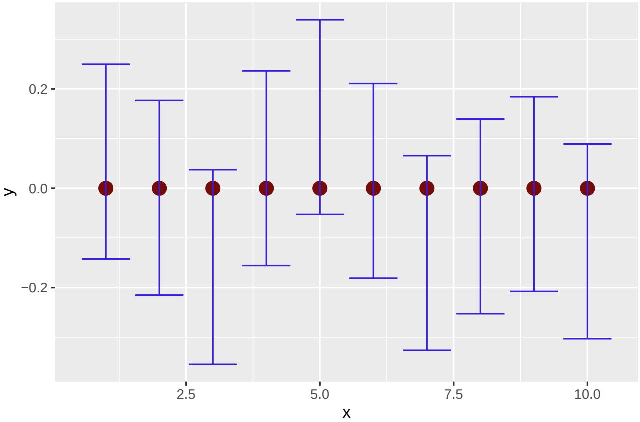
```
> normaldata = replicate(10, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)

TRUEIN
TRUE
  10
```

Confidence Intervals : 10 Trials



Confidence Intervals: 40 Trials

We generate 10 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

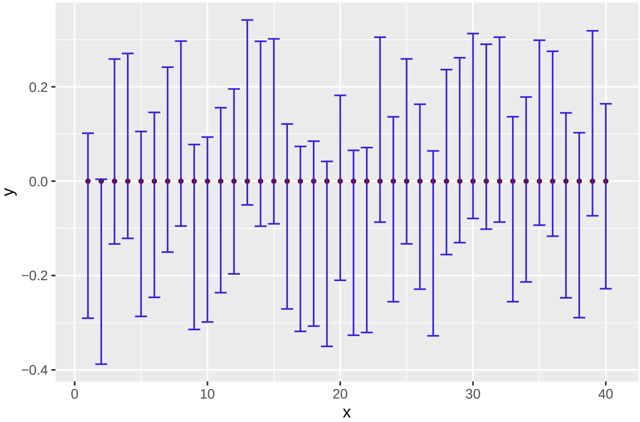
```
> normaldata = replicate(40, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)

TRUEIN
TRUE
  40
```

Confidence Intervals: 40 trials Plot



Confidence Intervals : 100 Trials

We generate 100 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

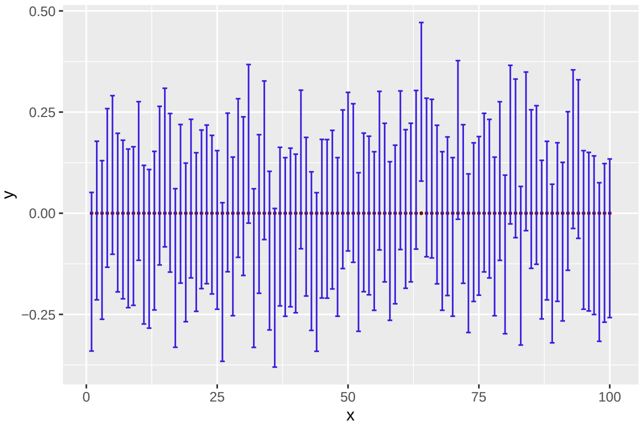
```
> normaldata = replicate(100, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)

TRUEIN
FALSE  TRUE
    1    99
```

Confidence Intervals : 100 Trials



Confidence Intervals : 1000 Trials

We generate 1000 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

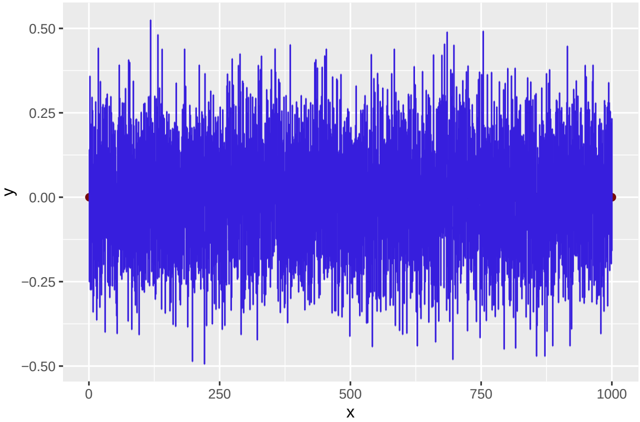
```
> normaldata = replicate(1000, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)

TRUEIN
FALSE  TRUE
   51   949
```

Confidence Intervals : 1000 Trials



Confidence Intervals

95% confidence interval for μ is $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$

Meaning: for n large if we did m (large) repeated trials and computed the above interval for each trial then true mean would belong to approximately 95% of m intervals calculated.

Thus numerically the above meaning seems to hold for a Normal population.