Strong Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Assume that X_1 has finite mean μ and $E \mid X_1 \mid < \infty$

$$A = \left\{ \lim_{n\to\infty} \underbrace{\frac{X_1 + X_2 + \ldots + X_n}{n}}_{\bar{X}_n} = \mu \right\},$$

then

$$P(A) = 1.$$

$A = \{ \omega \in \Omega \mid \forall \varepsilon > 0 \ \exists N \text{ s.t. } |\bar{X}_n - \mu| < \varepsilon \ \forall n \geq N \}$

$= \{ \omega \in \Omega \mid \forall k \geq 1 \ \exists N \text{ s.t. } |\bar{X}_n - \mu| < \frac{1}{k} \ \forall n \geq N \}$

$= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{ \omega \in \Omega \mid |\bar{X}_n - \mu| < \frac{1}{k} \}$

$A_n^k = \{ \omega \in \Omega \mid |\bar{X}_n - \mu| < \frac{1}{k} \}$

Fix $k \geq 1$.

Set: $B_N^k = \bigcap_{n=N}^{\infty} A_n^k$

Note: $\forall N \geq 1 \quad B_N^k \subseteq B_{N+1}^k$

$$B_N^{\kappa} \uparrow \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n^{\kappa} \quad -(1)$$

As $P(A)=1$ $\therefore \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n^{\kappa} \supseteq A$ we have

$$P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n^{\kappa}\right) = 1 \quad -(2)$$

From (1) and (2)

$$P(B_N^{\kappa}) \uparrow P\left[\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n^{\kappa}\right) = 1 \quad -(3)$$

as $N \to \infty$

$\Rightarrow \forall \delta > 0 \quad \exists M > 1$ s.t. $\forall N \geq M$

$$|P(B_N^{\kappa}) - 1| < \delta \quad \forall N \geq M \quad -(4)$$

For $N \geq M$

$$B_N^{\kappa} = \bigcap_{n=N}^{\infty} A_n^{\kappa} \subseteq A_N^{\kappa}$$

From (4), (3) we have $\exists M$:

$$\mathbb{P}(A_N^k) \geq 1-\delta \quad \forall N \geq M$$

$\Rightarrow \exists M$: $\mathbb{P}(\{\omega \in \Omega \mid |X_N - \mu| < \frac{1}{k}\}) \geq 1-\delta \quad \forall N \geq M$

$\Rightarrow \exists M$: $\mathbb{P}(\{\omega \in \Omega \mid |X_N - \mu| > \frac{1}{k}\}) < \delta \quad \forall N \geq M$

As $\delta > 0$ and $k \geq 1$ was arbitrary we have shown

Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Assume that X_1 has finite mean μ and finite variance σ^2. Then for any $\epsilon > 0$

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0,$$

Point Estimate

Let $p \in \mathbb{R}$ and $X_1, X_2, \ldots, X_n$ be from a population distributed as Normal with mean p and variance 1.

Method of moments Estimate:

$m_1(X_1, X_2, \ldots X_n) = \bar{X}_n$

$\mu_1 = p \Rightarrow$ $\boxed{\hat{p} = \bar{X}_n}$

Maximum likelihood Estimate:

- $L(p \mid x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - p)^2}$
- $\max_p L(p) \equiv \min_p \sum_{i=1}^{n}(x_i - p)^2$
- As: $\sum_{i=1}^{n}(x_i - p)^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{n}(\bar{x}_n - p)^2$

Ex: $\Rightarrow$ $\boxed{\hat{p} = \bar{X}_n}$

R-Simulation

```
> set.seed(2021)
> x= rnorm(100)
> mean(x)
[1] -0.1744844
```

$p = 0$, variance $= 1$

$= \bar{X}_n$

Q: Can we quantify $|\bar{X}_n - p|$?

A: $\bar{X}_n \overset{d}{=} \text{Normal}\left(p, \frac{1}{n}\right)$

$P\left(|\bar{X}_n - p| \leq \frac{3}{\sqrt{n}}\right) = ?$

Confidence Intervals

$\hat{p} - p \sim N(0, \frac{1}{n})$

As $\hat{p} \sim \text{Normal}(p, 1/n)$, we have $\sqrt{n}(\hat{p} - p) \sim \text{Normal}(0, 1)$. So,

$$P(-3 \leq \sqrt{n}(\hat{p} - p) \leq 3) = \Phi(3) - \Phi(-3)$$

$= P(-3 < Z < 3)$ where $Z \sim \text{Normal}(0,1)$

```
> pnorm(3) - pnorm(-3)
```

```
[1] 0.9973002
```

$$P\left(p \in \left(\hat{p} - \frac{3}{\sqrt{n}}, \hat{p} + \frac{3}{\sqrt{n}}\right)\right) = 0.9973$$

$|\sqrt{n}(\hat{p} - p)| < 3 \iff \hat{p} - \frac{3}{\sqrt{n}} < p < \hat{p} + \frac{3}{\sqrt{n}}$

$\iff p \in \left(\hat{p} - \frac{3}{\sqrt{n}}, \hat{p} + \frac{3}{\sqrt{n}}\right)$

$$P\left(p \in \left(\hat{p} - \frac{3}{\sqrt{n}}, \hat{p} + \frac{3}{\sqrt{n}}\right)\right) = 0.9973$$

$\mathbb{P}\left(p \in \left(\bar{X}_n - \frac{3}{\sqrt{n}}, \bar{X}_n + \frac{3}{\sqrt{n}}\right)\right) = 0.993$

We say:

$\left(\bar{X}_n - \frac{3}{\sqrt{n}}, \bar{X}_n + \frac{3}{\sqrt{n}}\right)$ is a 99% Confidence interval for p.

Shown:

$$\mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; p \in \left(\hat{p}(\omega) - \frac{3}{\sqrt{n}}, \hat{p}(\omega) + \frac{3}{\sqrt{n}}\right)\right\}\right) = 0.99$$

Experiment: $\omega_0 \in \Omega$ is fixed.

$$x_1^{(\omega_0)}, \dots x_n^{(\omega_0)} \rightsquigarrow \hat{p}_1^{(\omega_0)} = \bar{x}_n^{(\omega_0)}$$

$$\left(\hat{p}(\omega_0) - \frac{3}{\sqrt{n}}, \hat{p}(\omega_0) + \frac{3}{\sqrt{n}}\right) = I_n$$

Fact:

(X) That $p \in \left(\hat{p} - \frac{3}{\sqrt{n}}, \hat{p} + \frac{3}{\sqrt{n}}\right)$ has 99% probability

(V) Repeat the experiment say 100 times

- Fix $n \geq 1$.
- Construct $I_n^1, I_n^2, \dots, I_n^{100}$
- 99% of them contains p

Confidence Intervals

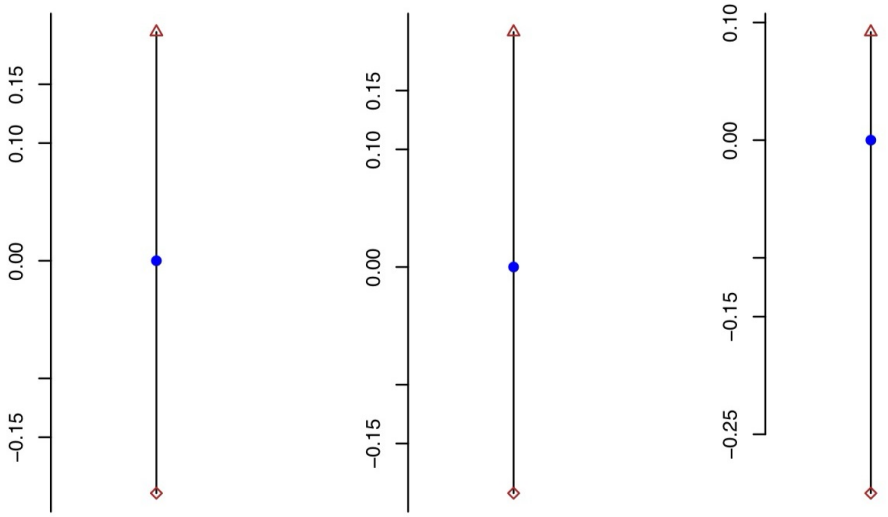
The below is code for finding the confidence interval for a data x.

```
> 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] -0.1975784  0.1944144
> x2 = rnorm(100,0,1);z = cifn(x2)
> z
[1] -0.1922503  0.1997425
> x3 = rnorm(100,0,1);w = cifn(x3)
> w
[1] -0.30008082  0.09191197
```

Does 0 belong to all the three confidence intervals ?



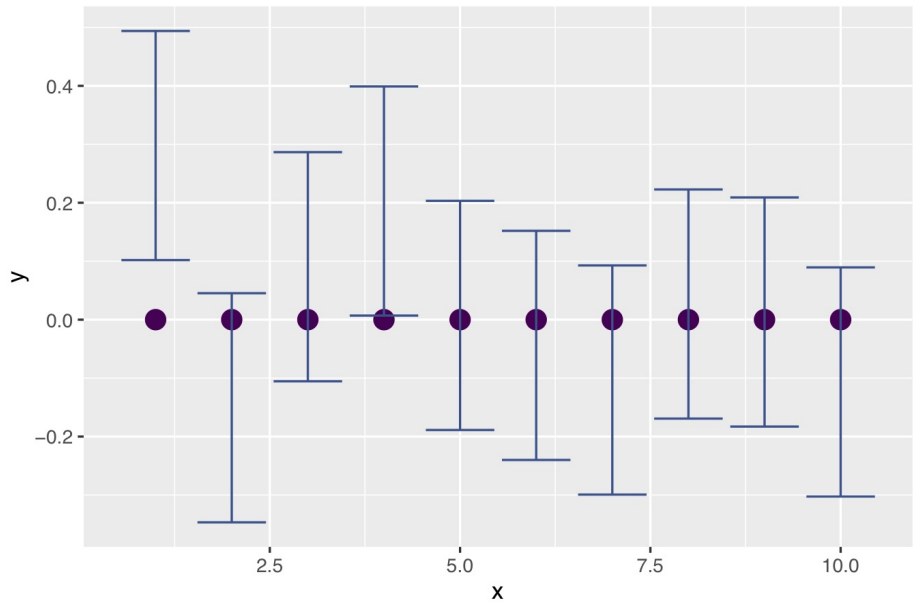
Confidence Intervals : 10 Trials

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
FALSE  TRUE
    2     8
```



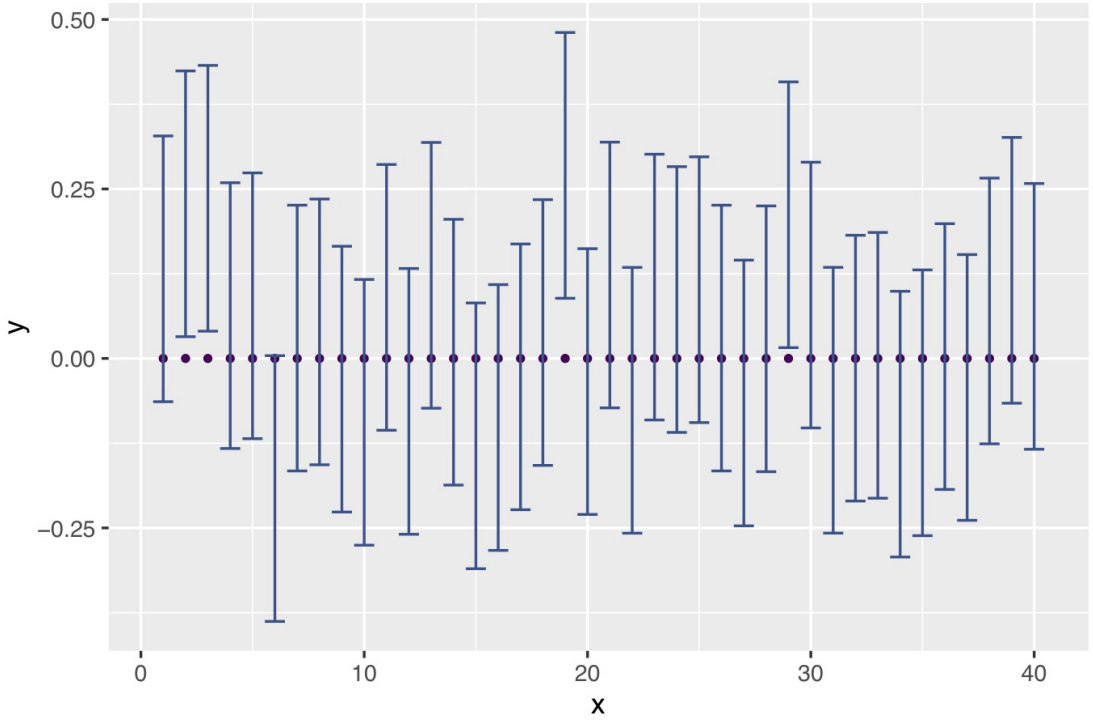
Confidence Intervals: 40 Trials

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

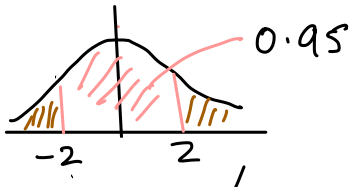
```
> normaldata = replicate(40, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

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

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)
TRUEIN
FALSE  TRUE
    4    36
```



Confidence Intervals



To get a confidence level of 95%, we want a factor z such that

$$P(p \in (\hat{p} - \tfrac{z}{\sqrt{n}}, \hat{p} + \tfrac{z}{\sqrt{n}})) = P(|\ Z\ | < z) = 0.95$$

That is we want a factor z such that

$$P(Z > z) = P(Z < -z) = 0.05/2 = 0.025,$$

Such a z is easily calculated as

```
> -qnorm(0.025)

[1] 1.959964
```

95% - C.I. for p

$$\left(\hat{p} - \frac{1.96}{\sqrt{n}},\ \hat{p} + \frac{1.96}{\sqrt{n}}\right)$$

Given: $\hat{p} = \bar{X}_n$

Find z: $P(Z \le -z) = \left(1 - \frac{\alpha}{100}\right)\frac{1}{2}$

$P(Z \le z) = \left(1 + \frac{\alpha}{100}\right)\frac{1}{2}$

Using R: $z = qnorm\left(\left(1 + \frac{\alpha}{100}\right)\frac{1}{2}\right)$

or

$z = -qnorm\left(\left(1 - \frac{\alpha}{100}\right)\frac{1}{2}\right)$

Similarly, for a confidence level of 80%,

α% - C.I. for p

$$\left(\hat{p} - \frac{z}{\sqrt{n}},\ \hat{p} + z/\sqrt{n}\right)$$

```
> -qnorm(0.10)

[1] 1.281552
```

80% - C.I for p

$$\left(\hat{p} - \frac{1.28}{\sqrt{n}},\ \hat{p} + \frac{1.28}{\sqrt{n}}\right)$$

Empirical Data

If we consider a specific empirical realisation of $X_1, X_2, \ldots X_n$, say $n = 15$ data points:

$$11.22, 9.56, 10.06, 10.21, 10.95, 10.03, 10.75, 10.71, 11.42, 9.61, 10.91, 8.14, 8.95, 10.57, 11.1.$$

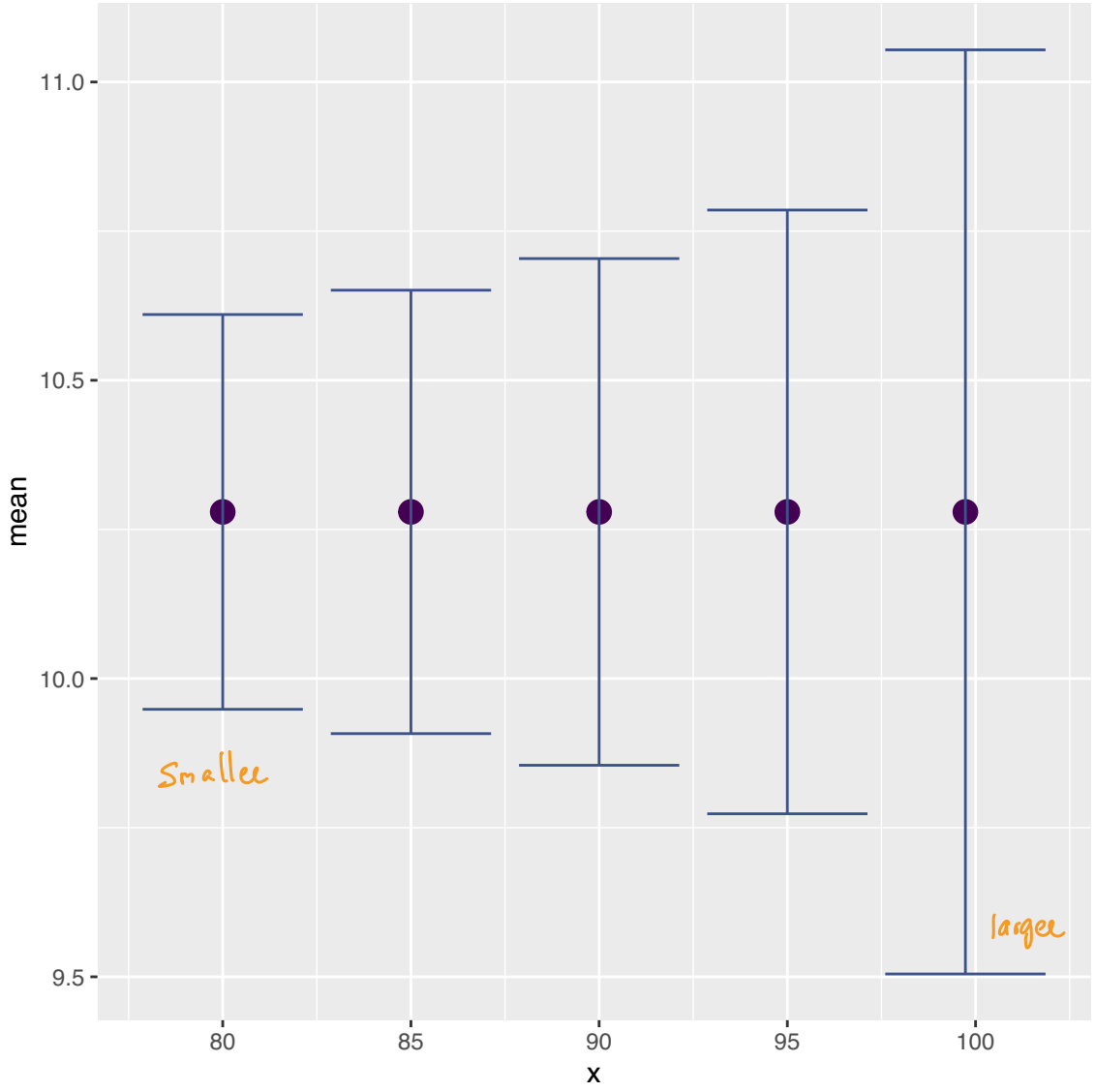
Here, $\hat{p} = 10.279$, so the 99.73% confidence interval is given by

```
> x = c(11.22, 9.56,
+   10.06, 10.21, 10.95, 10.03, 10.75, 10.71, 11.42, 9.61, 10.91,
+   8.95, 10.57, 11.1)
> mean= mean(x)
> x1 = c(mean(x) -(qnorm((1+0.9973)/2))/sqrt(length(x)),
+        mean(x) + (qnorm((1+0.9973)/2))/sqrt(length(x)))
> x2=c(mean(x) -(qnorm((1+0.95)/2))/sqrt(length(x)),
+      mean(x) + (qnorm((1+0.95)/2))/sqrt(length(x)))
> x3=c(mean(x) -(qnorm((1+0.9)/2))/sqrt(length(x)),
+      mean(x) + (qnorm((1+0.9)/2))/sqrt(length(x)))
> x4=c(mean(x) -(qnorm((1+0.85)/2))/sqrt(length(x)),
+      mean(x) + (qnorm((1+0.85)/2))/sqrt(length(x)))
> x5=c(mean(x) -(qnorm((1+0.8)/2))/sqrt(length(x)),
+      mean(x) + (qnorm((1+0.8)/2))/sqrt(length(x)))
> cidata= t(cbind(x1,x2,x3, x4,x5))
```

x1 - 95% C.I for $\hat{p}$ .... x5 - 80% C.I. for $\hat{p}$

Empirical Data

```
> require(ggplot2)
> df = data.frame(x= c(99.73, 95, 90, 85,80),
+                 z1 = cidata[,1], z2=cidata[,2], mean )
> ggplot(df, aes(x=x, y=mean)) +
+   geom_point(size=4, color="#440154") +
+   geom_errorbar(aes(ymax= z2, ymin=z1), color="#3b528b")
```



Pivotal Quantity

The key observation: $\sqrt{n}(\hat{p} - p) \sim \text{Normal}(0, 1)$.

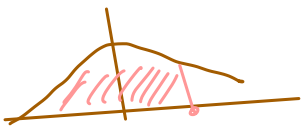
In other words, we have found a function of the data and the parameter of interest, namely,

$$T(X_1, X_2, \ldots, X_n, p) = \sqrt{n}(\hat{p} - p),$$

so that regardless of the value of p, $T(X_1, X_2, \ldots, X_n, p)$ has a completely known distribution.

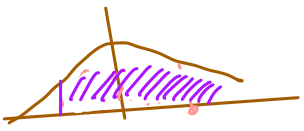
- Derivation of the interval works as $\sqrt{n}(\hat{p}-p) \overset{\cdot}{\sim} N(0,1)$ "free of p"
- odd ∴ we found symmetric $\alpha/100$ c. intervals. Actually we can find any a, b s.t. $P(Z \in [a,b]) = \alpha/100$
- but okay:

$a = \infty$, $b = \Phi^{-1}(\alpha/100)$



one-sided $\frac{\alpha}{100}$ - C.I.

$a = \Phi^{-1}(1 - \frac{\alpha}{100})$, $b = \infty$



Interval is $\{p : T(X_1, X_2, \ldots, X_n, p) \in [a, b]\}$.

Pivotal Quantity-Exponential

Let $\lambda > 0$ and $(X_1, X_2, \ldots, X_n)$ be an i.i.d. sample from an Exponential population with rate λ. We are interested in obtaining a confidence interval for λ. Take $X_i \overset{iid}{\sim} Exp(\lambda)$

$$T(X_1, X_2, \ldots, X_n, \lambda) = n\lambda\bar{X}.$$

$$= \lambda(X_1 + X_2 + \cdots + X_n)$$

$$\overset{d}{=} Gamma(n, 1)$$

$$\| Z$$

C.I: $= a < n\lambda\bar{x} < b$

```
> beta <- 0.95
> c(0, qgamma(beta, shape = 15))
[1]  0.00000 21.88649
> c(qgamma(1 - beta, shape = 15), Inf)
[1] 9.24633     Inf
```

As $\bar{x} > 0$

95% C.I $\left(0, \frac{21.88}{n\bar{x}}\right)$

95% C.I. $\left(\frac{9.24}{n\bar{x}}, \infty\right)$

```
> alpha <- 1 - beta
> c(qgamma(alpha / 2, shape = 15), qgamma(1 - alpha / 2,
+                                         shape = 15))
[1]  8.395386 23.489621
```

95% C.I.

$\left(\frac{8.39}{n\bar{x}}, \frac{23.48}{n\bar{x}}\right)$

Pivotal Quantity - Normal

$X_1, \ldots X_n \sim \text{Normal}(\mu, \sigma^2)$

If σ^2 is not known.

Earlier C.I. method wont work. why?

$\bar{X}$; $E[\bar{X}] = \mu$

S^2 ; $E[S^2] = \sigma^2$

If we take

$$T(X_1, X_2, \ldots, X_n, \mu) = \sqrt{n}\,\frac{(\bar{X} - \mu)}{S}. \quad \overset{d}{=} t_{n-1}$$

For 95% - C.I. :-

$n = 15$

① find t^*:

$P(|t_{n-1}| \le t^*) = 0.95$

- qt(1+0.95/2, 14)

```
> qt(0.975, df = 14)

[1] 2.144787 = t*
```

② 95% CI for μ is

$$\left(\bar{X} - \frac{S}{\sqrt{n}}(2.14),\ \bar{X} + \frac{S}{\sqrt{n}}(2.14)\right)$$

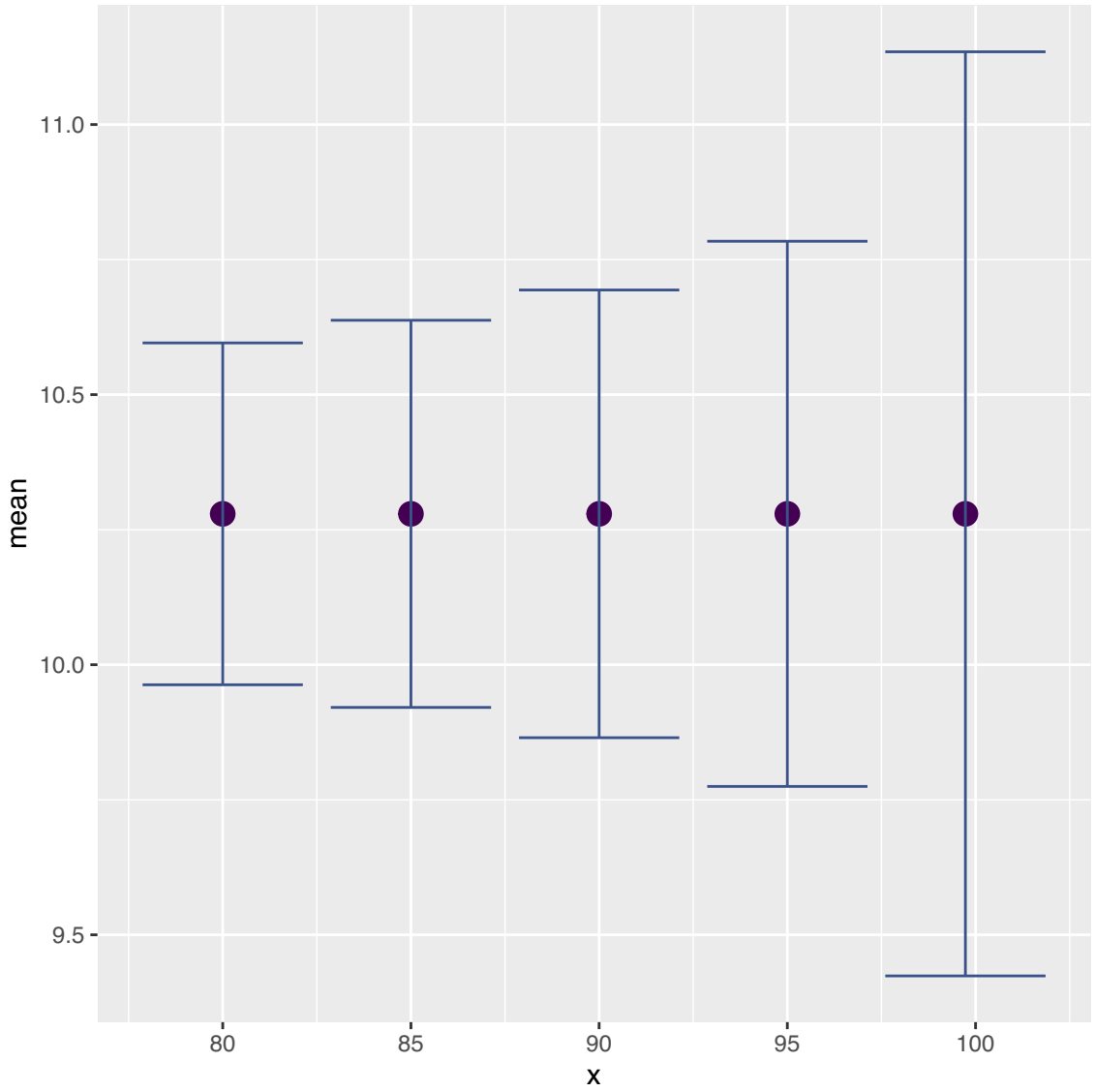
Empirical Data

Considering again the $n = 15$ data points 11.22, 9.56, 10.06, 10.21, 10.95, 10.03, 10.75, 10.71, 11.42, 9.61, 10.91, 8.14, 8.95, 10.57, 11.1,

```
> x = c(11.22, 9.56,
+   10.06, 10.21, 10.95, 10.03, 10.75, 10.71, 11.42, 9.61, 10.91,
+   8.95, 10.57, 11.1)
> mean= mean(x)
> sd = sd(x)
> x1 = c(mean(x) -(qt((1+0.9973)/2,14)*sd)/sqrt(length(x)),
+        mean(x) + (qt((1+0.9973)/2,14)*sd)/sqrt(length(x)))
> x2=c(mean(x) -(qt((1+0.95)/2,14)*sd)/sqrt(length(x)),
+      mean(x) + (qt((1+0.95)/2,14)*sd)/sqrt(length(x)))
> x3=c(mean(x) -(qt((1+0.9)/2,14)*sd)/sqrt(length(x)),
+      mean(x) + (qt((1+0.9)/2,14)*sd)/sqrt(length(x)))
> x4=c(mean(x) -(qt((1+0.85)/2,14)*sd)/sqrt(length(x)),
+      mean(x) + (qt((1+0.85)/2,14)*sd)/sqrt(length(x)))
> x5=c(mean(x) -(qt((1+0.8)/2,14)*sd)/sqrt(length(x)),
+      mean(x) + (qt((1+0.8)/2,14)*sd)/sqrt(length(x)))
> cidata= t(cbind(x1,x2,x3, x4,x5))
```

Empirical Data

```
> require(ggplot2)
> df = data.frame(x= c(99.73, 95, 90, 85,80),
+                 z1 = cidata[,1], z2=cidata[,2], mean )
> ggplot(df, aes(x=x, y=mean)) +
+   geom_point(size=4, color="#440154") +
+   geom_errorbar(aes(ymax= z2, ymin=z1), color="#3b528b")
```



Pivotal Quantity: Limitations

(a) Let $p > 0$ and $(X_1, X_2, \ldots, X_n)$ be from the Bernoulli(p) distribution. We are interested in a confidence interval for p.

(b) Let $\lambda > 0$ and $(X_1, X_2, \ldots, X_n)$ be from the Poisson(λ) distribution. We are interested in a confidence interval for λ.

In the above cases, there seems to be no obvious $T(X_1, X_2, \ldots, X_n; p)$ or $T(X_1, X_2, \ldots, X_n; \lambda)$ that is independent of parameter.

We need to use "large n" approximation provided by Central limit theorem.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. random variables with finite mean μ, finite variance σ^2. Then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z,$$

where $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ and $Z \sim$ Normal $(0, 1)$.

Confidence Intervals

$X_1, X_2, X_3, \ldots, X_n$ are from a population with mean μ (unknown) and variance σ^2 (known)

Use Central limit theorem (large n)

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \sim \text{Normal}[0,1)$$

Then

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

This implies that

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.