April 10:

Dataset 1:

$\{100, 90, 110\}$

Dataset 2:

$\{50, 100, 150\}$

Dataset 3:

$\{10, 100, 190\}$

For all $\bar{X}$ are same

i) Providing a single point estimate may not be desirable in many application

2) Interval estimation accounts for variability of the estimate

---

Definition: let θ be a parameter of interest. Based on a random sample $X_1, \ldots, X_n$, we obtain two statistics

$L = L(X_1, \ldots, X_n)$ and

$U = U(X_1, \ldots, X_n)$,

such that

$$\mathbb{P}(\theta \in [L, U]) = 1-\alpha,$$

where $\alpha \in (0,1)$.

Then $[L, U]$ is known as $(1-\alpha) \times 100\,\%$ "Confidence interval" for θ and the methodology to obtain the confidence interval is known as "interval estimation".

Remark: ① The interval

Remark. (i) ...

[L, U] is random.

(ii) Usually α is taken as 0.01, 0.05 or 0.1.

Remark: Two ways of statistical inference.

- Frequentist $\downarrow$ parameter is fixed
- Bayesian $\downarrow$ Parameter is random $\downarrow$

✓

out of scope

Interpretation:

$\mathbb{P}(\theta \in \underbrace{[L, U]})$

If L & U are fixed

then $\mathbb{P}(\theta \in [L, U])$

is either 0 or 1.

If we repeat experiment

1000 times then we expect $[L, U]$ contains the true parameter approximately $(1-\alpha) \times 1000$ times.

<u>Pivotal Quantity</u>: It's a function of $X_1, \ldots, X_n$ and the unknown parameter, such that it's distribution is completely known.

For ... X ... X_n is

Eg: Let $X_1, \dots, X_n$ be a random sample (IID) from $N(\mu, \sigma^2)$, where σ^2 is known.

Then, observe that

$$T_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

So, T_n is a pivotal Quanility.

Eg: Let $X_1, \dots, X_n$ be IID RVs from $N(\mu, \sigma^2)$

Where μ & σ^2 both are unknown.

We are interested in CI (confidence interval) of μ.

Here

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$ is not a relevant pivotal quantity.

Here

① $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ .}$$

(ii) $$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 ,$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

σ^{2}

(iii) $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ & $\dfrac{(n-1)s^2}{\sigma^2}$

are independent.

Therefore, using def of t-distribution.

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2}\cdot\dfrac{1}{(n-1)}}} \sim t_{(n-1)}$$

$$\Rightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{(n-1)}$$

completely
known

So, the pivotal quantity is

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S}.$$

<u>CI for normal mean when variance is known</u>

Let $X_1, \ldots, X_n$ are IID observations from $N(\mu, \sigma^2)$ where σ^2 is known.

... ested in CI of

we are interested in ...

μ.

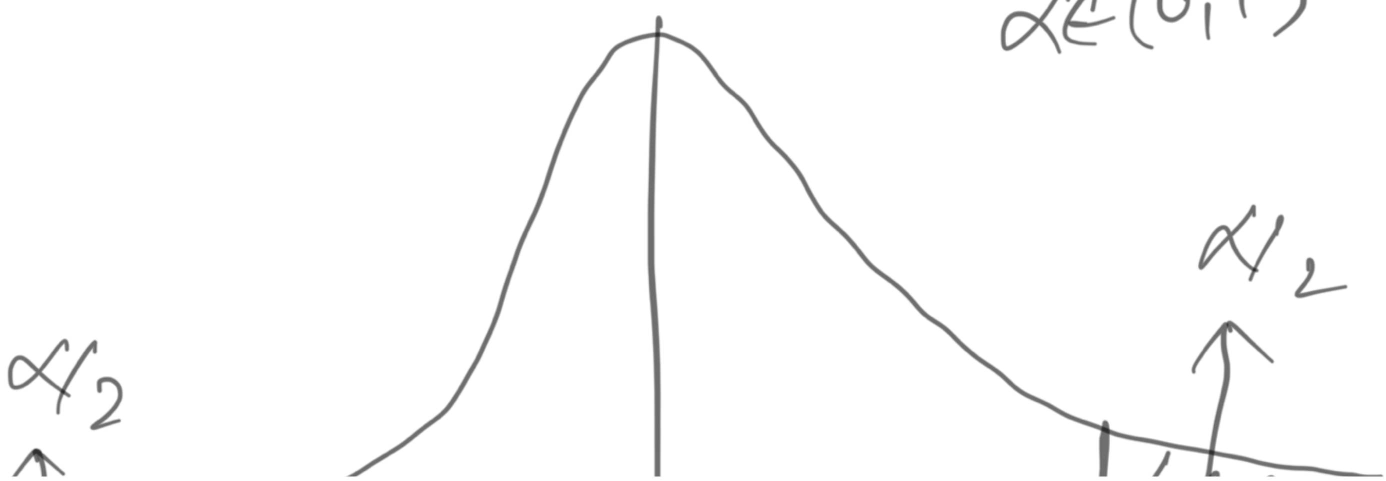
The pivotal quantity is

$$T_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

where $\bar{X}_n = \dfrac{X_1 + \dots + X_n}{n}$.

Using Z-table or comp. soft. or calculator, we can obtain $Z_{\alpha/2}$ such that

$\alpha \in (0,1)$



$$\mathbb{P}\left(|Z| \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(|T_n| \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(\left|\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right| \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right)$$

$$= 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(\bar{X}_n - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

$$= 1 - \alpha$$

$$\Rightarrow \mathbb{P}\left(\mu \in \left[\bar{X}_n - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]\right)$$

$$= 1 - \alpha$$

So, the $(1-\alpha) \times 100\%$ CI_n for μ is

$$\Big[\underbrace{\bar{X}_n - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}_{L}, \underbrace{\bar{X}_n + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}_{U}\Big]$$

In the above example if σ^2 is unknown.

the pivotal quantity is

$$T_n = \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{S} \sim t_{(n-1)}$$

Using prob. table/calculator we can find $t_{\alpha/2}$

such that

$$\mathbb{P}(|Y| < t_{\alpha/2}) = 1-\alpha$$

where $Y \sim t_{(n-1)}$

(t-dist is symmetric)

Therefore,

$$\mathbb{P}\left(\left|\frac{(\bar{X}_n-\mu)}{S/\sqrt{n}}\right| \le t_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(\bar{X}_n - t_{\alpha/2}\cdot\frac{S}{\sqrt{n}} \le \mu \le \bar{X}_n + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

$$= 1-\alpha$$

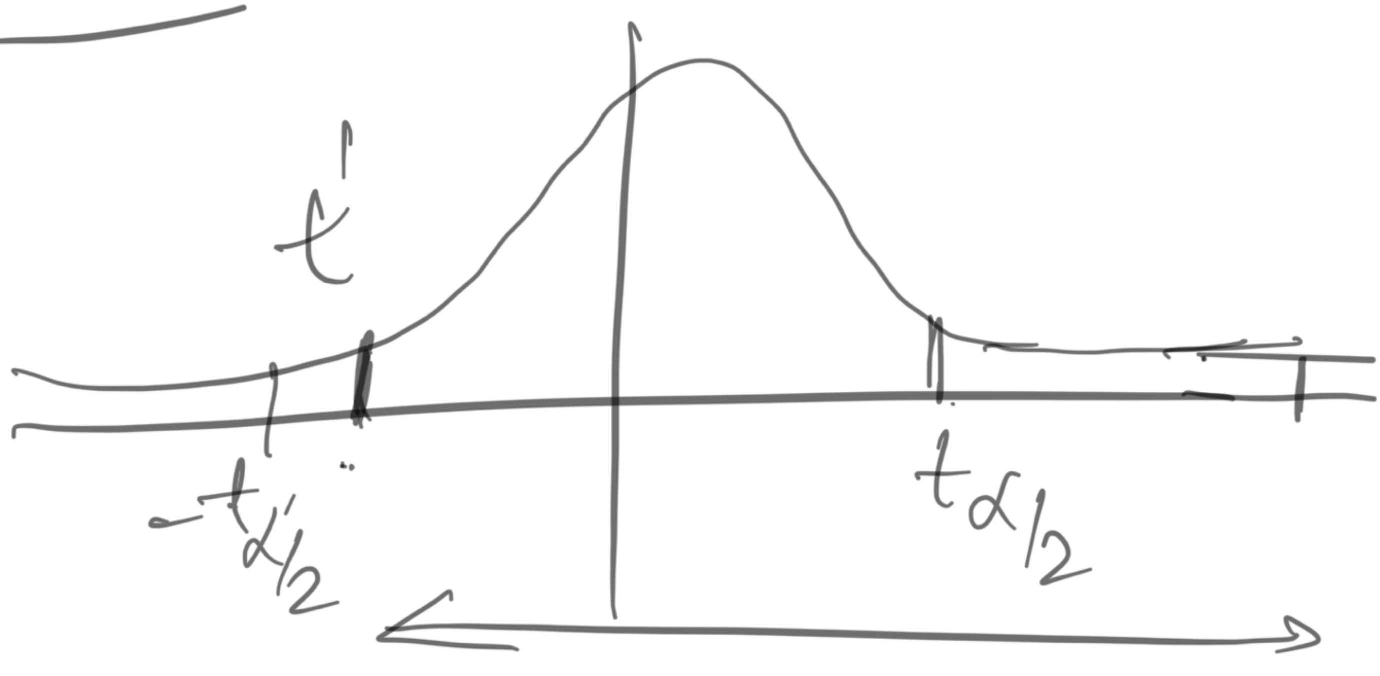
So, the $(1-\alpha)\times 100\,\%$ CI for μ is

$$\left[\bar{X}_n - t_{\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X}_n + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Sometimes it is written as

$$\bar{X}_n \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

Remark:



For normal & t-dist the central confidence interval are shortest.

If only higher values are of interest