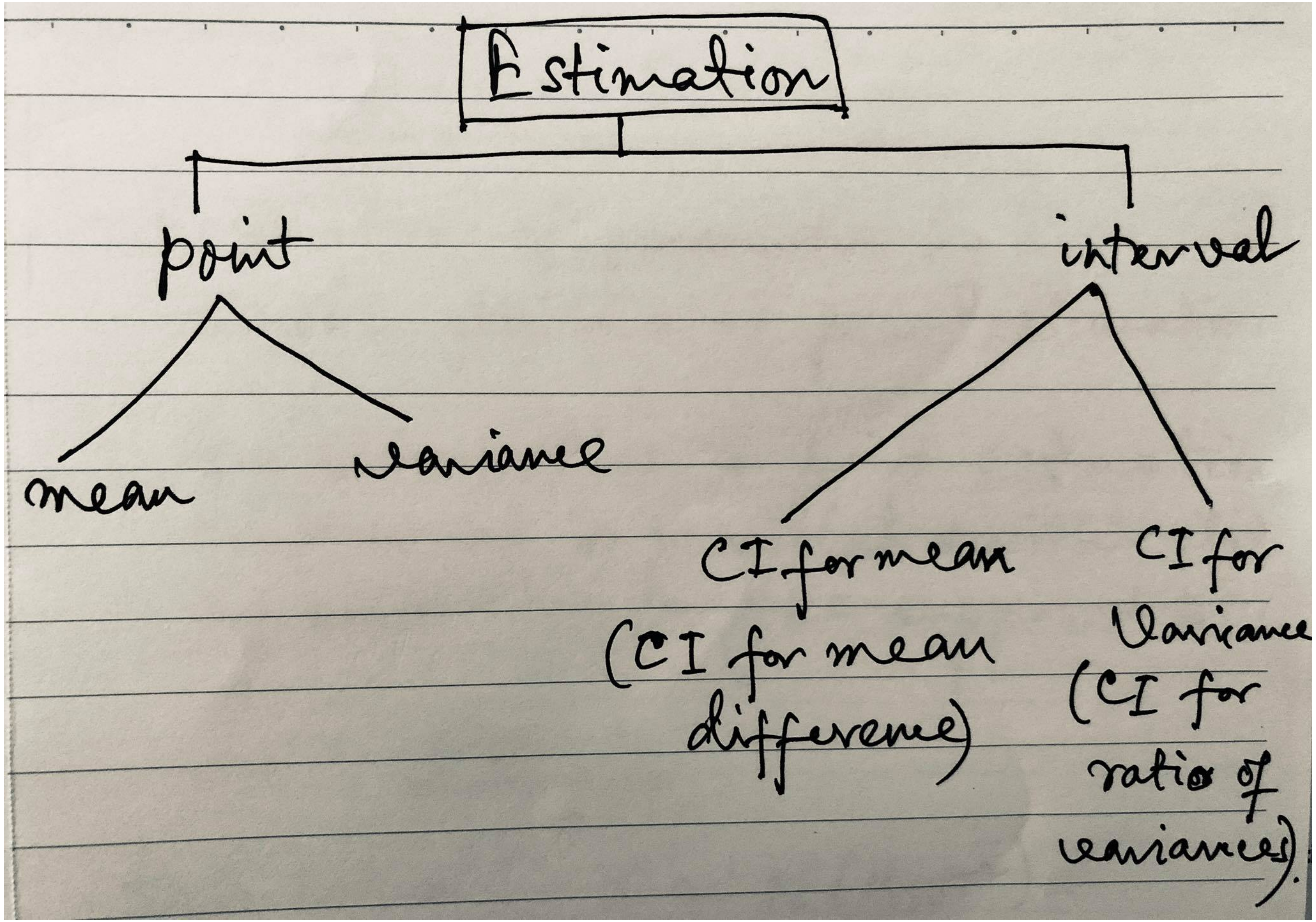
Confidence Interval

Mainak Thakur

14.09.2020

Estimation Types



Ref: https://nptel.ac.in/courses/111/105/111105091/

Interval Estimation

Interval Estimation.

- Estimation of parameter by a single value → point estimation
- An alternative to point estimation, an interval is provided within which the parameter may be supposed to lie.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Example

Example.

Population: $X \sim N(\mu, \sigma^2)$.

μ: unknown , σ^2: known

Sample: $X_1, X_2, ..., X_n$ i.i.d. random samples

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

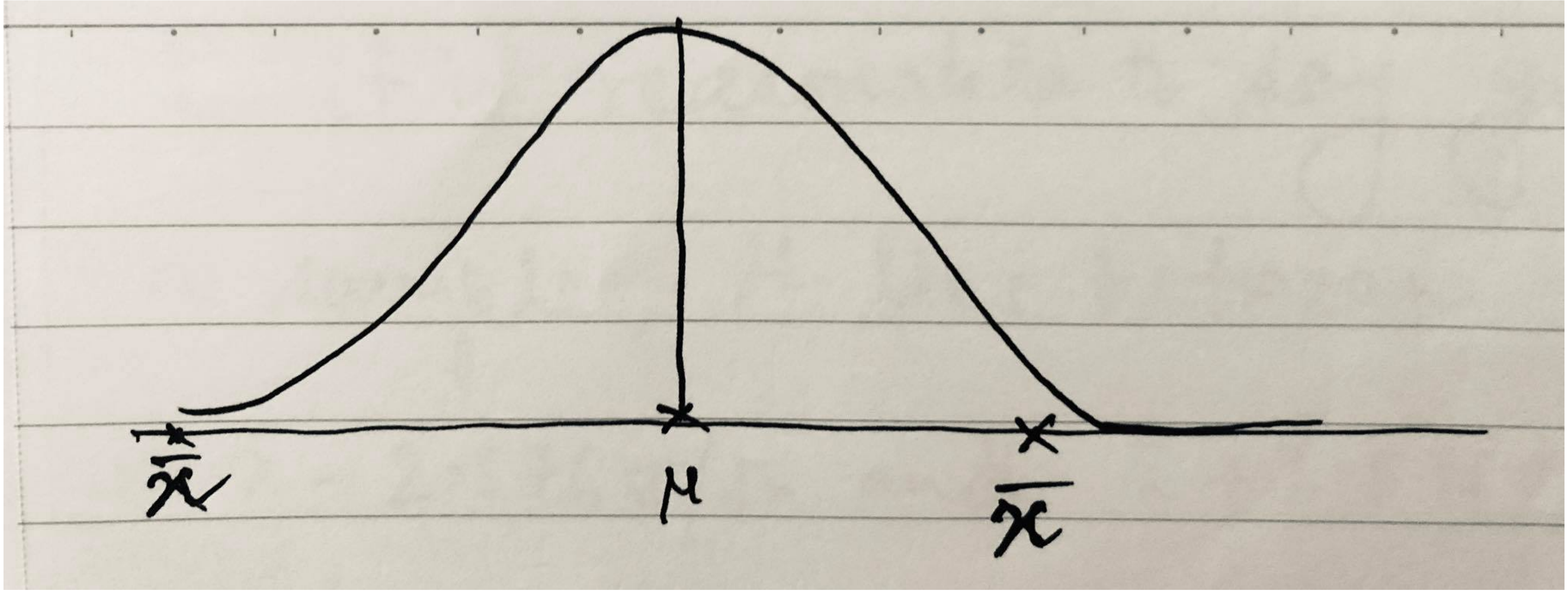
$$P\left[-2.576 \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq 2.576\right] = 0.99.$$

or $$P\left[\bar{X} - 2.576\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.576\frac{\sigma}{\sqrt{n}}\right] = 0.99.$$

1

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Concept



Concept

This means, in repeated sampling it is most likely that $(\bar{x} - 2.576\sigma/\sqrt{n}$, $\bar{x} + 2.576\sigma\sqrt{n})$ will include μ.

Conceptually, if a very large collection of samples, each of size n, are there each gathered from the same popn. For each sample, the interval is calculated as above, 99% of the times the interval includes μ.

Concept

So it is reasonable to say, given a sample, μ lies between $\bar{x} - 2.576\sigma/\sqrt{n}$ and $\bar{x} + 2.576\sigma/\sqrt{n}$, the limits computed based on the sample in hand. These are called 99% confidence limits to μ, 0.99 being the confidence coefficient - a measure of trust or confidence that we may have in the limits for actually including μ.

Concept

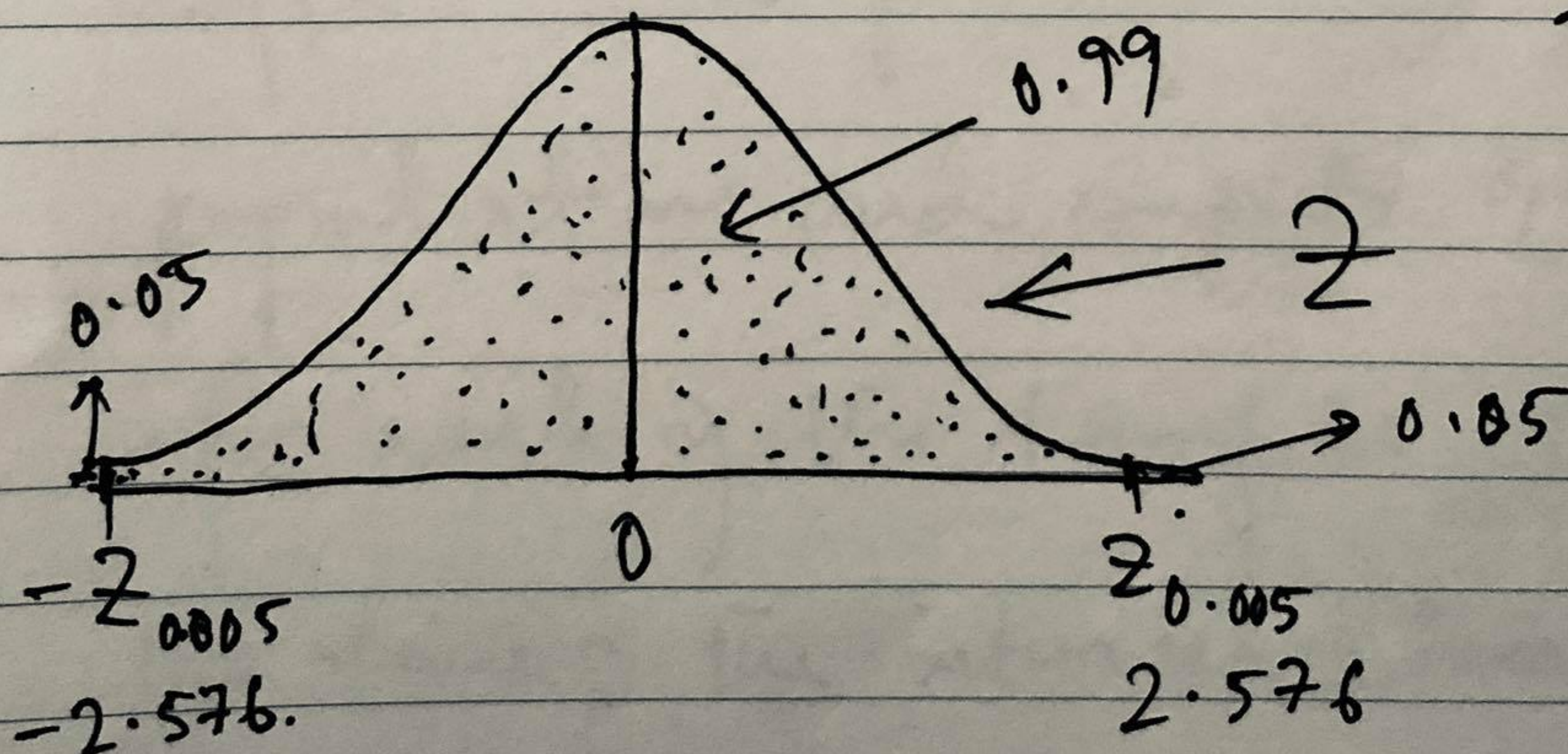
The choice of the confidence coeff depends on the study and the experimenter.

Convention, close to 1, such as

0.99, 0.95, 0.90

Notation, $0.99 = 1 - \alpha \Rightarrow \alpha = 0.01$

$\frac{\alpha}{2} = 0.005$

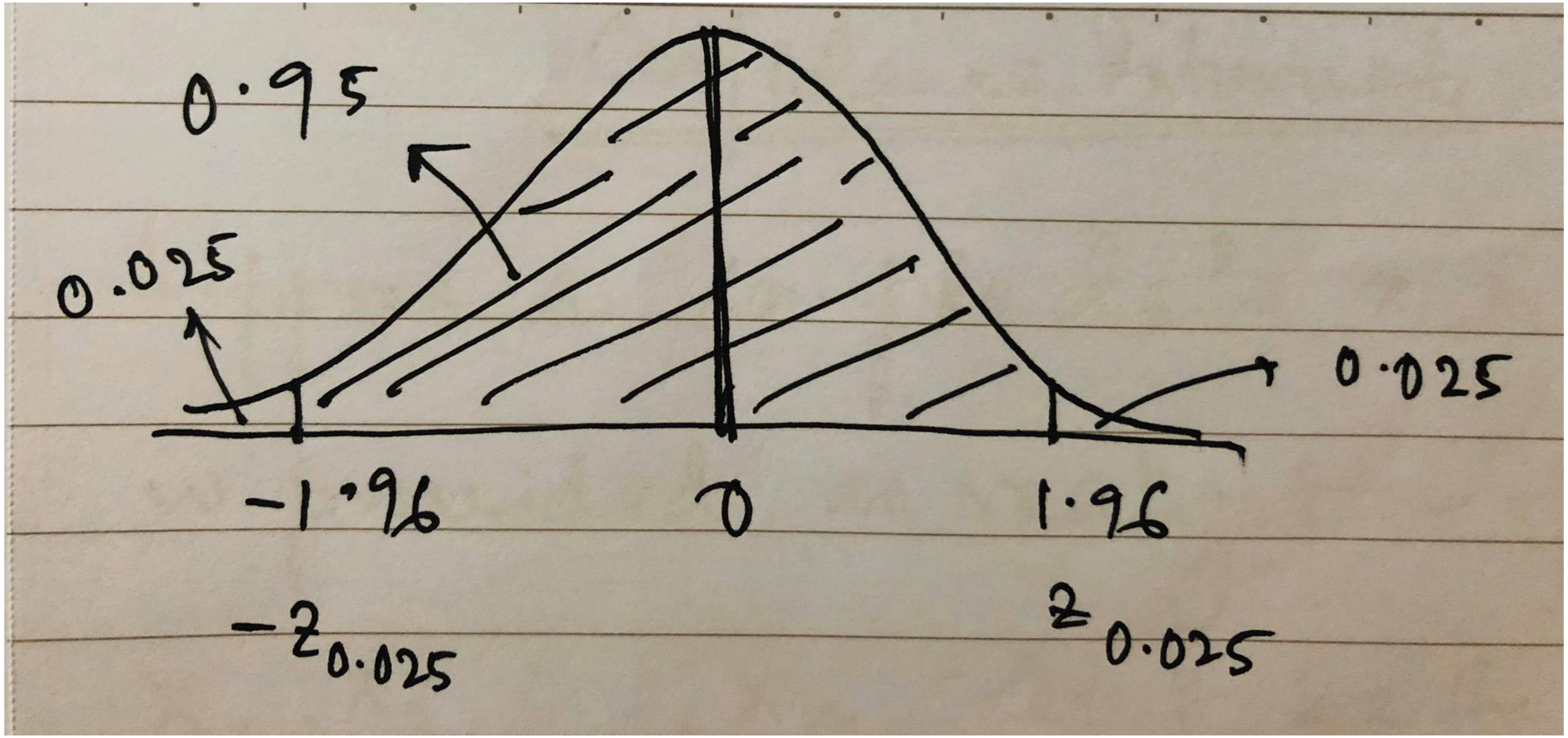


If $1 - \alpha = 0.95$, $\alpha = 0.05$

$\alpha/2 = 0.025$

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Concept



Concept

95% confidence interval implies that if you ~~collect~~ collect 100 different samples from the same population and compute 95% CI for each of the sample, 95 out of 100 times the interval should contain the true μ. In practice, only one sample is selected and CI is calculated based on the given sample.

Ref: Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Confidence Interval

Confidence Interval

Suppose a sample data $x_1, x_2, ..., x_n$ is provided, as realization of r.v.s $X_1, X_2, ..., X_n$. Let θ be the unknown parameter of interest, and $(1-\alpha)$ a no. between 0 and 1. If there exists sample statistics:

Ref: Wikipedia

Confidence Interval

$L_n = g(X_1, X_2, \ldots, X_n)$ and $U_n = h(X_1, X_2, \ldots, X_n)$

s.t. $P(L_n < \theta < U_n) = 1-\alpha$,

Then (l_n, u_n), where $g(x_1, x_2, \ldots, x_n) = l_n$

& $h(x_1, x_2, \ldots, x_n) = u_n$, is called

a $100(1-\alpha)\%$ confidence interval

for θ. $(1-\alpha)$ is called confidence level.

Ref: Wikipedia

CI for Single Population Mean

Confidence interval for mean: Single population (Sampling dist^{n}s)

Population.

- Normal
- Non-normal

Normal: 1. σ - known

- large sample ($n \geq 30$) → Z - dist^n
- small sample ($n < 30$) → Z - dist^n

2. σ - unknown

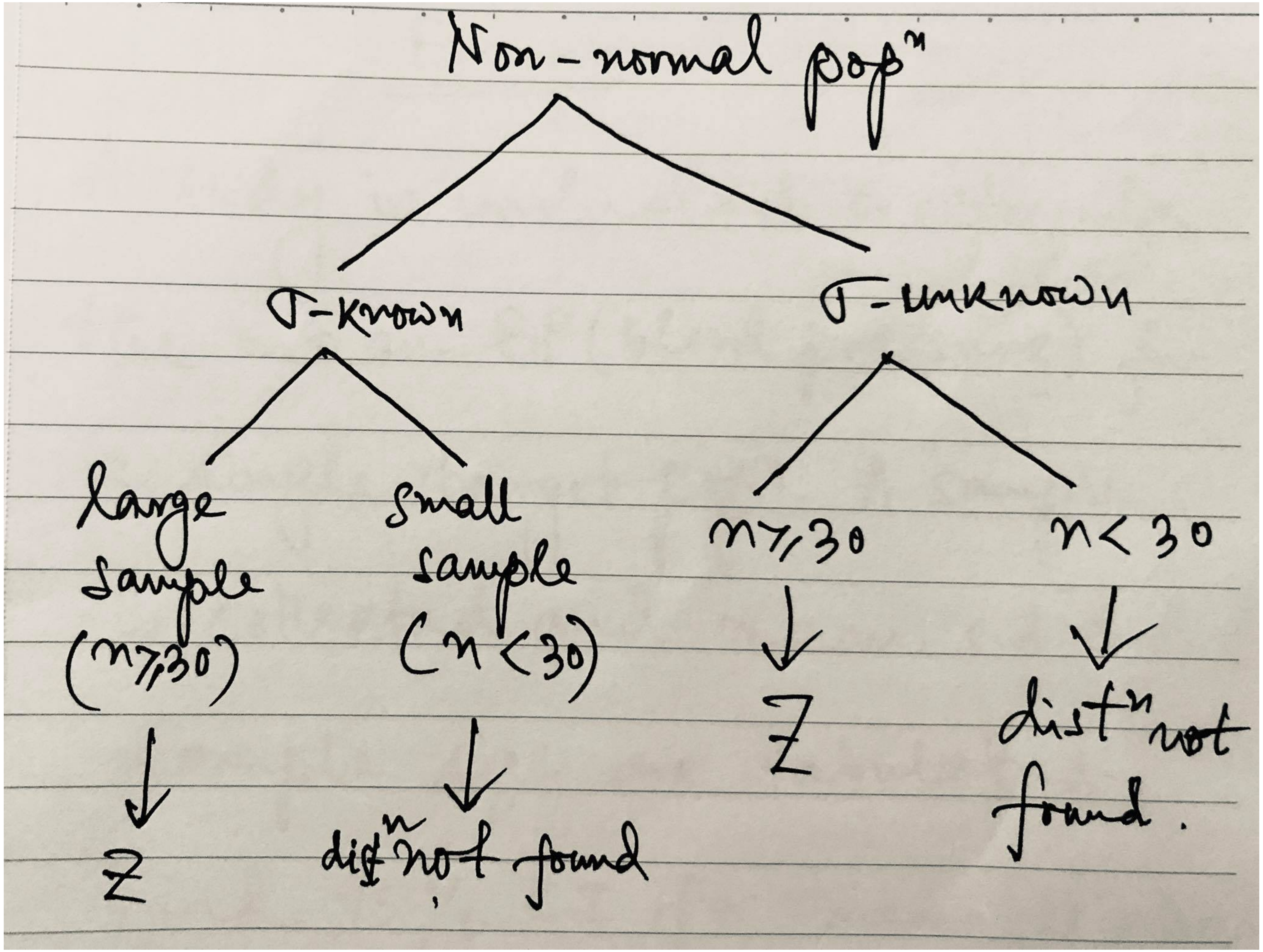
- large sample ($n \geq 30$) → Z
- small sample ($n < 30$) → t

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim Z$$

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Ref: https://nptel.ac.in/courses/111/105/111105091/

CI for Single Population Mean



Ref: https://nptel.ac.in/courses/111/105/111105091/