# Lezione 11 - Interval Estimation

The interval estimate is when we input the observed sample.

Confidence interval for Gaussians with KNOWN $\sigma^2$

$$\left( \bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \, , \, \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \right)$$

The length is: $L = 2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_0^2}{n}}$

For $L \to 0 \Rightarrow n \to \infty$

If $\alpha \to 0 \Rightarrow z_{1-\frac{\alpha}{2}} \to \infty \Rightarrow L \to \infty$

↳ it's not convenient to increase $1-\frac{\alpha}{2}$, since it increases the length, so it is less precise.

## Example

A lead ball falling 100 m.

$X$ = fall time (s)

$n = 4$

$$X_1, X_2, X_3, X_4 \overset{iid}{\sim} \mathcal{N}(\mu, (0{,}8)^2)$$

$x_1 = 4{,}4\,s \quad x_2 = 4{,}6\,s \quad x_3 = 4{,}7\,s \quad x_4 = 4{,}5$

$\hat{\mu} = \bar{x}_4 = 4{,}55$

CI for $\mu$ [two-sided] at level $1-\alpha = 95\%$

$$\left(\bar{x}_4 - z_{1-\frac{0,05}{2}}\sqrt{\frac{(0,8)^2}{4}}, \bar{x}_4 + z_{1-\frac{0,05}{2}}\sqrt{\frac{(0,08)^2}{4}}\right)$$

$$(3.766,\ 5.334) \rightarrow L = 1,568$$

<u>Remark</u>

↳ we can not say $P(3,766 < \mu < 5,334) = 0,95$, since there is no range variable involved, so there is no probability as there is no randomness, these are all observed and real values.

CI for $\mu$ of confidence level $1-\alpha$ is

$$\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_0^2}{n}}\right)$$

K → repeat the "experiment" k times, where k is large e.g. k = 1000, or k = 10000 or more

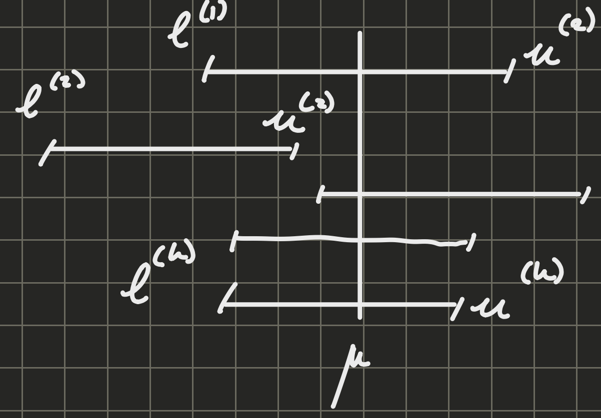
$k=1$ $\quad (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$ → first observed sample

we find $l^{(1)}, u^{(1)}$

$k=2$ $\quad (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)})$ → another real value interval $(l^{(2)}, u^{(2)})$

$k = K = 10000$ → $\qquad$ → we get $(l^{(k)}, u^{(k)})$

Suppose we know the true value of $\mu$:




Confidence level $1-\alpha$, means $(1-\alpha)\cdot 100\%$ of $k$ intervals contain the true value of $\mu$.

<u>Definition:</u>

A one sided lower confidence bound (limit) for $\theta$ of a confidence level $100(1-\alpha)\%$ is determined by the random interval $(L(X_1,\dots,X_n), \infty)$ such that:

$$P_\theta\left(L(X_1,\dots,X_n) < \theta\right) = 1-\alpha$$

There is a parallel for the upper limit.

<u>What changes if the variance is unknown</u>, while we are still trying to estimate $\mu$.

What changes is the distribution of the pivotal quantity, from $N(0,1)$ to the t-student distribution.

Definition:

The random variable $T$ is t-distributed with $k$ degrees of freedom if $T$ is absolutely continuous with density:

$$f_T(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{k}\right)^{\frac{k+1}{2}}} \quad t \in \mathbb{R}$$

Prop 1: If $k \geq 2$, $E(T) = 0$

Prop 2: If $k \geq 3$, $Var(T) = \frac{k}{k-2}$

Prop 3: If $k \to \infty$: $f_T(t) \to \phi(t)\ \forall t$

↳ density of standard normal.

Notation: $T \sim t(k)$, quantile $t_{1-\alpha}(k)$ (or $t_{1-\alpha,k}$)

$$P(T \leq t_{1-\alpha}(k)) = 1-\alpha \to \text{same definition as } z_{1-\alpha}$$

T-student density plot.

$$P(T < t_{1-\alpha}(k)) > P(T < z_{1-\alpha})$$

↳ the tails are heavier, the have more mass.

CI ($\mu$) of a Gaussian population with <u>unknown</u> variance

$X_1, \dots, X_n$ random sample from $N(\mu, \sigma^2)$, $\sigma^2$ unknown.

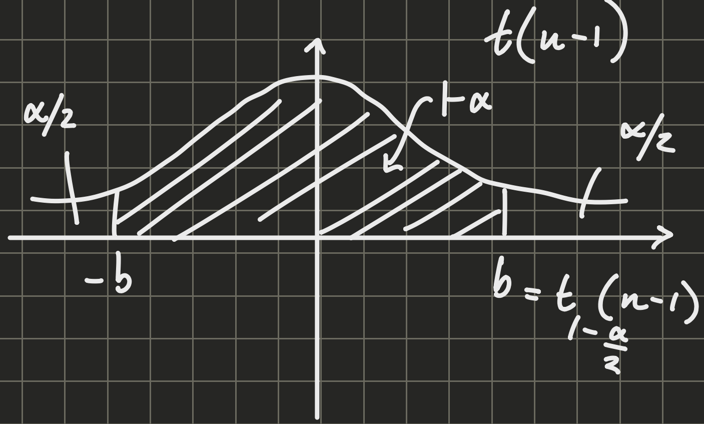
A two sided confidence interval (CI) for $\mu$ of confidence level $1-\alpha$ is:

$$\left(\bar{x}_n - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1),\ \bar{x}_n + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)\right)$$

estimate for $\sigma$, $\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x}_n)^2$

We get this from:

$$P\left(-b < \frac{\bar{X}_n - \mu}{\sqrt{\frac{S^2}{n}}} < b\right) = 1 - \alpha$$




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

Same as before we isolate $\mu$:

$$\left(\bar{X}_n + t_{1-\frac{\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}} < \mu < \bar{X}_n + t_{1-\frac{\alpha}{2}}(n-1)\sqrt{\frac{S^2}{n}}\right)$$

CI for $\mu$ of a population if $n$ is large

In general if the are not gaussian we cannot find the CI for $\mu$, unless $n$ is large ($n \geq 50$). if this is true the distribution of the pivotal statistic can be assumed to be large $N(0,1)$

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \overset{\text{approx}}{\sim} N(0,1)$$

## Example

A company manufactures a new type of solar panels

$n = 128$ tested, only 9 out of 128 are found to be defective.

a) Find a two-sided CI of 95% level for the Probability that a new solar panel is defective.

$n = 128$

$$X_1, \dots, X_{128} \overset{iid}{\sim} Be(p) \qquad X_i = \begin{cases} 1 & \text{if the } i\text{-th solar panel is defective} \\ 0 & \text{otherwise.} \end{cases}$$

$p$ = probability that a new solar panel is defective, $p \in (0,1)$

$p = E[X_1]$

CI for the mean of an UNKNOWN population

Pivot $$\frac{\bar{X}_n - p}{\sqrt{\frac{S^2}{n}}} \underset{\substack{\uparrow \\ n \text{ large}}}{\overset{\text{approx}}{\sim}} N(0,1)$$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}_n - p}{\sqrt{\sim\sim}} < z_{1-\frac{\alpha}{2}}\right) \simeq 1-\alpha$$

How can we approximate $S^2$?

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

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[X_i^2 - 2\bar{X}_n X_i - \bar{X}_n^2\right]$$

$$= \frac{1}{n-1}\left\{\left(\sum_{i=1}^{n} X_i^2\right) - 2n\bar{X}^2 + n(\bar{X})^2\right]$$

$$\boxed{S^2 = \frac{1}{n-1}\sum_i X_i^2 - n\bar{X}^2}$$

In this case $X_i^2 = X_i$ since $X_i$ is either 1 or 0.

$$= \frac{1}{n-1}\left[\sum_i X_i - n(\bar{X})^2\right] =$$

$$= \frac{1}{n-1}\left[n\bar{X} - n(\bar{X})^2\right] = \underbrace{\frac{n}{n-1}}_{\simeq 1}\bar{X}(1-\bar{X}) \simeq \bar{X}(1-\bar{X})$$

$\Rightarrow$ Pivot is $\dfrac{\bar{X} - p}{\sqrt{\dfrac{\bar{X}_n(1-\bar{X}_n)}{n}}} \overset{\text{appox}}{\sim} N(0,1)$

CI of $p$ with approximate confidence level $1-\alpha$ is

$$\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}},\ \bar{x}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\right.$$

$\hat{p} = \bar{x}_n = \frac{9}{128} \rightarrow (0.026, 0.1146) = (2{,}6\%, 11{,}46\%)$

$\hookrightarrow$ in formula only it might be written with $\hat{p}$

$L = 8{,}86\%$
$= 0{,}0886$

$1-\alpha = 95\%$ $\qquad z_{1-\frac{\alpha}{2}} = z_{1-\frac{0{,}05}{2}} = 1{,}96$

$n\hat{p} = n\bar{x}_n = 9 > 5$

$n > 50$

$n(1-\hat{p}) \geq 5$

b) Find the minimum number of solar panels such that the length of CI (with same $1-\alpha$) is not longer than 5%.

$n$

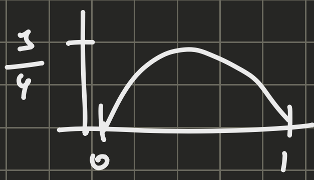
$$L = 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \leq 0{,}05$$

$$L = 2 z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \leq \cancel{2}\, z_{1-\frac{\alpha}{2}} \sqrt{\frac{\cancel{1/4}}{n}} \leq 0{,}05 \Rightarrow \sqrt{n} \geq \frac{1{,}96}{0{,}05} = \sqrt{1536{,}64}$$

Let's create the function

$x \mapsto x(1-x) \leq \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$x \in (0,1)$




$$\boxed{n \geq 1537}$$

since $x$ includes $n$, we found a limit for it so we didn't have to solve weirdly.

Remark

CI for $p$ approximate

If $\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < 0$

$\hat{p} = \bar{x}_n$

we do not write $CI(-0.\ldots, \ldots)$

We have to write $CI\left(0,\ \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

we limit to 0

Similarly:

If $\hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\phantom{x}} > 1$ $\quad CI\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1\right)$

We limit to 1.

## CI for paramalters different from $\mu$, using plug-in estimator

$$X_1, \dots, X_n \overset{iid}{\sim} Exp(\theta)$$

$$E_\theta[X_1] = \frac{1}{\theta} \qquad Var(X_i) = \frac{1}{\theta^2} = E_\theta([X_i])^2$$

$$\bar{X}_n = \widehat{\left(\frac{1}{\theta}\right)}$$

Plugin estimator

$$\widehat{Var}[X_i] = (\bar{X}_n)^2$$

How do we find the confidence interval for

If $n$ is large ($n \geq 50$) a CI for $\frac{1}{\theta}$ $(= E_\theta[X_i])$

$$\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{(\bar{x})^2}{n}} \;,\; \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\bar{x}}{\sqrt{n}}\right)$$

CI for $1/\theta$

$$\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\bar{x}}{\sqrt{n}} < \frac{1}{\theta} < \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\bar{x}}{\sqrt{n}}\right)$$

we can solve these inequalities with respect to $\theta$

$\Rightarrow$ CI for $\theta$ of approximate level $1-\alpha$

$$\left(\frac{1}{\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\bar{x}}{\sqrt{n}}} \;,\; \frac{1}{\bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\bar{x}}{\sqrt{n}}}\right)$$