Point Estimation

$X_1, \ldots, X_n$: iid random variables

$X_i \sim P(\theta)$ θ : unknown parameter

We want to estimate θ

$\Rightarrow$ we consider an estimator $\hat{\theta}_n$

$\left(\hat{\theta}_n = f(X_1, \ldots, X_n) \right)$

We prefer unbiased estimators

$\mathbb{E}[\hat{\theta}_n] = \theta$ for all n

assymptotical unbiased estimators

$$E[\hat{\Theta}_n] \xrightarrow[n\to\infty]{} \Theta$$

- if we have $X_1, \ldots X_n$ iid such that $\Theta = E[X_1]$, $\sigma^2 = V[X_1]$

  . an unbiased estimata for Θ is

$$\hat{\Theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X_n}$$

  . an unbiased estimata for σ^2 is

$$\hat{\sigma}^2_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

A second property for estimators is the convergence or the consistancy.

Let $\hat{\Theta}_n$ an estimata of Θ

we say that $\hat{\Theta}_n$ is consistant (or convergent) if

$$\hat{\Theta}_n \xrightarrow[\underline{P} \text{ or } \underline{a.s}]{} \Theta$$

Theorem:

Let $\hat{\theta}_n$ be an estimate of θ,

If:

- $\hat{\theta}_n$ is unbiased
- $V[\hat{\theta}_n] \xrightarrow[n \to \infty]{} 0$

Then $\hat{\theta}_n$ is consistant

Example:

Let $X_1, \dots, X_n$ iid random variables.

We denote by θ their expectation.

We know that $\hat{\theta}_n = \bar{X}_n$ is an unbiased estimata for θ.

Is $\hat{\theta}_n$ consistant? Yes

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

$$= \frac{1}{n^2} V\left[\sum X_i\right] = \frac{1}{n^2} \overbrace{\sum V[X_i]}^{n\sigma^2}$$

$$V\left[\bar{X}_n\right] = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$

Thanks to the previous theorem, we conclude that $\bar{X}_n$ is consistant

---

Rk: we can also prove that

$$\hat{\sigma}_n^2 \xrightarrow[P]{} \sigma^2$$

$$\hat{\sigma}_n^2 \xrightarrow[a.s]{} \sigma^2$$

Rk: We can get more properties in particular, we have some result about the best unbiased estimator.

If we just consider the family of unbiased estimator of an unknown parameter θ, there exists a 'best estimator' which is the one with the smallest variance.

the smallest variance is known,
this is the Cramer Rao bound.

This Cramer Rao bound is connected with the Fisher Information

$$\text{Cramer-Rao bound} = \frac{1}{\text{Fisher Information}}$$

Construction of estimators

2 main technics:

- method of moments
- maximum likelihood

a) Method of Moments

Remember:

X is a random variable ($k \in \mathbb{N}^*$)

$\mathbb{E}[X^k]$: the moment of order k

$\mathbb{E}[(X - \mathbb{E}[X])^k]$: the centered moment of order k

let consider $X_1, \dots, X_n$ a $\underbrace{\text{n sample}}_{\text{n random variables that are i.i.d}}$

with distribution P_θ.

We want to estimate θ.

<u>Method</u>:

- We find $k \in \mathbb{N}^*$ such that

$$E[X_1^k] = g(\theta)$$

- An estimator of $\hat{\theta}_n$ is such that

$$g(\hat{\theta}_n) = \boxed{\frac{1}{n} \sum_{i=1}^{n} X_i^k}$$

Explanation:

We consider $Y_1 = X_1^k$, $Y_2 = X_2^k$, ..., $Y_n = X_n^k$

$Y_1, \dots, Y_n$ iid r.v such that

$$E[Y_1] = \underbrace{g(\theta)}_{\mu} \text{ : unknown}$$

$\Rightarrow$ an estimate of μ is just

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Rk. because of the Law of Large Number, we know that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^{k} \xrightarrow[as]{} g(\theta)$$

Application :

$X_1, \dots, X_n$ iid with $\mathcal{U}([0,\theta])$

how to estimate θ ?

by method of moments :

$$E[X_1] = \frac{\theta}{2} = g(\theta)$$

with $g : \mathbb{R} \longrightarrow \mathbb{R}$
$\quad x \longrightarrow \frac{x}{2}$

we have $g(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$

it means that

$$\frac{\hat{\theta}_n}{2} = \overline{X}_n \iff \underbrace{\hat{\theta}_n = 2\overline{X}_n}$$

<u>Rk</u>: We have:

$$\mathbb{E}[X_1^2] = V[X_1] + (\mathbb{E}[X_1])^2$$

$$= \frac{\theta^2}{12} + \frac{\theta^2}{4}$$

$$= \frac{4\theta^2}{12} = \frac{\theta^2}{3} = h(\theta)$$

$h: \mathbb{R} \longrightarrow \mathbb{R}$

$\quad x \longrightarrow \frac{x^2}{3}$

we have that an estimator $\hat{\theta}_{n,2}$ is solution of

$$\mu\left(\hat{\theta}_{n,2}\right) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\frac{\hat{\theta}_{n,2}^2}{3} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\Rightarrow \hat{\theta}_{n,2} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$$

(if we assume that $\theta > 0$)

Rk: In the method of moments, we can use the centered moments

The method is:

1) We find k such that

$(k \geq 2)$ $$E\left[\left(X_i - E[X_i]\right)^k\right] = g(\theta)$$

2) An estimator $\hat{\theta}_n$ is such that

$$g(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^k$$

Application

$X_1, \dots, X_n$: n iid r.v with distribution
$$\mathcal{U}([-\theta, \theta])$$

we have $\mathbb{E}[X_1] = 0$

we use $V[X_1] = \frac{1}{12}(\theta - (-\theta))^2$
$$= \frac{\theta^2}{3}$$

$$\Rightarrow \frac{\hat{\theta}_n^2}{3} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

$$\hat{\theta}_n = \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

<u>Rk</u>: With the method of moments, we can construct several estimators for a same parameter

b) Maximum Likelihood

The idea: in the discrete case.

We observed $x_1, \dots, x_n$ that are realizations of $X_1, \dots, X_n$

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} P(X_1 = x_i) = f(\theta)$$

when we know the value of θ,
we look at the point $(x_1, \ldots, x_n)$
with the biggest probabability

$\rightarrow$ because we observe $(x_1, \ldots, x_n)$,
the estimata of θ will be
the value such that the
probability to observe $(x_1, \ldots, x_n)$
is maximal.

definition of the maximum likelihood estimate:

let $X_1, \ldots, X_n$ iid r.v with distribution P_θ

an estimate of θ by maximum likelihood is:

$\hat{\theta}_n$ such that:

$$\mathcal{L}(\hat{\theta}_n) = \max_{\theta} \mathcal{L}(\theta)$$

where $\mathcal{L}$: the likelihood

$$\mathcal{L}(\theta) := \mathcal{L}(x_1, \ldots, x_n ; \theta)$$

$$= \begin{cases} \prod_{i=1}^{n} P_\theta(X_i = x_i) & \text{if } X_i \text{ are discrete} \\ \prod_{i=1}^{n} f_\theta(x_i) & \text{if } X_i \text{ are continuous} \end{cases}$$

–

Application

$X_1, \dots, X_n$: iid r.v with distribution $\mathcal{B}(p)$

we want to estimate p.

$$\mathcal{L}(p) = \prod_{i=1}^{n} \underbrace{P(X_i = x_i)}_{= ?}$$

$$\left.\begin{aligned} P(X_i = 0) &= 1-p \\ P(X_i = 1) &= p \end{aligned}\right\} \quad (1-p)^{(1-x_i)} \, p^{x_i}$$

$$\mathcal{L}(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum x_i} (1-p)^{n - \sum x_i}$$

$$= \left(\frac{p}{1-p} \right)^{\sum x_i} (1-p)^n$$

We want to maximize this fn as a fn of p.

In general, we do not maximize the likelihood but the log likelihood.

Rk: log likelihood $= \ln \mathcal{L}$

let denote by g this likelihood for

$$g(p) = \ln \mathcal{L}(p) = \sum x_i \times \ln\left(\frac{p}{1-p}\right) + n \ln(1-p)$$

$$g(p) = \sum x_i \left(\ln p - \ln(1-p)\right) + n \ln(1-p)$$

we want to maximize g

we compute g'

$$g'(p) = \sum x_i \times \left(\frac{1}{p} + \frac{1}{1-p}\right) - \frac{n}{1-p}$$

because $\ln(1-p) = h(p) = f \circ \ell(p)$

$\ell(p) = 1-p$ and $f(p) = \ln p$

$$h'(p) = f'(\ell(p)) \times \ell'(p)$$

$$g'(p) = \sum x_i \frac{1}{p(1-p)} - \frac{n}{1-p}$$

We need to solve $g'(p) = 0$

$$\sum x_i \times \frac{1}{p(1-p)} - \frac{n}{1-p} = 0$$

$$\frac{\sum x_i}{n} = p$$

We have to compute g''
and see if $g''\left(\frac{\sum x_i}{n}\right) < 0$

$\Rightarrow$ we do not make the computation
but it is the case

$\Rightarrow$ we conclude that
$\hat{p}_n = \frac{\sum X_i}{n}$ is the maximum likelihood $\hat{\text{for}}$ estimate p

- $X_1, \dots, X_n$ iid r.v with distribution

$$\mathcal{E}(\lambda)$$

How to estimate λ ?

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i)$$

$$= \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i} \mathbb{1}_{]0,+\infty[}(x_i) \right)$$

$$= \lambda^n e^{-\lambda \sum x_i} \mathbb{1}_{\min(x_i) > 0}$$

does not depend on λ

we just need to maximize

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$

$$\ell(\lambda) = \ln L(\lambda)$$

$$= n \ln \lambda - \lambda \sum x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i$$

$$\ell'(\lambda) = 0 \iff \lambda = \frac{n}{\sum x_i} = \frac{1}{\overline{x}_n}$$

Since we can see that $\frac{1}{\bar{x}_n}$ is associated to the maximization of the likelihood, we conclude

the maximum likelihood estimator for λ is $\hat{\lambda}_n = \frac{1}{\overline{X_n}}$

Rk: if we apply the method of moments with the moment of order 1: same result

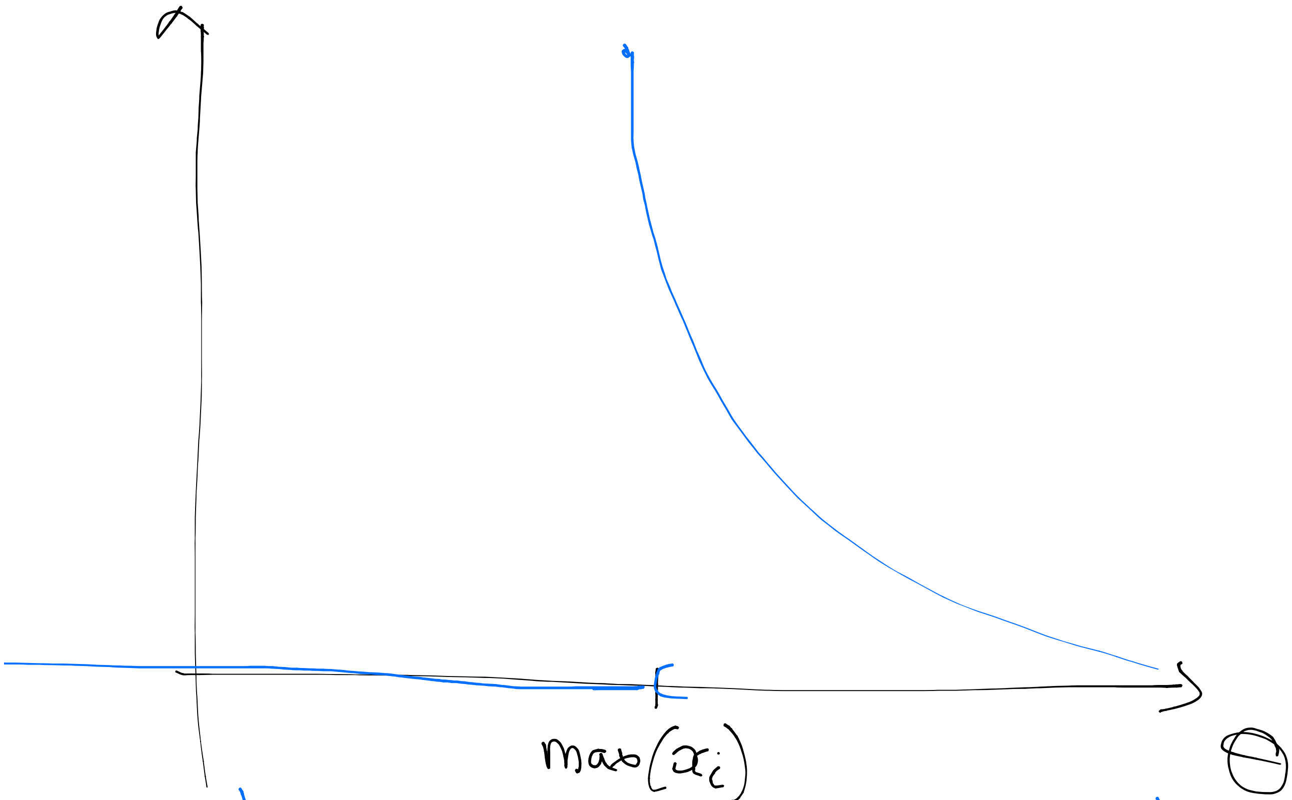
Application:

$X_1, \ldots, X_n$: n iid with $\mathcal{U}([0,\theta])$

estimation of θ ?

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i)$$

$$= \frac{1}{\theta^n} \mathbb{1}_{\min(x_i) \geq 0} \times \underbrace{\mathbb{1}_{\max(x_i) \leq \theta}}_{\text{depends on } \theta}$$



$\Rightarrow$ the maximization is in $\max(x_i)$

We conclude that, in this case the maximum likelihood estimate of Θ is:

$$\hat{\Theta}_n = \max(X_i)$$

Exercise:

Let $X_1, \dots, X_n$ iid r.v with $\mathcal{U}([0,\theta])$

We see that $\hat{\theta}_n = \max(X_i)$

is an estimator of θ

1) Is $\hat{\theta}_n$ an unbiased estimator?

2) If not, deduce an unbiased estimator.

for a given unknown parameter θ,
we can construct several estimators.

$\Rightarrow$ how to know which one is the best?

<u>definition</u>: quadratic error

let $\hat{\theta}_n$ an estimator of θ.

the quadratic error of $\hat{\theta}_n$ is

$$QE(\hat{\theta}_n) = E\left[\left(\hat{\theta}_n - \theta\right)^2\right]$$

property:

$$QE(\hat{\theta}_n) = V[\hat{\theta}_n] + \left(b(\hat{\theta}_n)\right)^2$$

$E[\hat{\theta}_n] - \theta$

Proof:

$$QE(\hat{\Theta}_n) = E\left[(\hat{\Theta}_n - \Theta)^2\right]$$

$$= E\left[\left(\left(\hat{\Theta}_n - E[\hat{\Theta}_n]\right) + \left(E[\hat{\Theta}_n] - \Theta\right)\right)^2\right]$$

$$= \underbrace{E\left[\left(\hat{\Theta}_n - E[\hat{\Theta}_n]\right)^2\right]}_{= V[\hat{\Theta}_n]}$$

$$+ 2E\left[\left(\hat{\Theta}_n - E[\hat{\Theta}_n]\right)\underbrace{\left(E[\hat{\Theta}_n] - \Theta\right)}_{b(\hat{\Theta}_n) =\ \text{constant}}\right]$$

$$+ E\left[\left(E[\hat{\Theta}_n] - \Theta\right)^2\right]$$

$$QE(\hat{\theta}_n) = V[\hat{\theta}_n] + 2b(\hat{\theta}_n) \times E[(\hat{\theta}_n - E[\hat{\theta}_n])] + (b(\hat{\theta}_n))^2$$

$$= E[\hat{\theta}_n] - E[E[\hat{\theta}_n]]$$

$$= E[\hat{\theta}_n] - E[\hat{\theta}_n]$$

$$= 0$$

$$QE(\hat{\theta}_n) = V[\hat{\theta}_n] + (b(\hat{\theta}_n))^2$$

<u>Rk</u>: If $\hat{\Theta}_n$ is an unbiased estimator of Θ

$\Rightarrow b(\hat{\Theta}_n) = 0$

$$QE(\hat{\Theta}_n) = V[\hat{\Theta}_n]$$

$\Rightarrow$ because of the definition of the quadratic error, we prefer unbiased estimator with a small variance

Comparison between 2 estimators

Let $\hat{\Theta}_{n,1}$ and $\hat{\Theta}_{n,2}$ two estimators of Θ.

If $\forall n$, $QE(\hat{\Theta}_{n,1}) \leq QE(\hat{\Theta}_{n,2})$

then $\hat{\Theta}_{n,1}$ is better than $\hat{\Theta}_{n,2}$.

Rq: If we have two estimators for θ, a way to get a better estimator consists in considering an agregated estimator

It means: $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$ 2 estimators of θ

$$\hat{\theta}_{n,3} = a\,\hat{\theta}_{n,1} + b\,\hat{\theta}_{n,2}$$

we find $\tilde{a}$ and $\tilde{b}$ such that

$$QE\left(\tilde{a}\hat{\theta}_{n,1}+\tilde{b}\hat{\theta}_{n,2}\right) = \min_{a,b} QE\left(a\hat{\theta}_{n,1}+b\hat{\theta}_{n,2}\right)$$

$\Rightarrow$ the best linear estimata construct with $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$ is $\tilde{a}\hat{\theta}_{n,1}+\tilde{b}\hat{\theta}_{n,2}$

Exercice:

$X_1, \dots, X_n$: n iid r.v with $\mathcal{U}([0,\theta])$

1) $\hat{\theta}_{n,1} = 2\overline{X}_n$

Compute the quadratic error of $\hat{\theta}_{n,1}$.

2) $\hat{\theta}_{n,2}$ the unbiased version of $\max(X_i)$

Compute the quadratic error of $\hat{\theta}_{n,2}$

3) Which one is the better one between $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$?

Main drawback with point estimation :

⇒ fluctuation due to the sample

Illustration of this phenomenon

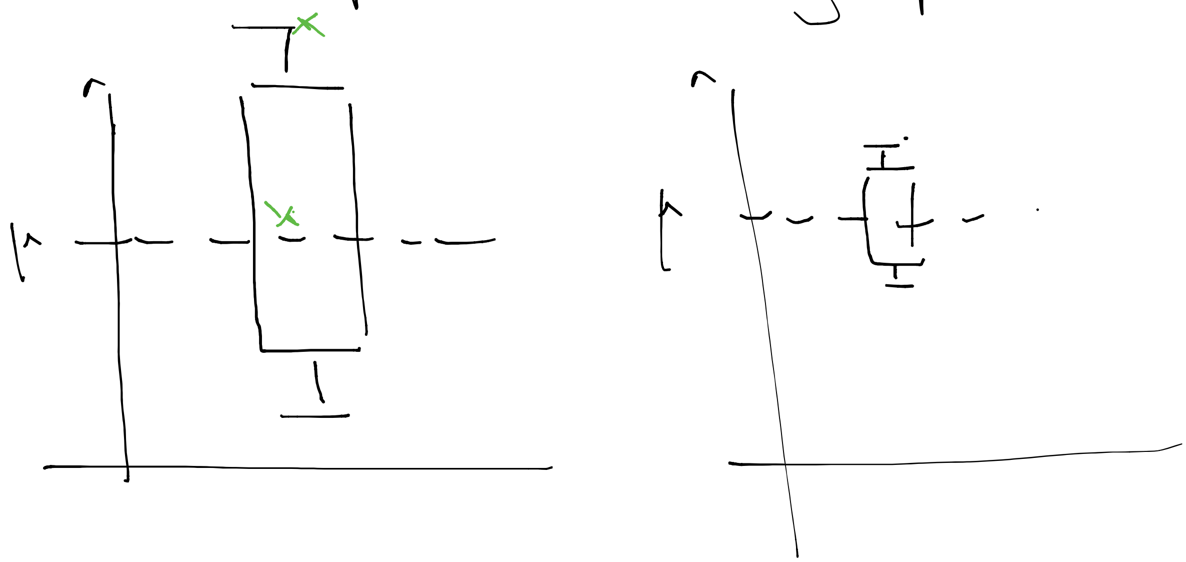
$X_1, \dots, X_n$: iid r.v with distribution
$\mathcal{N}(\mu, \sigma^2)$

we want to estimate μ

$\Rightarrow \hat{\mu}_n = \overline{X}_n$

$\Rightarrow$ we perform simulations

we create 50 datasets associated to $(X_1, \dots, X_n)$ and on each of them, we compute a value of $\hat{\mu}_n$



$n = 10$ $\qquad$ $n = 1000$

boxplot of the 50 values of $\hat{\mu}_n$ in both cases!

To have a quantification of the quality of the estimation, we prefer 'confidence interval'

What is a <u>confidence interval</u> ?

it is random object

to a confidence interval is associated a confidence level, denoted $100 \times (1-\alpha)\%$.

definition:

Let $X_1, \dots, X_n$ iid r.v with distribution P_Θ.

Let $\hat{\Theta}_n$ an estimata of Θ

A confidence interval, based on $\hat{\Theta}_n$, with confidence level $100 \times (1-\alpha)\%$, is a random interval $[A_n, B_n]$ such that:

- $A_n = f_1(\hat{\Theta}_n)$ and $B_n = f_2(\hat{\Theta}_n)$
- $P(\Theta \in [A_n, B_n]) = 1-\alpha$

RK: more and more α is small,
more and more the confidence
level is close to 1,
more and more the length of
the confidence interval is
big!

How to construct such confidence interval ?

a) $X_1, \ldots, X_n$ iid r.v with a distribution $\mathcal{N}(\mu, \sigma^2)$

μ is unknow and σ^2 is known

- we need at first a point estimator for μ.

$\Rightarrow \hat{\mu}_n = \overline{X}_n$ (because μ is the expectation)

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

we want to find A_n and B_n, random variables that are functions of $\overline{X}_n$ such that

$$P\left(A_n \leq \mu \leq B_n\right) = 1-\alpha$$

we make the choice: $A_n = \overline{X}_n - S_n$

$$B_n = \overline{X}_n + T_n$$

we need to find S_n and T_n such that:

$$P\left(\bar{X}_n - S_n \le \mu \le \bar{X}_n + T_n\right) = 1-\alpha$$

$$\iff P\left(-T_n \le \bar{X}_n - \mu \le S_n\right) = 1-\alpha$$

$\sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$ known

$$\iff P\left(-\frac{\sqrt{n}\,T_n}{\sigma} \le \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \le \frac{\sqrt{n}\,S_n}{\sigma}\right) = 1-\alpha$$

$\sim \mathcal{N}(0,1)$

Several solutions because
2 quantities to determine with
just one equation

$\Rightarrow$ we have to make a choice

We can write

$$Y_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}$$

$$\overbrace{P\left(Y_n < -\frac{\sqrt{n}\, T_n}{\sigma}\right)}^{\alpha_1} + \overbrace{P\left(Y_n > \frac{\sqrt{n}\, S_n}{\sigma}\right)}^{\alpha_2} = \alpha$$

3 main choices:

- $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$

$$\Leftrightarrow P\left(Y \leq \frac{\sqrt{n}\, S_n}{\sigma}\right) = 1 - \frac{\alpha}{2}$$

we have:

$$P\left(Y_n < -\frac{\sqrt{n}\,\overline{l_n}}{\sigma}\right) = P\left(Y_n > \frac{\sqrt{n}\, S_n}{\sigma}\right) = \frac{\alpha}{2}$$

one solution is to take

$$\frac{\sqrt{n}\,\sqrt{n}}{\sigma} = \frac{\sqrt{n}}{\sigma} S_n = \text{quantile of order } 1-\alpha/2 \text{ for } \mathcal{N}(0,1)$$

we denote this quantile by

$z_{1-\alpha/2}$

$\Rightarrow \delta_n = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$

$\Rightarrow$ confidence interval for μ at level $100 \times (1-\alpha)\%$ is

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \, , \, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right]$$

- $\alpha_1 = 0$ and $\alpha_2 = \alpha$

$$P\left(Y_n < -\frac{\sqrt{n}\;T_n}{\sigma}\right) = 0 \Leftarrow T_n = +\infty$$

$$P\left(Y_n > \frac{\sqrt{n}\,S_n}{\sigma}\right) = \alpha$$

$$\Leftrightarrow P\left(Y_n < \frac{\sqrt{n}\,S_n}{\sigma}\right) = 1-\alpha$$

$$\Rightarrow \frac{\sqrt{n}\,S_n}{\sigma} = z_{1-\alpha}$$

$$S_n = \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

In this case the confidence interval becomes:

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \, ; +\infty \right[$$

- $\alpha_2 = 0$ and $\alpha_1 = \alpha$

↳ the confidence interval becomes:

$$\left] -\infty \, ; \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right]$$

More generally, you take

$\alpha_1 \geq 0$, $\alpha_2 \geq 0$ such that $\alpha_1 + \alpha_2 = \alpha$

$$P\left(Y_n \leq -\frac{\sqrt{n}\, T_n}{\sigma}\right) = \alpha_1$$

$$\hookrightarrow \frac{\sqrt{n}\, T_n}{\sigma} = z_{1-\alpha_1}$$

$$P\left(Y_n \geq \frac{\sqrt{n}\, S_n}{\sigma}\right) = \alpha_2$$

$$\hookrightarrow \frac{\sqrt{n}}{\sigma} S_n = z_{1-\alpha_2}$$

The associated confidence interval is

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha_2}, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha_1}\right]$$

the length of the confidence interval is $\frac{\sigma}{\sqrt{n}}\left(z_{1-\alpha_2} + z_{1-\alpha_1}\right)$