Confidence interval:

def: I is a confidence interval of θ with level $100 \times (1-\alpha)\%$ if

$$P(\theta \in I) = 1-\alpha$$

Rk: I is a random interval

Construction on some examples

- $X_1, \dots, X_n$ be n r.v with $\mathcal{N}(\mu, \sigma^2)$
  we assume that σ^2 is known!

$\rightarrow$ we want a confidence interval for μ.

an unbaised estimator of μ is $\overline{X}_n$

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

$$\underbrace{\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}}_{T_n =} \sim \mathcal{N}(0,1)$$

aim: to find $a(X_1,\dots,X_n)$, $b(X_1,\dots,X_n)$ such that

$$P\left(a(X_1,\dots,X_n) \leq \mu \leq b(X_1,\dots,X_n)\right) = 1-\alpha$$

If we find t_1 and t_2 such that

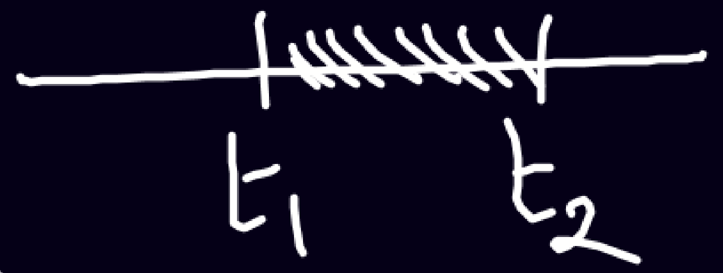
$$P\left(t_1 \leq T_n \leq t_2\right) = 1-\alpha$$

$$\Leftrightarrow P\left(t_1 \leq \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \leq t_2\right) = 1-\alpha$$

$$\Leftrightarrow P\left(\underbrace{\overline{X}_n - \frac{\sigma}{\sqrt{n}} t_2}_{a(X_1,\dots,X_n)} \leq \mu \leq \underbrace{\overline{X}_n - \frac{\sigma}{\sqrt{n}} t_1}_{b(X_1,\dots,X_n)}\right) = 1-\alpha$$

How to define t_1 and t_2

$$P(t_1 \le T_n \le t_2) = 1 - \alpha \quad \text{with } T_n \sim \mathcal{N}(0;1)$$



$$\Longleftrightarrow P(T_n \le t_1) + P(T_n \ge t_2) = \alpha$$

$$\Longleftarrow P(T_n \le t_1) = \alpha_1 \quad P(T_n \ge t_2) = \alpha_2$$

with $\alpha_1 + \alpha_2 = \alpha$

$\alpha_1 \ge 0, \ \alpha_2 \ge 0$

one choice to get some bilateral confidence interval is to take

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2}$$

with this choice, $t_1 = -t_2$

$\Rightarrow$ to find t_2, we say that $P(\mathcal{N}(0,1) > t_2)$

$= 1 - P(\mathcal{N}(0,1) \leq t_2)$

$$\Rightarrow P(\mathcal{N}(0,1) \leq t_2) = 1 - \frac{\alpha}{2}$$

$\Rightarrow$ t_2 : quantile of order $1-\frac{\alpha}{2}$ associated to a $\mathcal{N}(0;1)$

Confidence interval :

$$\left[\overline{X_n} - \frac{\sigma}{\sqrt{n}} t_2 \; ; \; \overline{X_n} + \frac{\sigma}{\sqrt{n}} t_2\right]$$

Rk:

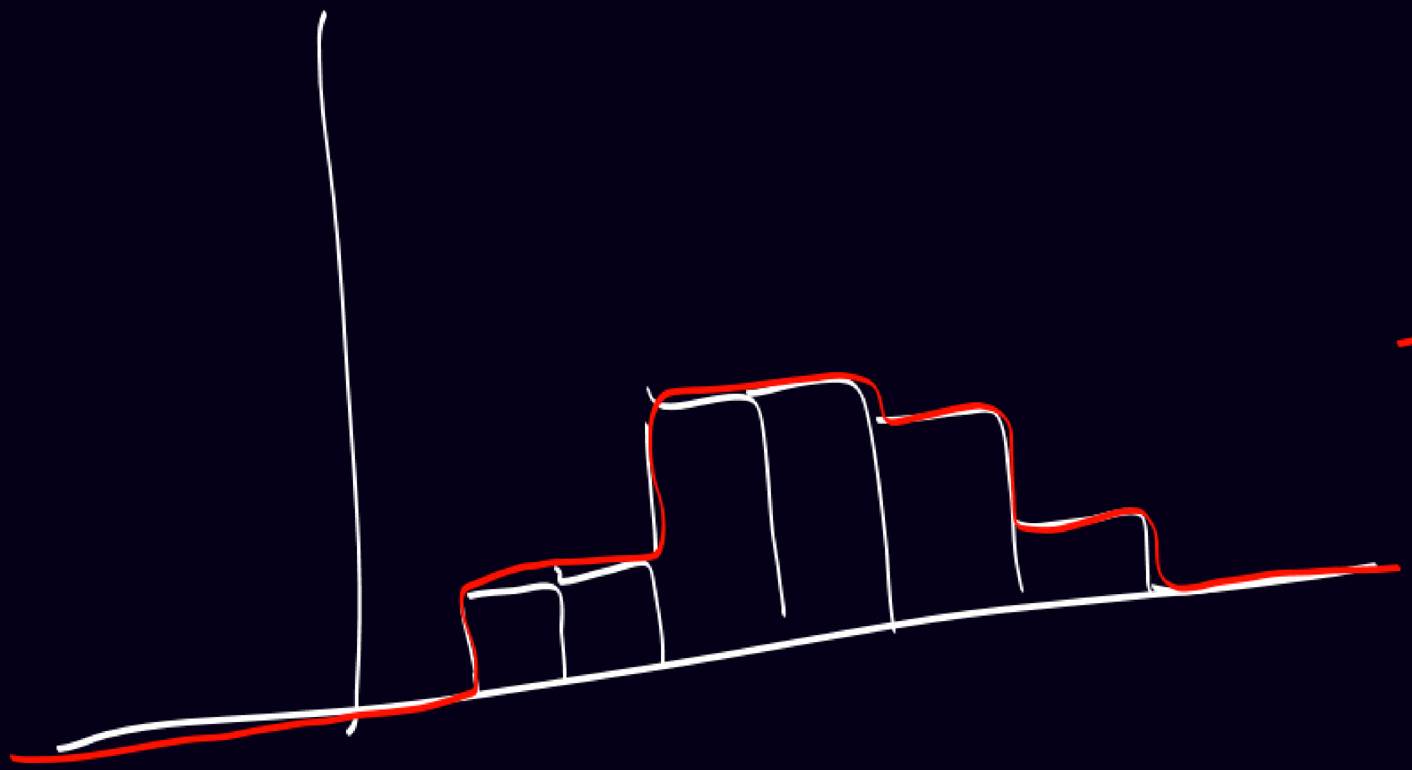
we have observations

$x_1, \dots, x_n$

↳ plot an histogram

↳ hypothesis on the distribution of the data

→ you may recognize a classical distribution (uniform, exponential, gaussian) or



→ f which is a density

Imagine that we assume a uniform distribution on $[0,\theta]$

→ by the theory, we know that

- $\max(X_i)$ is an estimator of θ
- $2 \times \overline{X_n}$ is another estimator

$x_1, \dots, x_n$

↳ an estimation for θ is given by:

- $\max(x_i)$
- $2 \times \bar{x}_n$

How to estimate μ when σ is unknown? (Still when $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$)

Pb: the previous confidence interval is not correct because it depends on an unknown parameter (σ).

$\overline{X}_n$: an estimator of μ

$$\sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

We want to replace σ by $\hat{\sigma}$ with

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

If $X_1, \dots, X_n$ are $\mathcal{N}(\mu, \sigma^2)$,

then we know the true distribution of $\hat{\sigma}^2$.

<u>def</u>: The definition of a Chi-Square r.v is the following.

$Y \sim \chi^2(k)$ $(k \in \mathbb{N}^*)$ if $Y = \sum_{i=1}^{k} Z_i^2$ with $Z_i \sim \mathcal{N}(0,1)$

→ independancy

properties:

Let $Y \sim \chi^2(k)$

- $E[Y] = k$
- $V[Y] = 2k$

<u>def:</u> Student r.v.

let U be a $\mathcal{N}(0,1)$ r.v.

let Z be a $\chi^2(k)$ r.v

with $U \perp\!\!\!\perp Z$.

Define $T = \dfrac{U}{\sqrt{Z/k}}$ $\Rightarrow$ T is a student r.v with parameter k.

Rk:

If $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2(n-1)$$

$$\sum_{i=1}^{n} (X_i - \bar{X}_n) = 0$$

If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

with μ known

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

$$\Rightarrow \frac{n \hat{\sigma}^2}{\sigma^2} \sim \chi^2(n)$$

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

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

We can prove that $\bar{X}_n$ and $\hat{\sigma}^2$ are independant r.v.

We define T_n by :

$$T_n = \frac{\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}}{\sqrt{\left(\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right) / (n-1)}} \sim \text{Student}(n-1)$$

$$T_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \times \frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}$$

$$T_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \times \frac{\sigma}{\hat{\sigma}} = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\hat{\sigma}}$$

⚠ $\sqrt{n}\,\dfrac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0;1)$

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\hat{\sigma}} \sim \text{Student}(n-1)$$

$$\Rightarrow E\left[\overline{X}_n\right] = \mu \quad \text{and} \quad \text{not} \quad \cancel{\overline{X}_n = \mu}$$

Confidence interval:

$$\left[\overline{X}_n - \frac{\hat{\sigma}}{\sqrt{n}} t_2 \, ; \, \overline{X}_n + \frac{\hat{\sigma}}{\sqrt{n}} t_2 \right]$$

t_2: quantile of order $1 - \frac{\alpha}{2}$ associated to a Student$(n-1)$

Rk:

If $X_1, \dots, X_n$ are iid r.v

and if we want to estimate

$\mathbb{E}[X_1] = \mu$

- if $V[X_1]$ is known! (σ^2)

$\to \overline{X}_n$ is an estimator of μ

By using the TCL, we say that we approximate the true distribution of $\overline{X}_n$ by $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

We do exactly the same by we replace the true distribution by $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

$$\Rightarrow \left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} t_2 \, ; \, \overline{X}_n + \frac{\sigma}{\sqrt{n}} t_2 \right]$$

with t_2 : quantile of order $1 - \frac{\alpha}{2}$ associated to a $\mathcal{N}(0;1)$

This interval is just an asymptotic confidence interval!

- If σ^2 is unknown.

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

Slutsky

is an unbiased estimator of σ^2 and consistent

In this case, an asymptotical confidence interval for μ is $\left[\overline{X}_n - \frac{\hat{\sigma}}{\sqrt{n}} t_2 ; \overline{X}_n + \frac{\hat{\sigma}}{\sqrt{n}} t_2\right]$

with t_2 the quantile of order $1-\frac{\alpha}{2}$

associated to a $\mathcal{N}(0;1)$

def: Fisher r.v

let $U \sim \chi^2(k)$
let $Z \sim \chi^2(p)$ } $U \perp\!\!\!\perp Z$

Define F by:

$$F = \frac{U/k}{Z/p} \sim \text{Fisher}(k, p)$$

- Let $X_1, \dots, X_n$ be $\mathcal{N}(\mu, \sigma^2)$

  1) Construct a confidence interval for σ^2 when μ is known

  2) Do the same when μ is unknown.

- let $X_1, \dots, X_n$ be $\mathcal{B}(p)$

  construct an confidence interval for p.

- density of $\max_i(X_i)$ $X_i \sim \mathcal{U}([0,\theta])$

iid

$$Y = \max_i(X_i)$$

$\forall t \in \mathbb{R}$, $P(Y \leq t) = P(\max(X_i) \leq t)$

$= P(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t)$

independance

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

$$P(Y < t) = \left(P(X_1 < t)\right)^n$$ because identically distributed

$$= \begin{cases} 0 & \text{if } t < 0 \\ \left(\dfrac{t}{\theta}\right)^n & \text{if } t \in [0, \theta] \\ 1 & \text{if } t > \theta \end{cases}$$

$$f_Y(t) = \begin{cases} \dfrac{n}{\theta^n} t^{n-1} & \text{if } t \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

$$P\left(Y-a \leq \theta \leq Y+b\right)$$

$$= P\left(Y \in [\theta-b, \theta+a]\right)$$

Rk: $Y = \min(X_i)$

$\forall t \in \mathbb{R}$, $P(Y \geq t) = P(X_1 \geq t, \dots, X_n \geq t)$

$$= \prod_{i=1}^{n} P(X_i \geq t) = \left(P(X_1 \geq t)\right)^n$$

$$F_Y(t) = P(Y \le t) = 1 - P(Y > t)$$

$$= 1 - \left(1 - P(X_1 \le t)\right)^n$$