Statistical Inference Course Notes

Yehor KOROTENKO

31 March 2026

Table of contents

Introduction

§1

1.1 Evaluation

- 0.4 Continuous Assessment +0.6 Exam.
- Breakdown : 80% midterm, 20% Quiz (scheduled for 26/01).

1.2 Statistical Model

Definition 1.1 (Statistical Model) – *A statistical model is a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *where* $\mathcal{P}$ *is a family of probability distributions* $\{P_\theta; \theta \in \Theta\}$.

- If $\exists p \in \mathbb{N}^*, \Theta \subset \mathbb{R}^p$: parametric model.
- Otherwise : non-parametric model.

Example 1.2 (Families of Distributions) –
- *Poisson distributions :* $\mathcal{P} = \{P(\lambda); \lambda > 0\}$.
- *Regular density :* $\mathcal{P} = \{\mathbb{P}; \mathbb{P} \text{ whose density admits a bounded second derivative}\}$.

◇

Definition 1.3 (Observation) – *An observation is a random variable (r.v.) whose distribution belongs to* $\{P_\theta, \theta \in \Theta\}$. *Our observation will have a structure of n-samples* $X_1, ..., X_n$ *i.i.d. (independent and identically distributed) with a common distribution* $\in \{P_\theta, \theta \in \Theta\}$.

Remark 1.4 – $(X_1, ..., X_n)$ *has distribution* $P_\theta^{\otimes n}$. *The sample contains all information about* P_θ*, thus about* θ. ◇

Definition 1.5 (Identifiability) – *A model is identifiable if and only if (iff) the mapping* $\theta \mapsto P_\theta$ *is injective.*

1.3 Estimators

Hypothesis : We observe $X_1, ..., X_n$ i.i.d. from a common distribution $\in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ (identifiable parametric model). Let θ^* be the true unknown value such that $P_{X_i} = P_{\theta^*}$.

Definition 1.6 (Estimator) – *An estimator of* θ *is a measurable function of the sample* $(X_1, ..., X_n)$ *and independent of* θ *(computable from the data).*

Notation : $\hat{\theta} = \hat{\theta}_n = h(X_1, ..., X_n)$. It is a random variable.
Examples : $\hat{\theta} = \overline{X}$, $\hat{\theta} = X_1 - X_3$, etc.

Fundamental Questions :

1. How to define a good estimator ?
2. How to construct a good estimator ?

1.4 Quadratic Risk

Idea : On average, $\hat{\theta}$ should be close to θ. We look at $\mathbb{E}\left[\hat{\theta} - \theta\right]$.

Definition 1.7 (Bias) – *The bias of $\hat{\theta}$ is defined by :*

$$B\left(\hat{\theta}, \theta\right) = \mathbb{E}\left[\hat{\theta}\right] - \theta$$

*We say that $\hat{\theta}$ is **unbiased** if $B\left(\hat{\theta}, \theta\right) = 0$.*

Definition 1.8 (Quadratic Risk / MSE) –

$$R\left(\hat{\theta}, \theta\right) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right]$$

*This is the **Mean Squared Error (MSE)** in English.*

We say that $\hat{\theta}_1$ is better than $\hat{\theta}_2$ if and only if $R\left(\hat{\theta}_1, \theta\right) \leq R\left(\hat{\theta}_2, \theta\right)$.

1.4.1 Example : Poisson Model

Let $X_1, ..., X_n$ be distributed according to a P_θ Poisson law, with $\theta > 0$. We seek an estimator for $\theta = \mathbb{E}[X_i]$.

Let's propose : $\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Bias Calculation :

$$\begin{aligned}
B\left(\hat{\theta}, \theta\right) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \theta \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] - \theta \quad \text{(by linearity)} \\
&= \frac{1}{n} \cdot n \cdot \mathbb{E}[X_1] - \theta \\
&= \theta - \theta = 0
\end{aligned}$$

Thus $\mathbb{E}\left[\overline{X}\right] = \theta$, is the unbiased estimator.

Risk Calculation :

$$\begin{aligned} R\left(\hat{\theta},\theta\right) &= \mathbb{E}\left[\left(\overline{X}-\theta\right)^2\right] = \mathbb{E}\left[\left(\overline{X}-\mathbb{E}\left[\overline{X}\right]\right)^2\right] \\ &= \operatorname{Var}\left(\overline{X}\right) = \operatorname{Var}\left(\frac{1}{n}\sum X_i\right) \\ &= \frac{1}{n^2}\sum \operatorname{Var}(X_i) \quad \text{(because i.i.d)} \\ &= \frac{1}{n^2}\cdot n\cdot \operatorname{Var}(X_1) = \frac{\operatorname{Var}(X_1)}{n} = \frac{\theta}{n} \end{aligned}$$

Theorem 1.9 (Bias-Variance Decomposition of Risk) –

$$R\left(\hat{\theta},\theta\right) = \left(B\left(\hat{\theta},\theta\right)\right)^2 + \operatorname{Var}\left(\hat{\theta}\right)$$

Proof.

$$\begin{aligned} R\left(\hat{\theta},\theta\right) &= \mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2\right] \\ &= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]+\mathbb{E}\left[\hat{\theta}\right]-\theta\right)^2\right] \\ &= \mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}\left[\hat{\theta}\right]-\theta\right)^2\right] + 2\mathbb{E}\left[\left(\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]\right)\left(\mathbb{E}\left[\hat{\theta}\right]-\theta\right)\right] \\ &= \operatorname{Var}\left(\hat{\theta}\right) + \left(B\left(\hat{\theta},\theta\right)\right)^2 + 2\left(\mathbb{E}\left[\hat{\theta}\right]-\theta\right)\underbrace{\mathbb{E}\left[\hat{\theta}-\mathbb{E}\left[\hat{\theta}\right]\right]}_{0} \\ &= \operatorname{Var}\left(\hat{\theta}\right) + B\left(\hat{\theta},\theta\right)^2 \end{aligned}$$

□

1.5 Consistency

Asymptotic property. We only consider consistent estimators.

Definition 1.10 (Consistency) – *Let $(X_1, ..., X_n)$ be i.i.d. from distribution P_θ. Let $\hat{\theta}_n = h(X_1, ..., X_n)$. $\hat{\theta}_n$ is a consistent (or convergent) estimator of θ if and only if :*

$$\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta$$

Remark 1.11 – *$\hat{\theta}_n$ is strongly consistent if and only if $\hat{\theta}_n \xrightarrow[n\to+\infty]{\text{p.s.}} \theta$.* ◇

1.5.1 Example : Revisiting the Poisson model

$\Theta = \mathbb{R}_+^*$, $\hat{\theta}_n = \overline{X}$.

- We can invoke the Law of Large Numbers (LLN) : $\overline{X} \xrightarrow{\mathbb{P}} \mathbb{E}[X_i] = \theta$.
- Via the quadratic risk :

$$R\left(\hat{\theta}_n,\theta\right) = \operatorname{Var}\left(\overline{X}\right) = \frac{\theta}{n} \xrightarrow[n\to+\infty]{} 0$$

According to Bienaymé-Chebyshev's inequality :

$$P\left(|\hat{\theta}_n - \theta| > \varepsilon\right) \leq \frac{\mathbb{E}\left[\left(\hat{\theta}_n - \theta\right)^2\right]}{\varepsilon^2} = \frac{R\left(\hat{\theta}_n, \theta\right)}{\varepsilon^2} \to 0$$

1.5.2 "Plug-in" Method

Let $(X_1, ..., X_n)$ be i.i.d. Poisson(θ). We want to estimate $\beta = P(X_i = 0) = e^{-\theta}$.

$$\hat{\beta} = e^{-\hat{\theta}} = e^{-\overline{X}}$$

$\hat{\beta}$ is consistent for estimating β.

Lemma 1.12 (Continuous Mapping Lemma) – *If $Z_n \xrightarrow{\mathbb{P}} Z$, then $h(Z_n) \xrightarrow{\mathbb{P}} h(Z)$ for any continuous function h.* ◇

Estimators

§2

2.1 Parametric Framework

2.1.1 Parametric statistical model

We have an observation $(X_1, ..., X_n)$, an i.i.d random variable sample (independent, identically distributed) with common distribution P belonging to a parameterized family of probability distributions $\{P_{\theta, \theta \in \Theta \subset \mathbb{R}^p}\}$.

Remark 2.1 – *If $\Theta \subset$ infinite-dimensional space $\to$ non-parametric model.* ◇

Estimating P is estimating $\theta \in \mathbb{R}^p$.

Example 2.2 – Bernoulli *(θ)*, Exp *(θ), $\mathcal{N}(\mu, \sigma^2)$, density distribution $f_\theta(x) = \theta x^{\theta-1} 1_{x \in [0,1]}$* ◇

Notation 2.3 – $E_{\theta_n}[h(X_1, ..., X_n)]$, $\Theta[h(X_1, ..., X_n)]$
Distribution of $(X_1, ..., X_n) \to P_\theta^{\otimes n}$ ◇

Definition 2.4 (Estimateur) –

$$\hat{\theta} = \hat{\theta}_n = h(X_1, ..., X_n)$$

Definition 2.5 (Qualité) –

- *Risque*

$$R\left(\hat{\theta}, \theta\right) = E_\theta\left[\left(\hat{\theta} - \theta\right)^2\right]$$

- *Consistance*

$$\hat{\theta}_n \underset{n \to +\infty}{\overset{P}{\to}} \theta$$

Definition 2.6 (Modèle identifiable) –

$$\theta \to P_\theta \quad \text{injective}$$

2.2 Method of moments

Definition 2.7 – *The theoretical moment of the distribution of X_i of order k is called:*

$$\mu_k = E\left[X_i^k\right], \quad k \geq 1$$

Definition 2.8 – *The empirical moment of the distribution of X_i of order k is called:*

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

By the law of large numbers $\hat{\mu}_k \underset{n\to+\infty}{\xrightarrow{P}} \mu_k$.

The method of moments: if we can write θ or $g(\theta)$, the parameter of interest, as a function of the k first theoretical moments.

$$\theta = \mathcal{L}(\mu_1, ..., \mu_k)$$

then the estimator

$$\hat{\theta} = \mathcal{L}(\hat{\mu}_1, ..., \mu_k)$$

is obtained by the method.

Example 2.9 (Calculations of estimators using the method of moments) –

- $X_i \sim \text{Bernoulli}(\theta)$ *with values 0-1,*

$$\theta = P(X_i = 1) = E[X_i] \to \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

- $X_i \sim \text{Exp}(\theta),, f_\theta(x) = \theta e^{-\theta x}1_{x\geq 0}$, $E[X] = \frac{1}{\theta} \iff \theta = \frac{1}{\mu_1}$, *by the method of moments,*

$$\hat{\theta} = \frac{1}{\hat{\mu}_1} = \frac{1}{\overline{X}}$$

$$\Theta(X_i) = \frac{1}{\theta^2} \iff \theta^2 = \frac{1}{E[X_i^2] - E[X_i]^2}$$

$$\iff \theta = \frac{1}{\sqrt{\mu_2 - \mu_1^2}}$$

$$\implies \hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}\right)^2}}$$

- $X_1, ..., X_n$ *i.i.d. from the distribution* P_θ *with density*

$$f_\theta(x) = \theta x^{\theta-1}1_{x\in[0,1]}$$

$$E_\theta[X_i] = \theta\int_0^1 x^\theta \, \mathrm{d}x = \frac{\theta}{\theta+1}$$

Method of moments:

$$(\theta + 1)\mu_1 = \theta \Longleftrightarrow \theta(1 - \mu_1) = \mu_1 \Longleftrightarrow \theta = \frac{E[X_i]}{1 - E[X_i]}$$

$$\Longrightarrow \hat{\theta}_{\mathbb{M}} = \frac{\overline{X}}{1 - \overline{X}},, P_\theta\left(\overline{X} = 1\right) = P_\theta(X_1 = X_2 = \dots = X_n = 1) = 0$$

◇

2.3 Rendered on the C.A.L.

(C.A.L. = Continuous Applications Lemma) $(X_n)_{n\geq 1}$ sequence of random variables. If X_n converges to X, what can be said about $g(X_n)_{n\geq 1}$? If g is continuous, C.A.L. applies.

- if $X_n \xrightarrow{P} X$ then $g(X_n) \xrightarrow{P} g(X)$
- if $X_n \xrightarrow{\mathcal{L}} X$ then $g(X_n) \xrightarrow{\mathcal{L}} g(X)$

Remark 2.10 (Sufficient condition) –

$$D_g = \{\text{points of discontinuity of } \ g\}$$

if $P\left(X \in D_g\right) = 0$, *the C.A.L. holds true.* ◇

Example 2.11 –

$$g(x) = \frac{x}{1 - x}$$

- *LLN:* $\overline{X} \xrightarrow{P} E[X]$
- *C.A.L.:* $g\left(\overline{X}\right) = \hat{\theta}_n \xrightarrow[n\to+\infty]{P} g(E[X]) = \theta$

◇

C.A.L. for pairs of sequences of random variables:

- if $(X_n, Y_n) \xrightarrow{P} (X, Y)$, then $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$, if $g : \mathbb{R}^2 \to \mathbb{R}$ or $\mathbb{R}^2$ is continuous
- if $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, then $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, Y)$

Example 2.12 –

$$\hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\overline{X}\right)^2}} \quad \text{consistent?}$$

LLN:

- $\overline{X} \xrightarrow{P} \mu_1$
- $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \mu_2$

therefore

$$\begin{pmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$g(x, y) = \frac{1}{\sqrt{y - x^2}} \Longrightarrow \hat{\theta}^{\mathbb{M}}$ *consistent for* θ, g *is continuous except at* $\left\{(x, y) \in \mathbb{R}^2, y = x^2\right\}$ *of measure zero.*

But this is false for convergence in distribution. ◇

PROPOSITION 2.13 (CONVERGENCE OF PAIRS) –

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix} \text{ iff } \begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases}$$

Proof.

- $\Longrightarrow$ then C.A.L. $g(x,y) = x$ is continuous, so $X_n \to X$ and $Y_n \to Y$
- $\Longleftarrow$ convergence of the pair?

$$\forall \varepsilon > 0, P(|X_n - X| + |Y_n - Y| > \varepsilon) \leq \underbrace{P\Big(|X_n - X| > \frac{\varepsilon}{2}\Big)}_{\to 0} + \underbrace{P\Big(|Y_n - Y| > \frac{\varepsilon}{2}\Big)}_{\to 0}$$

<u>This converse is false for convergence in distribution!</u>

□

2.3.1 EMPIRICAL VARIANCE

If the X_i have an expectation μ and a variance σ^2, we call the empirical variance

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 + \frac{1}{n}\sum_{i=1}^n \overline{X}^2 - \frac{2}{n}\sum_{i=1}^n X_i\overline{X} \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 + \overline{X}^2 - 2\overline{X}\overline{X} = \tilde{\sigma}^2 \end{aligned}$$

the moments estimator:

$$\sigma^2 = E[X_i^2] - E[X_i]^2$$

We replace theoretical moments with empirical moments

$$\to \widetilde{\sigma^2}^{\mathbb{M}} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\overline{X}\right)^2$$

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

$$\begin{cases} \overline{X} \xrightarrow{P} E[X] \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} E[X^2] \end{cases} \xrightarrow{\text{cv en proba}} \begin{pmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{pmatrix} \xrightarrow{\text{LAC}} \hat{\sigma}^2 \text{ which is consistent with } \operatorname{Var}(X) = E[X^2] - E[X]^2$$

EXAMPLE 2.14 –

- *calculate the bias of* $\hat{\sigma}_n^2$
- *calculate the risk of* $\hat{\sigma}_n^2$

◇

2.4 Maximum likelihood method

2.4.1 Given model

$(P_\theta)_{\theta\in\Theta}$ is given if there exists a measure μ (positive σ -defined $\to X_i$ with values in E, $E = \cup$ E_n with $\mu(E_n)$ finite) such that $\forall\theta, P_\theta$ admits a density with respect to μ.

2.4.2 In practice

- either E is at most countable: μ = counting measure. If $\exists, \{a_1, a_2, ...\}$ s.t. $\sum_{k\geq 1} P_\theta(X_i = a_k) = 1$, then $\mu = \sum_{k\geq 1}\delta_{a_k}$ with $\delta_a(\{a\}) = 1$ Dirac measure.

  Example 2.15 – Bernoulli *(θ), $X_i = 1$, probabilities $\theta \to \mu = \delta_0 + \delta_1$ We will write*

  $$f_\theta(x) = \underbrace{P_\theta(\{x\})}_{=1-\theta} - P_\theta(X_i = x) \quad \text{with} \quad x \in \{a_1, a_2, ...\}$$

  ◇

- or $E = \mathbb{R}^p$, then f_θ is the usual density

f_θ density of P_θ

Definition 2.16 – *We call the likelihood of the sample $(X_1, ..., X_n)$ the function*

$$\theta \to L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) \quad \text{(random variable)}$$

Definition 2.17 – *A maximum likelihood estimator $\hat{\theta}_{MV}$ is defined by:*

$$\forall\theta \in \Theta, L_n(\theta) \leq L_n\left(\hat{\theta}\right)$$

We often work with the **log-likelihood**

$$\log L_n(\theta) = \sum_{i=1}^{n} \ln f_\theta(X_i) \quad \text{sum of random variables}$$

$$\log L_n\left(\hat{\theta}\right) = \sup_{\theta\in\Theta} \log L_n(\theta)$$

Remark 2.18 – *$\hat{\theta}$ is a random variable*

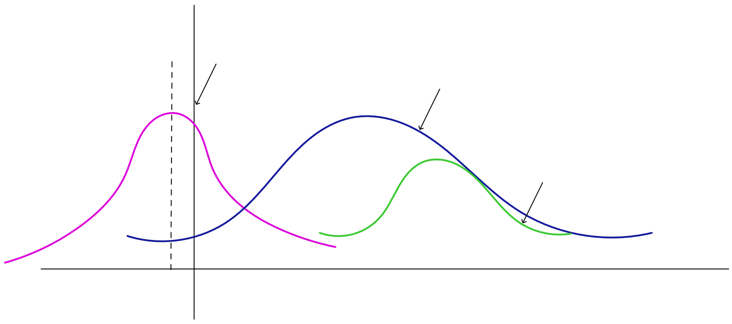
$f_\theta(x)$

θ_1

θ_2

θ_3

x



◇

EXAMPLE 2.19 –

- Bernoulli(θ), $f_\theta(x) = \theta^x(1-\theta)^{1-x}$, X_i *taking values 0-1*

$$L_n(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}$$

$$\log L_n(\theta) = \left(\sum_{i=1}^{n} X_i\right)\ln\theta + \left(n - \sum_{i=1}^{n} X_i\right)\ln(1-\theta)$$

$$(\log L_n)'(\theta) = \frac{\sum_{i=1}^n X_i}{\theta} - \frac{n - \sum_{i=1}^n X_i}{1-\theta} = \frac{\sum_{i=1}^n X_i - n\theta}{\theta(1-\theta)}\left(\overline{X} - \theta\right)$$

Likelihood equation:

$$(\log L_n)'(\theta) = 0 \Longleftrightarrow (1-\theta)\sum_{i=1}^{n} X_i = \left(n - \sum_{i=1}^{n} X_i\right)\theta$$

$$\Longleftrightarrow \sum_{i=1}^{n} X_i = n\theta \Longrightarrow \theta = \frac{\sum_{i=1}^n X_i}{n}$$

Is the critical point a maximum?

The derivative changes sign at $\overline{X}$ $\to$ *we indeed have a maximum* $\to$ $\hat{\theta}^{MV} = \overline{X}$
Second-order condition, if $(\log L_n)''(\theta) < 0$ *for all* $\theta \Longrightarrow \log L_n$ *is concave* $\Longrightarrow$ *global maximum*

$$(\log L_n)''(\theta) = -\frac{\sum_{i=1}^n X_i}{\theta^2} - \frac{n - \sum_{i=1}^n X_i}{(1-\theta)^2} < 0,, \forall\theta$$

◇

Fisher Information, efficiency

§3

Let $(P_\theta)_{\theta\in\Theta}$, $\Theta \subset \mathbb{R}^p$ (identifiable, given). Let f_θ be the density of P_θ

$$\operatorname{Supp} f_\theta = \left\{x \in E,, f_{\theta(x)} > 0\right\}$$

Given $(X_1, ..., X_n)$, i.i.d. from distribution P_θ and $\theta \mapsto L(\theta) = \prod_{i=1}^n f_{\theta(X_i)}$ as the likelihood of the sample. On $\operatorname{Supp} f_\theta$ we can calculate

$$\log L_{n(\theta)} = \sum_{i=1}^n \log f_{\theta(X_i)}$$

$$\hat{\theta} = \operatorname{argmax}_{\theta\in\Theta} \log L_{n(\theta)}$$

Proposition 3.1 – *If $\hat{\theta}$ MLE[1] for θ, $g\left(\hat{\theta}\right)$ is an MLE for $g(\theta)$*

Objective: what "better" estimator can we have? $\longrightarrow$ regular model

3.1 Regular Model

Definition 3.2 – *The model $(P_\theta)_{\theta\in\Theta}$ is said to be regular if*

1. *Θ is an open set and $\theta \mapsto f_{\theta(x)}$ is C^1*
2. *$\operatorname{Supp} f_\theta$ does not depend on θ: $S = \left\{x,\ f_{\theta(x)} > 0\right\}$*
3. *For all θ, the mapping*

$$x \mapsto \frac{\frac{\partial f_\theta}{\partial \theta}(x)}{f_{\theta(x)}} \mathbb{1}_{f_{\theta(x)}>0}$$

is integrable (L, μ) and the integral

$$I(\theta) = \int_S \frac{\frac{\partial f_\theta}{\partial \theta}(x)}{f_{\theta(x)}} \mathbb{1}_{f_{\theta(x)}>0} dx$$

is continuous on Θ.

Notation 3.3 – *We denote the derivative of $f_{\theta(x)}$ with respect to θ: $\frac{\partial f_\theta}{\partial \theta}(x)$ The quantity $I(\theta)$ is called the **Fisher Information of the model.*** ◇

Example 3.4 –

- *$f_{\theta(x)} = \theta e^{-x\theta}$ density with respect to $\mu(dx) = \mathbb{1}_{x\geq 0} dx$*

[1]**MLE** = Maximum Likelihood Estimator

$\theta \mapsto \theta e^{-x\theta}$ *is* C^∞ *on* $\Theta =]0, +\infty[$, $\operatorname{Supp} f_\theta = \mathbb{R}_+$

$$\frac{\partial f_\theta}{\partial \theta}(x) = (1 - x\theta)e^{-x\theta}$$

$$\frac{(1-x\theta)^2 \left(e^{-x\theta}\right)^2}{\theta e^{-x\theta}} = \frac{(1-x\theta)^2}{\theta} e^{-x\theta}$$

$$\begin{aligned} I(\theta) &= \int_\theta^\infty \frac{(1-x\theta)^2}{\theta^2} \theta e^{-x\theta} dx \\ &= \frac{1}{\theta^2} E_\theta (1 - X\theta)^2 \\ &= \frac{1}{\theta^2}[1 - 2\theta E(X) + \theta^2 E(X^2)] = \frac{1}{\theta^2} \end{aligned}$$

continuous on $]0, +\infty[$

◇

Example 3.5 – Bernoulli(θ), $x = 0.1$, $f_{\theta(0)} = 1 - \theta$, $f_{\theta(1)} = \theta$, *density with respect to* $\delta_0 + \delta_1$
For all $x \in \{0, 1\}$, $\theta \mapsto f_{\theta(x)}$ *is* C^1

$$\frac{\left(\frac{\partial f_{\theta(0)}}{\partial \theta}\right)^2}{f_{\theta(0)}} = \frac{1}{1-\theta}$$

$$\frac{\left(\frac{\partial f_{\theta(1)}}{\partial \theta}\right)^2}{f_{\theta(1)}} = \frac{1}{\theta} \Rightarrow I(\theta) = \frac{1}{1-\theta} + \frac{1}{\theta} = \frac{1}{\theta(1-\theta)}$$

continuous on $]0, 1[$

◇

Example 3.6 – $f_{\theta(x)} = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x) = \frac{1}{\theta}\mathbb{1}_{[x,+\infty[(\theta)}$ *non-regular model* ◇

3.2 Score and Fisher Information

$(X_1, ..., X_n)$ i.i.d. following the law of P_θ, f_θ

Definition 3.7 – *We call* ***score*** *or* ***score vector*** *the derivative of the log-likelihood* $\frac{\partial}{\partial \theta} \log L_{n(\theta)} = S_{n(\theta)} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(X_i)$

Example 3.8 – $X_i \sim \mathcal{E}(\theta)$, $L_{n(\theta)} = \theta^n e^{-\theta \sum_i X_i}$, $\log L_n(\theta) = n \log \theta - \theta \sum_i X_i$, *hence* $S - n(\theta) = \frac{n}{\theta} - \sum_{i=1}^n X - i$ ◇

Remark 3.9 –

$$E(S_n(\theta)) = E\left[n\left(\frac{1}{\theta} - \frac{\sum X_i}{n}\right)\right]$$

◇

Supplementary regularity hypothesis: (H) for any estimator $h(X)$ and any θ, the following integrals exist and are equal:

$$\frac{\partial}{\partial\theta}\int_S h(x)f_\theta(x)dx = \int_S h(x)\frac{\partial f_\theta}{\partial\theta}(x)dx$$

Remark 3.10 – *condition for applying Lebesgue's differentiation theorem.*

$$h \sup_{\tilde\theta \in V_\theta} |\frac{\partial f_\theta}{\partial\theta}(x)| \in L_1(\mu)$$

◇

Proposition 3.11 – *Under (H), the score is centered (P_θ), $n = 1$*

$$E_\theta\left[\frac{\partial}{\partial\theta}\log L_1(\theta)\right] = \int_S \frac{\partial}{\partial\theta}\log f_\theta(x)dx = \int_S \frac{\frac{\partial}{\partial\theta}f_{\theta(x)}}{\cancel{f_\theta(x)}}\cancel{f_\theta(x)}dx = \int_S \frac{\partial f_\theta}{\partial\theta}(x)dx \underbrace{=}_{(H)} \frac{\partial}{\partial\theta}\overbrace{\int f_\theta(x)dx}^{=1} = 0$$

Definition 3.12 – *The Fisher information associated with $(X_1, ..., X_n)$*

$$I_n(\theta) \underbrace{=}_{\text{def}} E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L_{n(\theta)}\right)^2\right] \overbrace{=}^{\text{cor. de la prop 1}} = \mathrm{Var}_\theta\left[\frac{\partial \log L_n(\theta)}{\partial\theta}\right]$$

$$(*) E_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right]^2 = \int_S\left(\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}\right)^2 f_\theta(x)dx = \int_S \frac{\left(\frac{\partial}{\partial\theta}f_\theta(x)\right)^2}{f_\theta(x)} = \text{"expression from definition 1"}$$

Example 3.13 – $(X_1, ..., X_n) \sim \mathcal{E}(\theta)$, $\frac{\partial}{\partial\theta}\log L_n(\theta) = \frac{n}{\theta} - \sum_{i=1}^n X_i$

$$I_n(\theta) = E\left(\left(\frac{n}{\theta} - \sum X_i\right)^2\right) = n^2 E\left[\left(\frac{1}{\theta} - \frac{\sum X_i}{n}\right)^2\right] = n^2\,\mathrm{Var}\left(\overline{X}\right) = n^2\frac{1}{n}\frac{1}{\theta^2} = \frac{n}{\theta^2}$$

◇

Proposition 3.14 –

$$I_n(\theta) = nI(\theta)$$

indeed,

$$I_n(\theta) = \mathrm{Var}\left(\frac{\partial}{\partial\theta}\log L_n(\theta)\right) = \mathrm{Var}\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right) \underbrace{=}_{\text{independance}} = \sum_{i=1}^{n}\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right) =$$

$$= n\underbrace{\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right)} = nI(\theta)$$

Example 3.15 – $(X_1, ..., X_n)$ *i.i.d* $\mathcal{P}(\theta)$, $f_\theta(x) = e^{-\theta}\frac{\theta^x}{x!}$

$$\log L_n(\theta) = -n\theta + \left(\sum X_i\right)\log\theta - \log\prod_{i=1}^{n} X_i!$$

$$\frac{\partial}{\partial\theta}\log L_n(\theta) = -n + \frac{\sum X_i}{\theta} \Rightarrow I_n(\theta) = \mathrm{Var}\left(\frac{\sum X_i}{\theta}\right) = \frac{1}{\theta^2}n\theta = \frac{n}{\theta}$$

◇

3.3 Fisher Information and Second Derivative

Proposition 3.16 – *Adding that $\theta \mapsto f_\theta(x)$ is C^2 and that (H) holds for $\frac{\theta^2}{\partial\theta^2}$, then Fisher's information can also be written as*

$$I_n(\theta) = -E_\theta\left[\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2}\right]$$

if $\hat\theta$ is MLE, $I_{n(\hat\theta)} > 0$

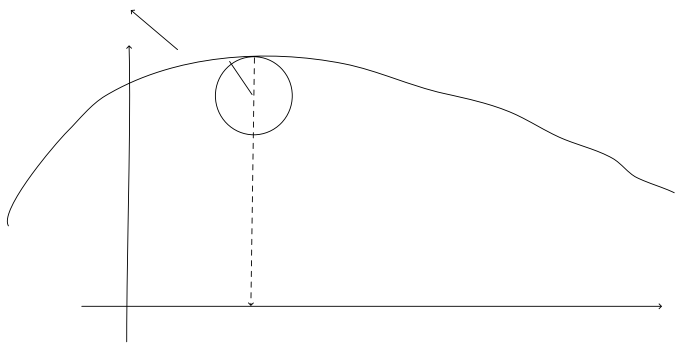
$n = 1$

$\frac{\partial^2}{\partial\theta^2}\log f_\theta(x) = \frac{\left(\frac{\partial^2 f_\theta(x)}{\partial\theta^2}\right)^2}{f_\theta(x)} - \frac{\left(\frac{\partial f_\theta}{\partial\theta}(x)\right)^2}{f_\theta^2(x)}$

$$E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_1)\right] = \int_S \frac{\frac{\partial^2 f_{\theta(x)}}{\partial\theta^2}}{\cancel{f_\theta(x)}}\cancel{f_\theta(x)}dx - \underbrace{\int_S \frac{\left(\frac{\partial f_\theta(x)}{\partial}\theta\right)^2}{f_\theta^2(x)}dx}_{I(\theta)}$$

$\log L(\theta)$

$\hat\theta$ θ



If the curve is very "peaked" at the MLE (i.e., Fisher information is large), then the MLE is precisely localized.

3.4 Cramer-Rao Inequality

Let $g(\theta)$ be the parameter of interest where $g : \Theta \to \mathbb{R}$

Proposition 3.17 – *Under the assumptions of a regular model, if for all θ, $I(\theta) > 0$, then for any <u>sans biais</u> estimator $T = T(X_1, ..., X_n)$, $E_\theta T^2 < +\infty$, we have*

$$\forall \theta \in \Theta, \underbrace{\mathrm{Var}_\theta(T) \geq \frac{(g'(\theta)^2)}{I_n(\theta)}} = \frac{(g'(\theta))^2}{nI(\theta)}$$

Proof.

$$\forall \theta \;\; E_{\theta(T)} = g(\theta)$$

$$\Rightarrow \frac{\partial}{\partial \theta} E_\theta(T) = g'(\theta)$$

$$\overset{T=T(X_1)}{\Leftrightarrow} \frac{\partial}{\partial \theta} \int_S T(x) f_\theta(x) dx = g'(\theta)$$

$$\overset{(H)}{\Leftrightarrow} \int_S T(x) \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx = g'(\theta)$$

$$\Leftrightarrow \int_S (T(x) - g(\theta)) \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) dx = g'(\theta)$$

□

Cauchy-Schwarz Inequality for $\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) f_\theta(x) dx$ with $h_1(X)$ and $h_2(X)$ centered

$$\left(\langle T(X) - g(\theta), \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \rangle_\theta \right)^2 = (g'(\theta))^2 \underbrace{\leq}_{\text{c.s}} \underbrace{\int (T(x) - g(\theta))^2 f_\theta(x) dx}_{= \mathrm{Var}_\theta(T)} \times \underbrace{\int \left(\frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) dx}_{= I(\theta)}$$

DEFINITION 3.18 – *If T attains the equality, then T is called* ***efficient****.*

Asymptotic study of estimators

§4

In a regular parametric model, if $\hat{\theta}_n$ is an estimator of θ, then

$$\operatorname{Var}\left(\hat{\theta}_n\right) \geq \frac{1}{I_{n(\theta)}} = \frac{1}{nI(\theta)}$$

if $\operatorname{Var}\left(\hat{\theta}_n\right) = \frac{1}{nI(\theta)}$ and it is unbiased, $\hat{\theta}_n$ is efficient efficient

Asymptotic: $n \to +\infty$,

$$n \operatorname{Var}\left(\hat{\theta}_n\right) \underset{n\to+\infty}{\longrightarrow} \frac{1}{I(\theta)}$$

4.1 Convergences

$(X_n)_{n\geq 0}$ sequence of real random variables $\left(\mathbb{R}^d\right)$

-
- convergence in distribution: $X_n \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} X$ iff $P(X_n \leq x) \longrightarrow P(X \leq x)$ at every continuity point of x.

Lemma 4.1 (lemme de Portmanteau) – *Equivalent characterizations:*

- *For any bounded continuous function h,*

$$E[h(X_n)] \longrightarrow E[h(X)]$$

$\Rightarrow$ convergence in distribution is stable under continuous mappings (CMT) MAIS it is generally faux that if $X_n \overset{\mathcal{L}}{\longrightarrow} X$ and $Y_n \overset{\mathcal{L}}{\longrightarrow} Y$ then $\binom{X_n}{Y_n} \overset{\mathcal{L}}{\longrightarrow} \binom{X}{Y}$

This is true in 3 cases:

1. si $\begin{cases} \forall n, X_n \text{ et } Y_n \text{ sont indépendantes} \\ X \text{ et } Y \text{ sont indépendantes} \end{cases}$ alors $\begin{cases} \text{convergence en loi de } X_n \text{ et } Y_n \\ \text{convergence en loi du couple } \binom{X_n}{Y_n} \end{cases}$

2. $$\text{si } \begin{cases} X_n \overset{P}{\longrightarrow} X \\ Y_n \overset{p}{\longrightarrow} Y \end{cases} \Longrightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{P}{\longrightarrow} \begin{pmatrix} X \\ Y \end{pmatrix} \Longrightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{\mathcal{L}}{\longrightarrow} \begin{pmatrix} X \\ Y \end{pmatrix}$$

3. *(Slutsky's Lemma) (the most important)*

$$\text{si } \begin{cases} X_n \overset{\mathcal{L}}{\longrightarrow} X \\ Y_n \overset{\mathcal{L}}{\longrightarrow} c \end{cases} \text{ alors } \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{\mathcal{L}}{\longrightarrow} \begin{pmatrix} X \\ c \end{pmatrix}$$

by applying the CMT,

$$\begin{aligned} h(x,y) &= x+y & X_n + Y_n &\xrightarrow{\mathcal{L}} X + c \\ &= xy & X_n Y_n &\xrightarrow{\mathcal{L}}\xrightarrow{\mathcal{L}} cX \\ &= \frac{x}{y} & \frac{X_n}{Y_n} &\xrightarrow{\mathcal{L}} \frac{X}{c} \end{aligned}$$

◇

4.2 Consistency of Estimators

Definition 4.2 – $\hat{\theta}_n$ *is asymptotically unbiased if and only if*

$$\text{Biais}\left(\hat{\theta}_n, \theta\right) = E\left[\hat{\theta}_n\right] - \theta \underset{n \to +\infty}{\longrightarrow} 0$$

Remark 4.3 – *Convergence in probability does not imply convergence of expectations.*
If $X_n \xrightarrow{P} X$, $|X_n| \leq Y \in L'$, *then by dominated convergence* $X_n \longrightarrow X$ *in* L_1 ◇

Example 4.4 – $\hat{\tau}_n = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum X_i^2 - \left(\overline{X}\right)^2$ *moment estimator for* $\tau^2 = E[X^2)] - (E[X])^2$

$$\text{Biais}(\hat{\tau}_n^2, \tau^2) = -\frac{1}{n}\tau^2 \quad \text{asymptotically unbiased}$$

Consistency of $\hat{\tau}_n^2$*?*

Tools to show consistency:

- *LLN*
- *if* $R\left(\hat{\theta}_n, \theta\right) \longrightarrow 0$ *then* $\hat{\theta}_n$ *is consistent because* $L^2 \Rightarrow$ *convergence implies convergence in probability*
- *return to the definition of convergence in probability*

- *if* (X_i) *are i.i.d., then* (X_i^2) *is i.i.d.*

  $$E[X_i^2] < +\infty$$

  LLN: $\frac{1}{n}\sum_i X_i^2 \xrightarrow{P} E[X^2] = \tau^2 - \mu^2$
- $\overline{X} \xrightarrow{P} \mu$ *(LLN), CMT with* $h(x) = x^2$*:* $\left(\overline{X}\right)^2 \xrightarrow{P} \mu^2$
- *Therefore* $\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^2 \\ \frac{1}{n}\sum X_i \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \tau^2 - \mu^2 \\ \mu \end{pmatrix}$
- *CMT* $h(x,y) = x - y^2$

Therefore $\frac{1}{n}\sum X_i^2 - \left(\overline{X}\right)^2 \xrightarrow{P} \tau^2 + \mu^2 - \mu^2 = \tau^2$ ◇

4.3 Asymptotic normality of $\hat{\theta}_n$ for θ.

→ Question: what is the convergence rate of $\hat{\theta}_n$ towards θ ?

$(X_1, ..., X_n)$ i.i.d., with expectation θ, $\hat{\theta} = \overline{X}$ with variance $\tau^2(\theta)$

<u>CLT</u> $\sqrt{n}\left(\overline{X}-\theta\right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,\tau^2(\theta))$ <u>regardless of the distribution of</u> X_i

Definition 4.5 – *$\hat{\theta}_n$ is an <u>asymptotically normal</u> estimator if and only if*

- *convergence rate in $\sqrt{n}$*
- *convergence in distribution*
- *limiting distribution is normal*

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,\tau^2(\theta))$$

Example 4.6 – *Is $\hat{\tau}_n^2$ asymptotically normal?*

$(X_1,\dots,X_n)$ i.i.d. with expectation μ, with variance τ^2

$$\hat{\tau}_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2+\left(\overline{X}-\mu\right)^2+\underbrace{\frac{2}{n}\sum_{i=1}^{n}(X_i-\mu)\left(\mu-\overline{X}\right)}_{=2\left(\mu-\overline{X}\right)\left(\overline{X}-\mu\right)}$$

- *CLT: if (X_i) are i.i.d., then $(X_i-\mu)^2$ are i.i.d. with expectation τ^2,*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-\tau^2\right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,u_4-\tau^4)$$

$$\operatorname{Var}(X_i-\mu)^2=E\left[(X_i-\mu)^4\right]-\mu^4=\mu_4-\tau^4$$

- *CLT: $\sqrt{n}\left(\overline{X}-\mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\tau^2)$*
- $\sqrt{n}(\hat{\tau}_n^2-\tau^2)=\sqrt{n}\left(\frac{1}{n}\sum_i(X_i-\mu)^2-\tau^2\right)-\underbrace{\sqrt{n}\left(\overline{X}-\mu\right)^2}_{\sqrt{n}\underbrace{\left(\overline{X}-\mu\right)}_{\xrightarrow{\mathcal{L}}\mathcal{N}(0,\tau^2)}\times\underbrace{\left(\overline{X}-\mu\right)}_{\xrightarrow{\mathcal{L},P}0}}$

$$\left.\begin{array}{l}\overline{X}-\mu \xrightarrow{\mathcal{L}} 0\\ \sqrt{n}\left(\overline{X}-\mu\right) \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0,1)\end{array}\right\} \overset{\text{lemme de Slutsky}}{\Longrightarrow} \sqrt{n}\left(\overline{X}-\mu\right)^2 \xrightarrow[P]{\mathcal{L}} 0$$

$$\sqrt{n}(\hat{\tau}_n^2-\tau^2) \xrightarrow[n\to+\infty]{\mathcal{L}} Z+0$$

Therefore $\hat{\tau}_n^2$ is an asymptotically normal estimator ◇

Remark 4.7 – $\sqrt{n}\left(\hat{\theta}_n-\theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\tau^2) \iff \frac{\sqrt{n}\left(\hat{\theta}_n-\theta\right)}{\tau} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$ *Application of Slutsky's Lemma: if $\hat{\tau}^2$ is a consistent estimator of τ^2, then we still have*

$$\frac{\sqrt{n}\left(\hat{\theta}_n-\theta\right)}{\hat{\tau}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

◇

Proof.

$$\frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\hat{\tau}} = \underbrace{\left(\sqrt{n}\frac{\left(\hat{\theta}-\theta\right)}{\tau}\right)}_{\xrightarrow{\mathcal{L}} Z\sim\mathcal{N}(0,1)} \times \underbrace{\left(\frac{\tau}{\hat{\tau}}\right)}_{\xrightarrow{P} 1}$$

$$\xrightarrow{\mathcal{L}} 1 \times Z \text{ by Slutsky's Lemma and consistency of } \hat{\tau}$$

□

4.4 δ-METHOD

$\hat{\theta}$ asymptotically normal estimator: what is the asymptotic distribution of $g(\theta)$?

LEMMA 4.8 (MÉTHODE DÉLTA) – *Let Z_n be a sequence of real random variables s.t.*

$$\sqrt{n}(Z_n - \mu) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \tau^2)$$

Let g be a differentiable function, $g'(\mu) \neq 0$. Under these assumptions, we have

$$\sqrt{n}[g(Z_n) - g(\mu)] \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \tilde{Z} \sim \mathcal{N}\left(0, (g'(\mu))^2\tau^2\right)$$

$$g(x) = g(\mu) + g'(\mu)(x-\mu) + (x-\mu)R(x-\mu) \quad \text{where} \quad R(y) \underset{y\to 0}{\to} 0$$

$$\sqrt{n}(g(Z_n) - g(\mu)) = \underbrace{g'(\mu)\underbrace{\sqrt{n}(Z_n-\mu)}_{\xrightarrow{\mathcal{L}} Z\sim\mathcal{N}(0,\tau^2)}}_{\to\mathcal{N}\left(0,(g'(\mu))^2\tau^2\right)} + \underbrace{\underbrace{(\sqrt{n})(Z_n-\mu)}_{\xrightarrow{\mathcal{L}}\mathcal{N}(0,\tau^2)}\underbrace{R(Z_n-\mu)}_{\xrightarrow{P} 0?}}$$

Do we have $Z_n \xrightarrow{P} \mu$?

$$\begin{aligned}
P(|X_n - \mu| > \varepsilon) &= P\left(\frac{\sqrt{n}|Z_n-\mu|}{\tau} > \frac{\sqrt{n}\varepsilon}{\tau}\right) \\
&= P\left(\frac{\sqrt{n}(Z_n-\mu)}{\tau} > \frac{\sqrt{n}\varepsilon}{\tau}\right) + P\left(\frac{\sqrt{n}(Z_n-\mu)}{\tau} < -\frac{\sqrt{n}\varepsilon}{\tau}\right) \\
&\sim= 1 - \Phi_n\left(\frac{\sqrt{n}\varepsilon}{\tau}\right) + \Phi_n\left(-\frac{\sqrt{n}\varepsilon}{\tau}\right) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\tau}\right)\right)
\end{aligned}$$

◇

Empirical Distribution Function

§5

$(X_1, ..., X_n)$ i.i.d. real-valued sample from an unknown distribution F.

$$\forall x \in \mathbb{R}, F(x) = P(X_1 \leq x) = E\left[\mathbb{1}_{X_1 \leq x}\right]$$

Definition 5.1 – *The* *<u>empirical</u>* *distribution function associated with* $(X_1, ..., X_n)$ *is defined by:*

$$\begin{aligned} \hat{F}_n : \mathbb{R} &\longrightarrow [0, 1] \\ x &\mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} \end{aligned}$$

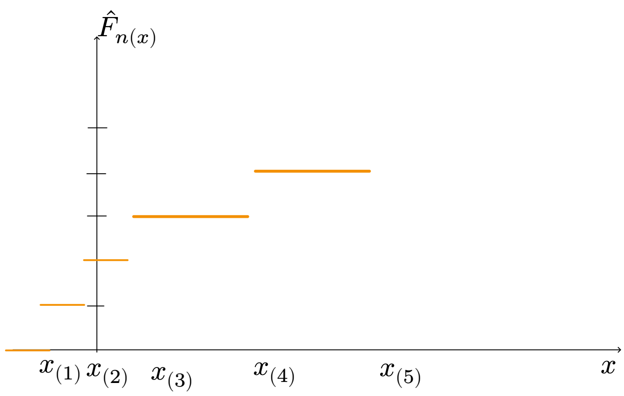
$\forall x \in \mathbb{R}, \hat{F}_n(x)$ *is a random variable, an estimator of* $F(x)$*.*

Definition 5.2 – *Empirical Law* $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ *is a discrete uniform law on* $\{X_1, ..., X_n\}$*.*

Graphical Representation

Conditionally $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)} \quad \text{ordered values}$$



Proposition 5.3 (Immediate Properties) –

- $n\hat{F}_n(x) = \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x}$ *follows the binomial law*$(n, F(x))$

- $R\left(\hat{F}_n(x), F(x)\right) = 0 + \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq x}\right) \underbrace{=}_{\text{indep}} \frac{1}{n}F(x)(1-F(x)) \underset{n\to+\infty}{\longrightarrow} 0$ *therefore* $\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow{P} F(x)$
- *or* *LLN*: $\hat{F}_n(x)$ *estimator* *consistent* *of* $F(x)$.
- *We have a uniform convergence result:*

$$\sup_{x\in\mathbb{R}}|\hat{F}_n(x) - F(x)| \underset{n\to+\infty}{\xrightarrow{P}} 0 \quad \text{(Glivenko-Cantelli Theorem)}$$

.[2]

- $\hat{F}_n(x)$ *is it asymptotically normal?*

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \leq x}$$

CLT: the X_i *are i.i.d., so the* $\left\{\mathbb{1}_{X_i\leq x} = Y_i\right\}$ *are i.i.d.*

$$\forall x, F(x) \in]0,1[, \quad \sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, F(x)(1-F(x)))$$

$$\Longleftrightarrow \frac{\hat{F}_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0,1)$$

5.1 Empirical Estimation

"plug-in" or substitution method, parameter of interest $\theta = c(F)$, the empirical method defines $\hat{\theta}$, an empirical estimator by replacing F with $\hat{F}_n \to \hat{\theta}_n = c\left(\hat{F}_n\right)$.

Example 5.4 – $\theta = E_F(X) \to \hat{\theta}_n = E_{\hat{F}_n}(X) = \sum_{i=1}^n X_i \times \frac{1}{n} = \overline{X}$ *if* X_i *are distinct*

$$\theta = \operatorname{Var}_F(X) \to \hat{\theta}_n = \operatorname{Var}_{\hat{F}_n}(X) = \frac{1}{n}\sum\left(X_i - \overline{X}\right)^2$$

◇

5.2 Generalized Inverse

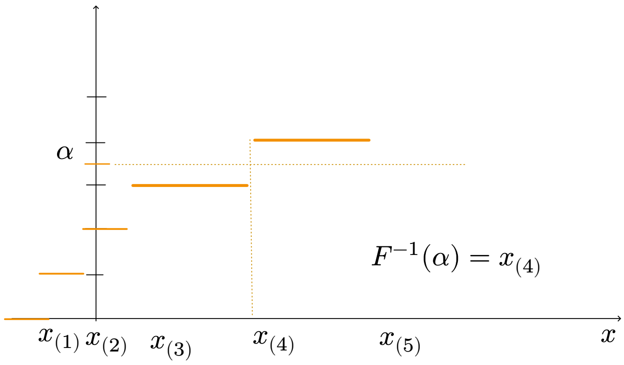
Definition 5.5 – *The generalized inverse of* F *is defined by:*

$$F^{-1} : [0,1] \longrightarrow \mathbb{R}$$

$$\forall \alpha \in [0,1], F^{-1}(\alpha) = \inf\{x \in \mathbb{R}, F(x) \geq \alpha\}$$

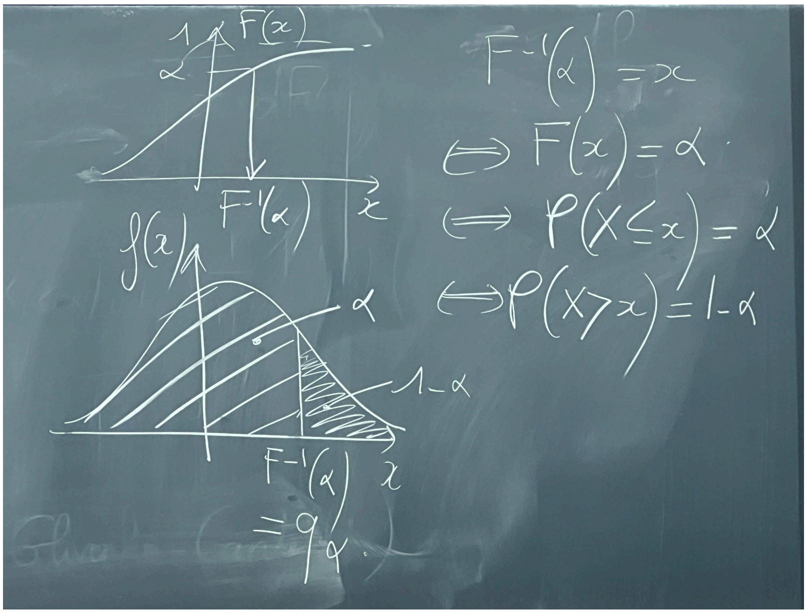
If F *is strictly increasing,* $\inf x$ *such that* $F(x) \geq a \Leftrightarrow x \geq F^{-1}(\alpha)$*, if* F *is the function of a discrete distribution.*

[2] Glivenko-Cantelli Thm: https://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_de_Glivenko-Cantelli



EXAMPLE 5.6 –

$$F^{-1}(\alpha) = x \Leftrightarrow F(x) = \alpha \Leftrightarrow P(X \leq x) = \alpha \Leftrightarrow P(X > x) = 1 - \alpha$$



◇

Vocabulary:

- F^{-1} is also called the <u>quantile function</u>
- $F^{-1}(\alpha) = \alpha$-order quantile, of the distribution F
- $F^{-1}\left(\frac{1}{4}\right) =$ 1st quantile
- $F^{-1}\left(\frac{1}{2}\right) =$ median
- $F^{-1}\left(\frac{3}{4}\right) =$ 3rd quantile

LEMMA 5.7 – *U a random variable on $[0, 1]$, F a c.d.f., then $F^{-1}(U)$ is a random variable with distribution F* ◇

- If F is bijective:

$$P(F^{-1}(U) \leq x) \underbrace{=}_{F \text{ bijective}} P(U \leq F(x)) \underbrace{=}_{\text{car } P(U \leq x) = x \text{ sur } [0,1]} F(x)$$

- If F is discrete: F^{-1} generalized inverse: $F^{-1}(y) \leq x \Leftrightarrow y \leq F(x)$

5.3 Empirical Quantile

Definition 5.8 – *We define the* <u>*empirical quantile*</u> *(sample quantile) of order α, as the quantile of $\hat{F}_n$:*

$$\hat{q}_{n,\alpha} = \hat{F}_n^{-1}(\alpha) = \inf\left\{x, \hat{F}_n(x) \geq \alpha\right\}$$

Proposition 5.9 –

- *It can be shown that $\hat{q}_{n,\alpha} = X_{([n\alpha])}$ where $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ is the ordered sample of $(X_i)_{1 \leq i < n}$*

$$[u] = \text{the smallest integer} \geq u$$

Example 5.10 – $\alpha = \frac{1}{2}$, $[\frac{n}{2}]$,

$$\begin{cases} \text{si } n = 2k \quad \text{medianne} = \hat{q}_{n,\frac{1}{2}} = X_{(k)} \\ \text{si } n = 2k+1 \quad \text{medianne} = \hat{q}_{n,\frac{1}{2}} = X_{(k+1)} \end{cases}$$

- *Consistency*

  if $\alpha \in]0,1[$, if F is strictly increasing in the neighborhood of α

◇

Confidence Intervals

§6

6.1 Definitions

$(X_1, ..., X_n)$ i.i.d. from distribution $P \in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$, we are interested in $\theta \in \mathbb{R}$ or $g(\theta) : \mathbb{R}^p \longrightarrow \mathbb{R}$.

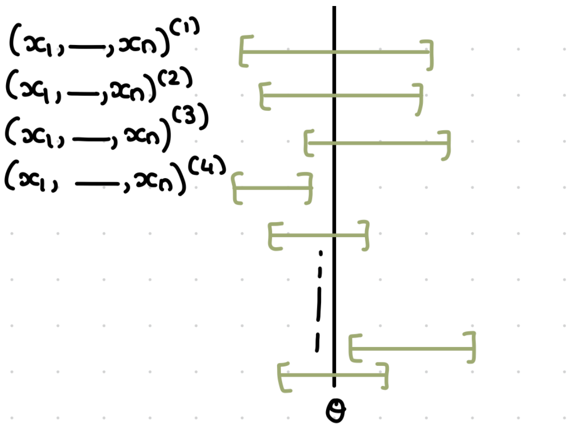
A confidence interval for θ, with a confidence level of $1 - \alpha, \alpha \in]0, 1[$ is an interval whose bounds are random, functions of the sample and do NOT depend on the unknown parameters of the model, and such that

$$P([B\inf(X_1, ..., X_n); B\sup(X_1, ..., X_n)] \ni \theta) \geq 1 - \alpha$$

[3]

- A CI is computable from the data
- if the inequality is an equality $=$, the confidence level is exact.
- if we have $P(\theta \in [B\inf, B\sup]) \underset{n \to +\infty}{\longrightarrow} 1 - \alpha$, the level is asymptotic.
- generally $\alpha = 1\%, 5\%$

6.2 Interpretation



$IC = [B\inf(X_1, ..., X_n), B\sup(X_1, ..., X_n)]$ mathematical formula that guarantees the level $1 - \alpha$. We observe $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$, a realization of the random sample. We calculate $IC = [2.3; 5.1]$ with a confidence level 95% ($\alpha = 5\%$).

On average, out of 100 calculated intervals (using the same formula), there are 5 intervals that do not contain θ.

$$P(\theta \in [B\inf, B\sup]) = 1 - \alpha$$

$\cancel{P(\theta \in [2.3, 5.1]) = 95\%}$ because θ is a number

6.3 Pivotal Method

$(X_1, ..., X_n)$ i.i.d. with expectation $\theta \in \mathbb{R}$, with variance $\sigma^2(\theta)$. Let $\hat{\theta}$ be asymptotically normal:

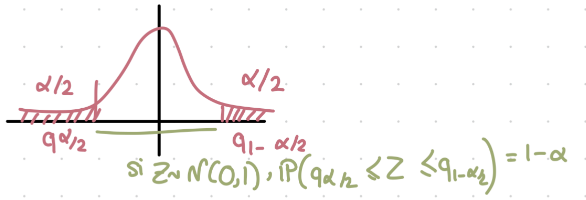
$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \underset{n \to +\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \sigma^2(\theta))$$

$$\Leftrightarrow \frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sigma(\theta)} \underset{n \to +\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, 1)$$

By definition of Gaussian quantiles, $q_\alpha = \Phi^{-1}(\alpha)$ where Φ is the c.d.f. of $\mathcal{N}(0, 1)$

[3] $B\inf$ for lower bound and $B\sup$ for upper bound

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\sigma(\theta)} \leq q_{1-\frac{\alpha}{2}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha \tag{1}$$



- **pivot or pivotal statistic** $= \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\hat{\sigma}}$ a centered and reduced statistic derived from $\hat{\theta}$, where $\sigma^2(\theta)$ is estimated by $\hat{\sigma}^2$, <u>consistent</u> for estimating $\sigma^2(\theta)$.

  If this is the case,

$$\underbrace{\frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\sigma(\theta)}}_{\underset{\hat{\theta}\text{ as. normal}}{\xrightarrow{\mathcal{L}}}\mathcal{N}(0,1)} \times \underbrace{\frac{\sigma^2(\theta)}{\hat{\sigma}^2}}_{\underset{\text{estimateur consistant}}{\xrightarrow{P}}1} \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0,1) \quad \text{by Slutsky's lemma}$$

- we deduce

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\hat{\sigma}}(\theta) \leq q_{1-\frac{\alpha}{2}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha$$

$$P\left(\hat{\theta}-\frac{1}{\sqrt{n}}\hat{\sigma}q_{1-\frac{\alpha}{2}} \leq \theta \leq \hat{\theta}-\frac{1}{\sqrt{n}}\hat{\sigma}q_{\frac{\alpha}{2}}\right) \longrightarrow 1-\alpha$$

Remark 6.1 (why $\frac{\alpha}{2}$?) – *We can observe that the quantiles in Equation 1 are of order $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$. To understand why, it is enough to perform a simple calculation. First, we note $\frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\sigma(\theta)} =: Z \sim \mathcal{N}(0,1)$.*

$$P\left(q_{\frac{\alpha}{2}} \leq Z \leq q_{1-\frac{\alpha}{2}}\right) = P\left(Z \leq q_{1-\frac{\alpha}{2}}\right) - P\left(Z \leq q_{\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}-\frac{\alpha}{2} = 1-\alpha$$

◇

Supplements (before midterm)

§7

1. Review of asymptotic normality
2. Example
3. Asymptotic pivot
4. Example 2

7.1 Asymptotic properties of a sequence of estimators $\left(\hat{\theta}_n\right)_{n\geq 1}$

- Consistency $\hat{\theta}_n \xrightarrow{P} \theta$
- Asymptotic normality, if there exists $\sigma^2 > 0$

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, \sigma^2)$$

In general, if there exists $v_n \underset{n\to+\infty}{\longrightarrow} +\infty$

$$v_{n\left(\hat{\theta}_n - \theta\right)} \xrightarrow{\mathcal{L}} Y$$

We say that $\hat{\theta}_n$ converges at rate $\frac{1}{v_n}$

Remark 7.1 – *If $\hat{\theta}_n$ is asymptotically normal $\Rightarrow$ $\hat{\theta}_n$ is consistent*

$$\hat{\theta}_n - \theta = \underbrace{\frac{1}{\sqrt{n}}}_{\to 0} \underbrace{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)} \underset{\text{Slutsky}}{\xrightarrow{\mathcal{L} \text{ ou } P}} 0$$

$$U_n = \frac{1}{\sqrt{n}} \to 0$$

◇

<u>δ-method</u>

$$\sqrt{n}(X_n - 1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$$\sqrt{n}(X_n - 1) \overset{\mathcal{L}}{\approx} Z \sim \mathcal{N}(0, 1)$$

$$X_n \overset{\mathcal{L}}{\approx} 1 + \frac{1}{\sqrt{n}} Z$$

If g is differentiable at 1,

$$g(1 + h) = g(1) + hg(1)$$

$$g(X_n) \approx g(1) + \frac{1}{\sqrt{n}} g'(1) Z$$

$$\sqrt{n}(g(X_n) - g(1)) \approx g'(1) Z$$

δ-method

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0,1)$$

g differentiable at θ

$$g(x) = g(\theta) + g'(\theta)[(x-\theta) + r(x)] \text{ where } r(x) \underset{x\to 0}{\to} 0$$

$\hat{\theta}_n \xrightarrow{P} \theta$ thus (LAC) $r\left(\hat{\theta}_n\right) \longrightarrow r(\theta) = 0$

$$g\left(\hat{\theta}_n\right) = g(\theta) + \left(\hat{\theta}_n - \theta\right)\left[g'(\theta) + r\left(\hat{\theta}_n\right)\right]$$

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right) - g(\theta)\right) = \underbrace{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}_{\xrightarrow{\mathcal{L}} Z}\left[\underbrace{g'(\theta) + r\left(\hat{\theta}_n\right)}_{\xrightarrow{P} g'(\theta)}\right] \underset{\text{Slutsky}}{\Rightarrow} \sqrt{n}\left(g\left(\hat{\theta}_n\right) - g(\theta)\right) \xrightarrow{\mathcal{L}} g'(\theta)Z \sim \mathcal{N}\left(0, (g'(\theta))^2\right)$$

Example 7.2 – *$X_1, .., X_n$ with density law $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}, x \geq 0, \mu = E[X_i] > 0$*

μ

$\hat{\mu} = \overline{X}$ *estimated by* $\log L_n(\mu) = -n\log\mu - \frac{1}{\mu}\sum_{i=1}^n X_i$ *efficient?*

$$\operatorname{Var}(\hat{\mu}) = \frac{1}{n^2}\operatorname{Var}\left(\sum_i X_i\right) \underbrace{=}_{\text{indép}} \frac{1}{n^2}\sum_i \operatorname{Var}(X_i) \underbrace{=}_{\text{i.i.d.}} \frac{1}{n}\operatorname{Var}(X_i) = \frac{\mu^2}{n}, E[\hat{\mu}] = \mu$$

$$\frac{\partial}{\partial\mu}(\log L_n)(\mu) = -\frac{n}{\mu} + \frac{1}{\mu^2}\sum(X_i)$$

$$\begin{aligned} I_{n(\mu)} &= \operatorname{Var}\left(-\frac{n}{\mu} + \frac{1}{\mu^2}\sum X_i\right) \\ &= \frac{1}{\mu^4}\operatorname{Var}\left(\sum X_i\right) \\ &= \frac{n}{\mu^4}\operatorname{Var}(X_i) \end{aligned}$$

$$I_{n(\mu)} = \frac{n}{\mu^2}$$

$\hat{\mu}$ $\operatorname{Var}(\hat{\mu}) = \frac{1}{I_{n(\mu)}}$ *is unbiased and* $\hat{\mu}$. *Therefore,*

is efficient.$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^2) \Rightarrow \operatorname{Var}(\hat{\mu}_n) = \frac{\mu^2}{n} =$ *CLT:*

$\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\mu}$ *variance of the asymptotic Gaussian distribution*$\mathcal{N}(0,1)$

- *has the asymptotic distribution* $(X_1, ..., X_n)$ *another parametrization:* $f(x) = \theta e^{-\theta x}, x \geq 0$ *i.i.d.*

$$EX_i = \frac{1}{\theta}, \operatorname{Var} X_i = \frac{1}{\theta^2}$$

$$\log L_n(\theta) = n \log \theta - \theta \sum_{i=1}^{n} X_i$$

$$\frac{\partial}{\partial \theta}(\log L_n)(\theta) = \frac{n}{\theta} - \sum X_i \hookrightarrow \hat{\theta}^{\mathrm{MV}} = \frac{1}{\overline{X}}$$

$$\hookrightarrow I_{n(\theta)} = \mathrm{Var}\left(\frac{n}{\theta} - \sum X_i\right) = \mathrm{Var}\left(\sum X_i\right) = \frac{n}{\theta^2}$$

Remark 7.3 – $n\overline{X} \sim \Gamma(n, \theta)$ cf. TD1:

$$E\left[\frac{1}{n\overline{X}}\right] = \frac{\theta}{n-1} \qquad \mathrm{Var}\left(\frac{1}{\left(n\overline{X}\right)^2}\right) = \frac{\theta^2}{(n-1)(n-2)}$$

and ◇

$$E\left[\frac{1}{\overline{X}}\right] = n\frac{\theta}{n-1}$$

$$\hookrightarrow \tilde{\theta} = \frac{n-1}{n}\hat{\theta}$$

unbiased

$$\mathrm{Var}\left(\tilde{\theta}\right) = \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\left(\frac{1}{\overline{X}}\right) = \frac{(n-1)^2}{n^2}\left[E\left[\frac{1}{\left(\overline{X}\right)^2}\right] - \left(E\left[\frac{1}{\overline{X}}\right]\right)^2\right]$$

$$= \frac{\cancel{(n-1)^2}}{\cancel{n^2}} \times \frac{\cancel{n^2}\theta^2}{\cancel{(n-1)}(n-2)} - \frac{(n-1)^2}{n^2}\frac{n^2}{(n-1)^2}\theta^2$$

$$= \theta^2\frac{n-1}{n-2} - \theta^2 = \frac{\theta^2}{n-2} \underbrace{\geq}_{\mathrm{BCR}} \frac{1}{I_{n(\theta)}} \quad \text{not efficient}$$

$$\sqrt{n}\left(\overline{X} - \frac{1}{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

$\hat{\theta}$ is asymptotically efficient

$\overline{X}$ asymptotically normal (CLT). $g(x) = \frac{1}{x}$ on $]0, +\infty[$, $g'(x) = -\frac{1}{x^2} \neq 0$, delta method:

$$\sqrt{n}\left(\frac{1}{\overline{X}} - \theta\right) \xrightarrow{\mathcal{L}} \underbrace{g'\left(\frac{1}{\theta}\right)}_{=\theta^2}\mathcal{N}\left(0, \frac{1}{\theta^2}\right) = \mathcal{N}\left(0, \frac{\theta^4}{\theta^2} = \theta^2\right)$$

◇

7.2 Pivotal (asymptotic) or pivotal statistic

Definition 7.4 – *A statistic whose distribution does not depend on unknown parameters*

Example 7.5 – $X_1, ..., X_n$ *i.i.d.* Bernoulli(θ) *with* $\theta \in]0, 1[$*:*

$$\sqrt{n}\left(\overline{X}-\theta\right) \xrightarrow[\text{TLC}]{\mathcal{L}} \mathcal{N}(0,\theta(1-\theta))$$

$$\Leftrightarrow \underbrace{\sqrt{n}\frac{\overline{X}-\theta}{\sqrt{\theta(1-\theta)}}}_{\text{pivot ou stat. pivotale}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Pivotal method for confidence intervals: We estimate $\sqrt{\theta(1-\theta)}$ *by* $\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}$ *using the "plug-in" method with the continuous function* $g(x)=\sqrt{x(1-x)}$ *where* $x\in]0,1[$, $\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}$ *is a consistent estimator of* $\sqrt{\theta(1-\theta)}$

$$\sqrt{n}\frac{\hat{\theta}-\theta}{\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}} = \underbrace{\sqrt{n}\frac{\hat{\theta}-\theta}{\sqrt{\theta(1-\theta)}}}_{\xrightarrow[\text{TLC}]{\mathcal{L}}\mathcal{N}(0,1)} \times \underbrace{\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}}}_{\xrightarrow[\text{consistant}]{P} 1}$$

◇

Example 7.6 – $(X_1,...,X_n)$ *with density* $\theta>0$. $f_\theta(x)=\frac{3}{\theta}x^2\exp\left(-\frac{x^3}{\theta}\right)\mathbb{1}_{x\geq 0}$

. MLE?

$$\log L_n(\theta) = n(\log^3-\log\theta)+\sum_{i=1}^{n}\log(X_i^2)-\frac{1}{\theta}\sum_{i=1}^{n}X_i^3$$

$$(\log L_n)'(\theta) = -\frac{n}{\theta}+\frac{1}{\theta^2}\sum X_i^3 \Rightarrow \hat{\theta}=\frac{\sum X_i^3}{n}$$

$$(\log L_n)''(\theta)=\frac{n}{\theta^2}-\frac{2}{\theta^3}\sum X_i^3; (\log L_n)''\left(\hat{\theta}\right)=\frac{n}{\hat{\theta}^2}-\frac{2}{\hat{\theta}^3}n\hat{\theta}=-\frac{n}{\hat{\theta}^2}<0+\text{uniqueness}$$

$\Rightarrow$ *global maximum*

CLT:

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\theta^2)$$

$$\underbrace{\Leftrightarrow}_{\text{pivot asymptotique}} \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\theta}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1)$$

$$\underbrace{\Rightarrow}_{\text{Slutsky}} \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\hat{\theta}}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1)$$

$q_{\frac{\alpha}{2}}$ *and* $q_{1-\frac{\alpha}{2}}$ *quantiles of* $\mathcal{N}(0,1)$

$$P\left(q_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}\left(\hat{\theta}-\theta\right)}{\hat{\theta}} \leq 1_{1-\frac{\alpha}{2}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha$$

$$P\left(q_{\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}} \leq \hat{\theta}-\theta \leq q_{1-\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}}\right) \longrightarrow 1-\alpha$$

$$P\left(\underbrace{\hat{\theta}-q_{1-\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}} \leq \theta \leq \hat{\theta}-q_{\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}}}_{\Rightarrow IC(\theta) \text{ de niveau asymptotique } (1-\alpha)}\right) \longrightarrow 1-\alpha$$

◇

Estimation in Gaussian samples

§8

1. Normal distribution and derived distributions
2. Distribution of empirical estimators
3. CI of parameters
4. Exercise

8.1 TL;DR

$(X_1, ..., X_d)$ i.i.d. random variables following $\mathcal{N}(0,1)$ and $X \sim \mathcal{N}(0,1)$,

$$Y = X_1^2 + ... + X_d^2 \sim \chi^2(d)$$

$$\frac{X}{\sqrt{\frac{Y}{d}}} \sim \text{Student}(d)$$

8.2 Normal distribution and derived distributions

Definition 8.1 – *Z is said to be standard Gaussian (normal) if its distribution has the density function*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

We denote $Z \sim \mathcal{N}(0,1)$.

X is said to follow a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if and only if

$$X = \mu + \sigma Z$$

denoted

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Other characterizations of the normal distribution:

- by its density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- by the moment generating function

$$M(t) = E[e^{tX}] = e^{t\mu + \frac{1}{2}\sigma^2 t^2}, \forall t \in \mathbb{R}$$

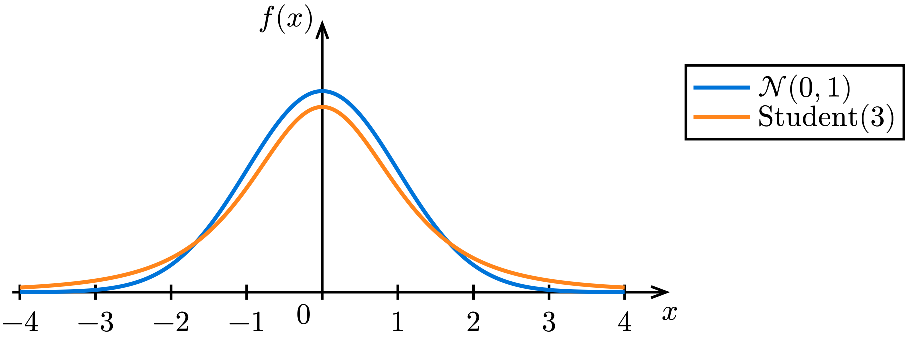
Remark 8.2 –

- *If $\sigma^2 = 0 \to X = \mu$ almost surely*

- *if* $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ *and* $\lambda \in \mathbb{R}$, *then* $\lambda X_1 + X_2 \sim \mathcal{N}(\lambda\mu_1 + \mu_2, \lambda^2\sigma_1^2 + \sigma_2^2)$

◇

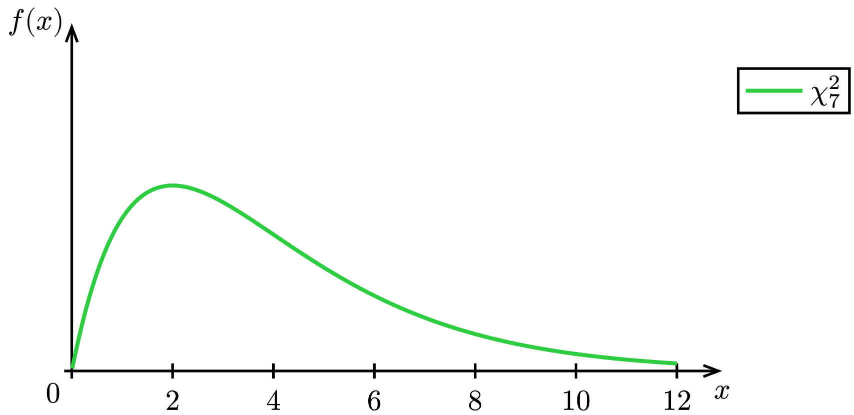
Central moments: symmetric density with respect to μ



centered moments: $E[(X-\mu)^k]$

- all odd-order centered moments are zero
- $\mu_{2k} = \frac{(2k)!}{2^k k!}\sigma^{2k}$
  - $E[(X-\mu)^4] = 3\sigma^4$
  - $\mathrm{Var}(X) = E[(X-\mu)^2] = \sigma^2$

Definition 8.3 – $(X_1, ..., X_d)$ *i.i.d. sample from* $\mathcal{N}(0,1)$*. The distribution of* $X_1^2 + X_2^2 + ... + X_d^2$ *is called the* χ^2 *(chi-squared) distribution with* d *degrees of freedom (df).*



Corollary 8.4 –

- *if* Y *follows a* $\chi^2(d)$ *distribution,* $E[Y] = d$, $\mathrm{Var}(Y) = 2d$

$$\mathrm{Var}(X_1^2 + ... + X_d^2) \underbrace{=}_{\text{indep}} d \quad \underbrace{\mathrm{Var}(X_i^2)}_{EX_i^h = E[X_i^2]^2 = 3-1=2}$$

- *support* $\mathbb{R}_+$
- $M(t) = (1-2t)^{-\frac{d}{2}}$, $\left(t < \frac{1}{2}\right)$

◇

Definition 8.5 – *if $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(d)$ are independent, the distribution of $Z = \frac{X}{\sqrt{\frac{Y}{d}}}$ is called Student's t-distribution with d df.*

Remark 8.6 – *if $d \to +\infty$, the Student distribution converges to the distribution $\mathcal{N}(0,1)$*

$$\frac{Y}{d} \underset{\text{def}}{\overset{\mathcal{L}}{=}} \frac{1}{d}\sum_{i=1}^{d} U_i^2 \quad \text{\#rmk[where } U_i \sim \mathcal{N}(0,1) \text{ are mutually independent from } X$$

$$\xrightarrow[\text{LGN}]{P} E(U_i^2) = 1$$

therefore (LAC)

$$g(x) = \sqrt{x}\frac{1}{\sqrt{\frac{Y}{d}}} \xrightarrow{P} 1$$

by Slutsky's Lemma $Z \xrightarrow{\mathcal{L}} 1 \cdot X \sim \mathcal{N}(0,1)$ ◇

We introduce $(X_1, ..., X_n)$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are unknown parameters.

- $\hookrightarrow \mu = E[X_i] \rightsquigarrow \hat{\mu} = \overline{X}$
- $\hookrightarrow \sigma^2 = \text{Var}(X_i) \rightsquigarrow \widehat{\sigma^2} = \frac{1}{n}\sum\left(X_i - \overline{X}\right)^2$

Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X - i - \overline{X}\right)^2$ <u>unbiased</u>

8.3 Empirical estimators law

Theorem 8.7 (law of $\hat{\mu}$ and $\hat{\sigma}^2$) –

- $\overline{X}$ *and* $\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ *are random variables <u>independent</u>*
- $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{1}{\sigma^2}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 \sim \chi^2(n-1) \Rightarrow \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$ *and* $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$
- $\frac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}} \sim \text{Student}(n-1)$
- $\overline{X}$ *and* $\overbrace{\left(\overline{X}, X_1 - \overline{X}, ..., X_n - \overline{X}\right)}^{T}$ *are independent*

Proof.

$$\begin{aligned}
M(u, t_1, ..., t_n) &= E\left[e^{u\overline{X}+t_1\left(X_1-\overline{X}\right)+...+t_n\left(X_n-\overline{X}\right)}\right] \\
&= E\left[e^{\left(\frac{u}{n}+\left(\underbrace{\frac{t_1+t_2+...+t_n}{n}}_{\bar{t}}\right)\right)X_1} ...e^{\left(\frac{u}{n}+\bar{t}\right)X_n}\right] \\
&= E\left[\prod_{i=1}^{n} e^{\left(\frac{u}{n}+t_i-\bar{t}\right)X_i}\right] \\
X_i \text{ indep.} &= \prod_{i=1}^{n} \underbrace{E\left[e^{\left(u_n+t_i-\bar{t}\right)X_i}\right]}_{M\left(u_n+t_i-\bar{t}\right)} \\
&= \prod_{i=1}^{n} e^{\mu\left(\frac{u}{n}+t_i-\bar{t}\right)+\frac{\sigma^2}{2}\left(u_n+t_i-\bar{t}\right)^2} \\
&= e^{\sum_{i=1}^{n} \mu\left(\frac{u}{n}+t_i-\bar{t}\right)+\frac{\sigma^2}{2}\left(u_n+t_i-\bar{t}\right)^2} \\
&= e^{\mu u+\mu\overbrace{\sum_i\left(t_i-\bar{t}\right)}^{0}+\frac{\sigma^2}{2}\sum_i\left(\frac{u^2}{n^2}+\left(t_i-\bar{t}\right)^2+2\frac{u}{n}\left(t_i-\bar{t}\right)\right)} \\
&= e^{\mu u+\frac{\sigma^2}{2}\left(\frac{u^2}{n}+\sum_i\left(t_i-\bar{t}\right)^2\right)} \\
&= \underbrace{e^{\mu u+\frac{\sigma^2 u^2}{2n}}}_{M_{\overline{X}}(u)} \underbrace{e^{\frac{\sigma^2}{2}\sum_i\left(t_i-\bar{t}\right)^2}}_{M_T(t_1,...,t_n)}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2+\frac{n}{\sigma^2}\left(\overline{X}-\mu\right)^2+\frac{2}{\sigma^2}\sum_i\left(X_i-\overline{X}\right)\left(\overline{X}-\mu\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2+\frac{n}{\sigma^2}\left(\overline{X}-\mu\right)^2+\frac{2}{\sigma^2}\left(\overline{X}-\mu\right)\underbrace{\sum_i\left(X_i-\overline{X}\right)}_{=0} \\
&= \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}_{=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)} + \underbrace{\frac{n}{\sigma^2}\left(\overline{X}-\mu\right)^2}_{=\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2\sim\chi^2(1)}
\end{aligned}$$

$$\text{by independence} \Rightarrow M_{\chi^2(n)}(t) = M_?(t)M_{\chi^2(1)}(t) \Rightarrow M_?(t) = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{(n-1)}{2}}$$

which characterizes the $\chi^2(n-1)$

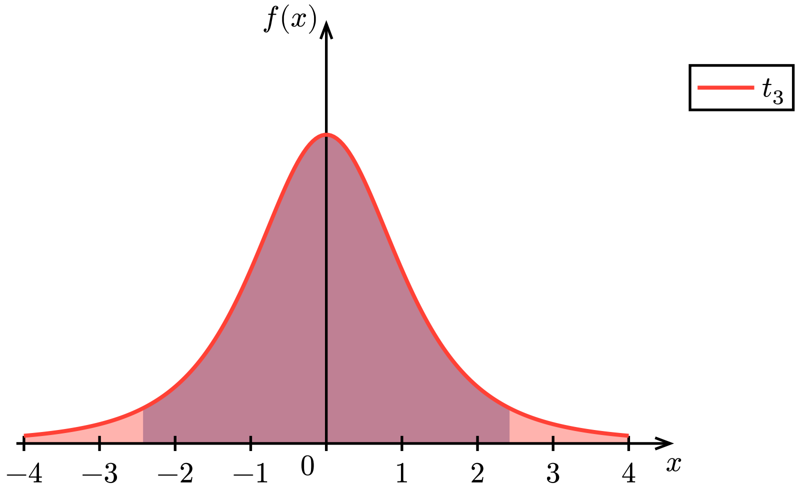
$$\frac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}} = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S_n}{\sqrt{n}}\times\frac{\sqrt{n}}{\sigma}} = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{S_n^2}{\sigma^2}}}$$

$$\frac{S_n^2}{\sigma^2} = \left(\frac{\sum\left(X_i-\overline{X}\right)^2}{\sigma^2}\right) \sim \chi^2(n-1)$$

therefore$\overline{X}+S_n^2$ are independent $\underbrace{\Rightarrow}_{\text{def Student}}$ Student$(n-1)$ □

8.4 IC of the parameters

Pivot. $\frac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}} \underset{\text{loi exacte}}{\sim}$ Student$(n-1)$



$$P\Big(q_{\frac{\alpha}{2}}t(n-1)\Big) \leq \frac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}} \leq q_{1-\frac{\alpha}{2}}t(n-1))$$

$$\Leftrightarrow P\left(\overline{X} - \frac{S_n}{\sqrt{n}}q_{1-\frac{\alpha}{2}}t(n-1) \leq \mu \leq \overline{X} + \frac{S_n}{\sqrt{n}}q_{1-\frac{\alpha}{2}}t(n-1)\right) = 1-\alpha$$

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

$$P\left(q_{\frac{\alpha}{2}}\chi^2(n-1) \leq \frac{n\hat{\sigma}^2}{\sigma^2} \leq q_{1-\frac{\alpha}{2}}\chi^2(n-1)\right) = 1-\alpha$$

$$= P\left(\frac{n\hat{\sigma}^2}{q_{1-\frac{\alpha}{2}}\chi^2(n-1)} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{q_{\frac{\alpha}{2}}\chi^2(n-1)}\right) = \alpha - 1$$

$\rightsquigarrow$ CI $= \left[\frac{n\hat{\sigma}^2}{q_{1-\frac{\alpha}{2}}\chi^2(n-1)}, \frac{n\hat{\sigma}^2}{q_{\frac{\alpha}{2}}\chi^2(n-1)}\right]$

Remark 8.8 – $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$ *and* $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$

$$\frac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}} \sim \text{Student}(n-1)$$

◇

8.5 Exercise

- Show that $\left(\hat{\mu}, \widehat{\sigma^2}\right)$ are the MLEs of μ and σ^2
- $R(S_n^2, \sigma^2) > R(\widehat{\sigma_n}^2, \sigma^2)$ where R represents a risk

Introduction to Statistical Tests

§9

9.1 Example

9.1.1 Quality control: industrial.

Produces "parts"

- ↳ of good quality
- ↳ defective

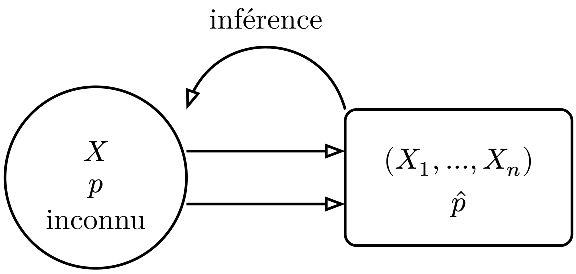
For the manufacturer, a proportion of 20% defective parts is assumed acceptable.

To control: randomly select n parts, verified ($p \leq 20\%$)

9.1.2 Modeling

$i^{\text{ème}}$ item $X_i = \begin{cases} 0 \text{ si bonne qualité} \\ 1 \text{ si défectueuse} \end{cases}$ $p = P(X_i = 1)$

↳ we sample n items and observe a sample $(X_1, ..., X_n)$ whose observed values are $(x_1, ..., x_n)$.



What is p? → we estimate

- empirical proportion
- $X_i \underset{\text{indep}}{\sim} \text{Bernouili}(p) \to \hat{p} = \overline{X}$

We observe $\overline{x} = 0.22$, $n = 100$

We proceed with a confidence interval for p. We define $\hat{p} = \overline{X}$, CLT:

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

we estimate the standard deviation by $\frac{\hat{p}(1-\hat{p})}{n}$ (consistent). Slutsky's Lemma:

$$\frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \times \frac{\underbrace{\overbrace{\sqrt{\frac{p(1-p)}{n}}}^{\text{LGN: } \overline{X} \xrightarrow{P} p \text{ , LAC: } g(x) = \sqrt{\frac{x(1-x)}{n}}}}_{\xrightarrow{P} 1}}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} \xrightarrow[n \to +\infty]{\mathcal{L}} 1.\mathcal{N}(0,1)$$

$$P\left(\mathrm{qnorm}_{\frac{\alpha}{2}} \leq \frac{\overline{X}-p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} \leq \mathrm{qnorm}_{1-\frac{\alpha}{2}}\right) \underset{n\to+\infty}{\longrightarrow} 1-\alpha$$

$$\Leftrightarrow P\left(\overline{X} - q_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \leq p \leq \overline{X} - q_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right) \underset{n}{\longrightarrow} 1-\alpha$$

$\mathrm{IC}(p) = \overline{X} \pm \mathrm{qnorm}_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}$ with an asymptotic level $1-\alpha$.

Ex: $\overline{x} = 0.22$, $\alpha = 5\%$, $n = 100$, $\mathrm{IC} = [0.14, 0.30]$

Question: is $p \leq 0.2$ or $p > 0.2$?

9.2 Principle of a test

$\Theta \subset]0,1[$. We want to test if $p \leq 0.2$ or $p > 0.2$.

$\Theta = \underbrace{\Theta_0}_{]0,0.2]} \cup \underbrace{\Theta_1}_{]0.2,1[}$ disjoint subsets.

We test H_0: $p \in \Theta_0$, $p \leq 0.2$ against H_1: $p \in \Theta_1$, $p > 0.2$

Conclusion:

- Either we retain H_0: ($p \leq 0.2$)
- Or we reject H_0 (we conclude $p > 0.2$)

Definition 9.1 – *A test of H_0 against H_1 is defined by the construction of a rejection region for H_0, $\mathcal{R}$*

- *if $(X_1, ..., X_n) \in \mathcal{R}$, we reject H_0 (in favor of H_1)*
- *if $(X_1, ..., X_n) \not\in \mathcal{R}$, we retain H_0*

Often $\mathcal{R} = \{(X_1, ..., X_n), T(X_1, ..., X_n) > c\}$

- T: test statistic (real-valued)
- c: test threshold

Remark 9.2 – *the decision of a test is random (depends on random T)* ◇

How to relate $\mathcal{R}$ to the tested hypotheses?

9.3 Error Risk

Definition 9.3 – *Type* $1^{\text{ère}}$ *error or Type I risk is the function defined on*

$$\begin{aligned}\Theta_0 &\longrightarrow [0,1]\\ p &\mapsto P_p((X_1, ..., X_n) \in \mathcal{R}) = P_p(\text{we reject } H_0)\end{aligned}$$

The test is said to be of level α if

$$\sup_{p\in\Theta_0} P_p(\text{rejection of } H_0) \leq a$$

R_q Type I error $= P(\text{rejection of } H_0 \text{ wrongly})$

reality / decision	H_0 true	H_1 true
H_0 true	ok	Type I error
H_1 true	Type II error	ok

Definition 9.4 – *The Type II error is the function defined on Type II risk*

$$\Theta_1 \longrightarrow [0,1]$$
$$\beta : p \mapsto P_p((X_1, ..., X_n) \not\in \mathcal{R}) = P_p(\text{we retain } H_0)$$

Remark 9.5 – *Type II error is* $P(\text{to retain } H_0 \text{ wrongly})$ ◇

power of the test: = 1 - Type II error

$$\prod : p \in \Theta_1 \longrightarrow P_p((X_1, ..., X_n) \in \mathcal{R})$$

Choice: the 2 errors cannot be minimized simultaneously. In general, α increases when β decreases.

Test: We choose to control the Type 1^{ere} error ($\Rightarrow$ the Type II error is generally unknown)

9.4 Construction of a Test

Principle: determine $\mathcal{R}$ such that the type I error $\leq \alpha$ (if we have multiple tests, we will choose (from a theoretical point of view) the one whose type II error is the smallest (or whose power is the greatest)). Based on an asymmetry between H_0 and H_1 in the construction.

Example 9.6 – $H_0 : p \leq 0.2$ *versus* $H_1 : p > 0.2$ *(*$\overline{x} = 0.22$*)*

- *p is unknown, so we estimate it as* $\hat{p} = \overline{X}$
- *Idea: under* H_1*,* $\hat{p}$ *takes larger values than under* H_0

↳ *$\mathcal{R}$ of the form $\hat{p} > c$ with c such that $P_p(\hat{p} > c) \leq \alpha$? (calculation? limit distribution of parameter p?)*

$$\hat{p} = \overline{X} \longrightarrow \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately distributed as $\mathcal{N}(0,1)$

$$P(\hat{p} > c) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \overbrace{\tilde{c}}^{=\left(\frac{c-p}{\sqrt{\frac{p(1-p)}{n}}}\right)}\right)$$

We want that

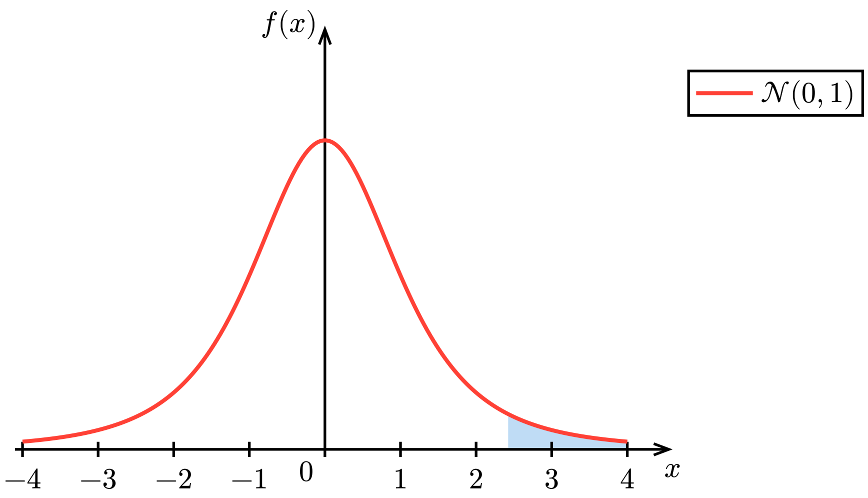
$$\sup_{\substack{p \in \Theta_0 \\ p \leq 0.2}} P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > c\right) \leq \alpha$$

↳ *the supremum is reached at* $p = 0.2$

$$\mathcal{R} = \left\{ (X_1, ..., X_n), \frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} > c \right\}$$

Find c *such that* $P((X_1, ..., X_n) \in \mathcal{R}) \underset{n \to +\infty}{\longrightarrow} \alpha$

$$P\left(\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} > c\right) \underset{n \to +\infty}{\longrightarrow} \alpha \text{ iff } c = \text{qnorm}_{1-\alpha}$$



- *rejection of* H_0 *iff*

$$T = \underbrace{\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}}}_{\substack{\text{statstique} \\ \text{de test}}} > \text{qnorm}_{1-\alpha}$$

Numerical Application: $\alpha = 5\%$, $\text{qnorm}_{1-\alpha} = 1.645$, $n = 100$, $\hat{p} = \overline{x} = 0.22$

↳

$$T = \frac{\overline{x} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{100}}} = \frac{0.02}{\sqrt{\frac{0.2(1-0.2)}{100}}} = \frac{0.2}{\sqrt{0.2 \cdot 0.8}} = \frac{0.2}{0.4} = \frac{1}{2} < 1.645$$

Conclusion: we retain H_0 *(the associated risk is unknown)*

$$\text{rejection } H_0 \Leftrightarrow \overline{x} > 0.2 + 1.645\sqrt{\frac{0.2(1-0.2)}{100}}$$
$$\Leftrightarrow \overline{x} > 0.266$$

◇

Hypothesis Tests (on a parameter)

§10

10.1 Test Formalism

10.1.1 Introduction

Definition 10.1 (statistical test) – *A hypothesis test is a (measurable) function of the sample $(X_1, ..., X_n)$ taking values in $\{0, 1\}$.*

- *H_0 is accepted if $\varphi(X_1, ..., X_n) = 0$*
- *H_0 is rejected if $\varphi(X_1, ..., X_n) = 1$*

The domain $\{(X_1, ..., X_n),\ \varphi(X_1, ..., X_n) = 1\} =: \mathcal{R}$ is the rejection region of the test, $\mathcal{R}^c$ is the acceptance region. We can write: $\varphi(X_1, ..., X_n) = \mathbb{1}_{\mathcal{R}}(X_1, ..., X_n)$

Very often, $\mathcal{R}$ is constructed from $T = T(X_1, ..., X_n)$, a test statistic Definition 10.1, itself based on an estimator $\hat{\theta}_n$ of θ, the parameter of interest.

↳ The question is: how to construct $\mathcal{R}$?

10.1.2 Risks of test error

Risk of 1$^{\text{ere}}$ kind.

Generally, we will test

$$H_0 :\ \theta = a_{a\in\Theta} \text{ against } H_1 : \theta \neq a$$
$$H_0 :\ \theta \leq a \text{ against } H_1 : \theta > a \text{ (e.g., quality control)}$$

If we consider a partition $\Theta_0 \cup \Theta_1 = \Theta_{\text{espace des paramètres}}$, $\Theta_0 \cap \Theta_1 = \emptyset$, then the hypotheses are: H_0: $\theta \in \Theta_0$ against H_1: $\theta \in \Theta_1$

Remark 10.2 (Vocabulary) –

- *Two-sided test*

  $\Theta_0 = \{a\}$ *H_0 is a simple hypothesis.*

  $\Theta_1 = \Theta \setminus \{a\}$, *$H_1$ is a two-sided hypothesis*
- *One-sided test*

  if $\Theta_0 =]-\infty, a]$ and $\Theta_1 =]a, +\infty[$ H_1 and H_0 are one-sided

  $H_0 = \theta = a$ against H_0: $\theta > a \longrightarrow$ One-sided test

◇

Definition 10.3 (error of 1$^{\text{ere}}$ kind) – *– the one we want to control*

$$\begin{aligned} \alpha : \Theta_0 &\longrightarrow [0,1] \\ \theta \mapsto &\quad P_\theta((X_1, ..., X_n) \in \mathcal{R}) = E_\theta[\varphi(X)] \\ &\quad = \underset{\text{vrai}}{P_{H_0}}(\text{rejection of } \ H_0) \end{aligned}$$

- *level α iff*

$$\sup_{\theta \in \Theta_0} P_\theta((X_1, ..., X_n) \in \mathcal{R}) \text{ op } \alpha$$

where for

$$\text{op} = \begin{cases} \leq \text{pour lois discrètes} \\ = \text{pour lois continues exactes} \\ \longrightarrow \text{pour lois asymptotiques} \end{cases}$$

Definition 10.4 (error of 2^nde kind) –

$$\begin{aligned} \beta : \Theta_1 &\longrightarrow [0,1] \\ \theta &\mapsto P_\theta((X_1, ..., X_n) \in \mathcal{R}^c) = \underset{\text{vrai}}{P_{H_1}}(\text{retaining } \ H_0) \end{aligned}$$

Definition 10.5 (Power functions) –

$$\begin{aligned} \Pi : \Theta &\longrightarrow [0,1] \\ \theta &\mapsto P_\theta((X_1, ..., X_n) \in \mathcal{R}) \end{aligned}$$

- *if $\theta \in \Theta_0$: $\Pi(\theta) = \alpha(\theta)$*
- *if $\theta \in \Theta_1$: $\Pi(\theta) = P_\theta((X_1, ..., X_n) \in \mathcal{R}) = 1 - P_{H_1}((X_1, ..., X_n) \in \mathcal{R}^c) = 1 - \beta(\theta)$*

10.2 Example

$(X_1, ..., X_n)$ are i.i.d. from distribution $\mathcal{N}(\theta, 1)$. Hypotheses to test:

$$H_0 : \ \theta \leq 0 \ \text{ versus } \ H_1 : \ \theta > 0$$

Since $E[X_i] =: \theta$ is unknown, we estimate it with $\hat{\theta} = \overline{X}$.

1. First idea: rejection of H_0 if $\hat{\theta} > 0$

$$\mathcal{R} = \left\{(X_1, ..., X_n), \hat{\theta}(X_1, ..., X_n) > 0\right\}$$

Let $\theta \leq 0$,

$$\alpha(\theta) = P_\theta\left(\hat{\theta} > 0\right) = P_\theta\left(\overline{X} > 0\right)$$

What is the distribution of $\overline{X}$?

$\overline{X} \underset{\substack{\text{loi} \\ \text{exacte}}}{\sim} \mathcal{N}()$ because any linear combination of Gaussian random variables is a Gaussian.

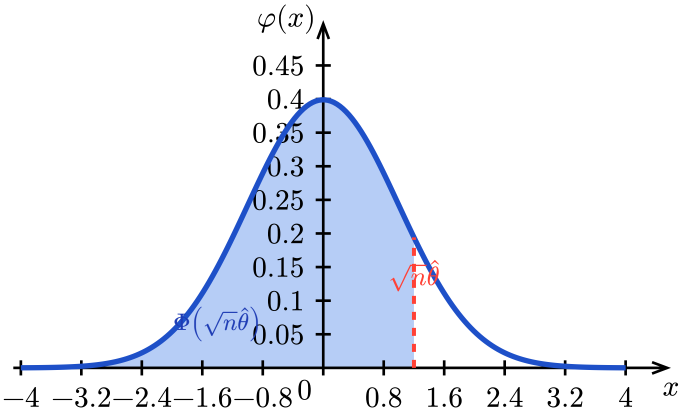
$E\left[\overline{X}\right] = E[X_i] = \theta$, $\text{Var}\left(\overline{X}\right) = \frac{1}{n}\text{Var}(X_i) = \frac{1}{n}$

reflex: standardize the normal distribution:

$$\alpha(\theta) = P_\theta\left(\frac{\overline{X}-\theta}{\sqrt{\frac{1}{n}}} > -\sqrt{n}\theta\right) = P\big(\mathcal{N}(0,1) > -\sqrt{n}\theta\big) = 1-\Phi\big(-\sqrt{n}\theta\big) = \Phi\big(\sqrt{n}\theta\big)$$

Where Φ is the cumulative distribution function of the $\mathcal{N}(0,1)$

distribution.



$$\text{level} = \sup_{\theta\leq 0}\Phi\big(\sqrt{n}\theta\big) = \Phi(0) = \frac{1}{2} = 50\%$$

Therefore, we have a 1 in 2 chance of being wrong – which is not acceptable!

$\longrightarrow$ we want α to be small: $\alpha = 5\%$:

- $\mathcal{R} = \left\{\hat{\theta} > 0\right\} \longrightarrow \mathcal{R} = \left\{\hat{\theta} > c\right\}$ $(c > 0)$
- value of $c = c(\alpha)$ such that $\sup_{\theta\leq 0}\alpha(\theta) \leq \alpha$

$$\begin{aligned}\alpha(\theta) = P_{\theta\leq 0}\big(\overline{X} > c\big) &= P_{\theta\leq 0}\left(\frac{\overline{X}-\theta}{\sqrt{\frac{1}{n}}} > \frac{c-\theta}{\sqrt{\frac{1}{n}}}\right) \\ &= P\big(\mathcal{N}(0,1) > \sqrt{n}(c-\theta)\big)\end{aligned}$$

<u>Level Condition</u>:

Find c such that

$$\begin{aligned}&\sup_{\theta\leq 0} P\big(\mathcal{N}(0,1) > \sqrt{n}(c-\theta)\big) \underset{\substack{\text{loi}\\\text{continue}}}{=} \alpha \\ &\Leftrightarrow P\big(\mathcal{N}(0,1) > \sqrt{n}c\big) = \alpha \\ &\Leftrightarrow 1-\Phi\big(\sqrt{n}c\big) \\ &\Leftrightarrow \Phi\big(\sqrt{n}c\big) = 1-\alpha \\ &\Leftrightarrow \sqrt{n}c = \Phi^{-1}(1-\alpha) \Rightarrow c_\alpha = \frac{1}{\sqrt{n}}\,\text{qnorm}_{1-\alpha}\end{aligned}$$

We constructed a test of level α with

$$\mathcal{R} = \left\{(X_1,...,X_n), \overline{X} > \frac{\text{qnorm}_{1-\alpha}}{\sqrt{n}}\right\}$$

Numerical application: $\alpha = 5\% \Rightarrow \text{qnorm}_{1-\alpha} = 1.645, n = 100 \to c_\alpha = 0.1645$

experiment $\to \overline{X}^{\text{obs}}$ = realization of $\overline{X}$ on my data

- if $\overline{X}^{\text{obs}} = 0.1 < c_\alpha$ we do not reject H_0
- if $\overline{X}^{\text{obs}} = 0.3 > c_\alpha \Rightarrow$ rejection of H_0

10.3 Construction of a test

1. - Define the hypotheses H_0 and H_1
   - Identify the parameter of interest
2. - Define the form of $\mathcal{R}$; the form of $H_1 \Rightarrow$ implies the form of $\mathcal{R} = \{T > c\}$ or $\{T < c\}$
   - Find a test statistic
   - T = normalized version of

$$\hat{\theta} = \frac{\hat{\theta} - \theta}{\sqrt{\operatorname{Var}\left(\hat{\theta}\right)}}$$

3. Find the threshold c to obtain a level α

Tests of a Gaussian parameter

§11

11.1 Summary of construction

$X = (X_1, ..., X_n)$ i.i.d. with law P_θ

1. Specify the hypotheses tested:

$$H_0\colon \theta \leq \theta_0 \text{ against } H_1\colon \theta > \theta_0$$

2. Test statistic: $T(X)$: under H_0 $T(X)$ is computable.

The distribution of T under H_0 allows distinguishing between H_0 and H_1.

↳ $\mathcal{R} = \{T(X) > c\}$ (under H_1, if the distribution of ... of T deviates from H_0 to the right); if $H_1\colon \theta < \theta_0 \to \mathcal{R} = \{T(X) < c\}$, if two-tailed test $H_1 : \theta \neq \theta_0 \to \mathcal{R} = \{|T(X)| > c\} = \{T(X) > c \text{ or } T(X) < -c\})$

3. Decision rule

α fixed level,

- Level condition:

$$\begin{aligned}\sup_{\theta\leq\theta_0} P_{H_0}(T(X) > c) &= \alpha \text{ (if the distribution of } T \text{ under } H_0 \text{ is continuous)}\\ &\leq \alpha \text{ (if the distribution of } T \text{ is discrete)}\\ &\underset{n\to+\infty}{\longrightarrow} \text{ (if the distribution of } T \text{ is asymptotic)}\end{aligned}$$

4. Numerical application:

- calculation of the threshold
- calculation of the realization of $T = T^{\text{obs}} = T(X)$ if $x = (x_1, ..., x_n)$

is a realization of $(X_1, ..., X_n)$ in our experiment

- if $T^{\text{obs}} > c_\alpha$ then we reject H_0, with a risk of being wrong of α.
- if $T^{\text{obs}} \leq c_\alpha$, we retain H_0, with an unknown risk of being wrong unknown (in general)

Remark 11.1 – *The test of $H_0\colon \theta \geq \theta_0$ against $H_1\colon \theta < \theta_0$ is the same as the test of $H_0\colon \theta = \theta_0$ against $H_1\colon \theta < \theta_0$, $\mathcal{R} = \{T < c\}$* ◇

11.2 P-value

Example 11.2 – *$(X_1, ..., X_n)$ i.i.d. following $\mathcal{N}(0, 1)$; test $H_0\colon \theta = 0$, against $H_1\colon \theta > 1$*

$\hookrightarrow$ $T = \frac{\hat{\theta}-0}{\frac{1}{\sqrt{n}}} = \sqrt{n}\hat{\theta} \underset{H_0}{\sim} \mathcal{N}(0,1)$ *where* $\hat{\theta} = \overline{X}$; $\mathcal{R} = \{T > c\}$ *with the level condition*

$P_{\theta=0}\left(\underbrace{T}_{\to\sim\mathcal{N}(0,1)} > c\right) = \alpha \Rightarrow c_\alpha = \text{qnorm}_{1-\alpha} = \Phi^{-1}(1-\alpha) \Rightarrow \mathcal{R} = \left\{\sqrt{n}\hat{\theta} > \Phi^{-1}(1-\alpha)\right\}$

$H_0 \Leftrightarrow \alpha > 1 - \Phi\left(\sqrt{n}\hat{\theta}\right)$

$\alpha = 5\%$ 5%? 10%? 1%?

Numerically: $\hat{\theta} = 0.3$, $n = 100$, $\sqrt{n}\hat{\theta} = 3$, $1 - \Phi(3) \approx 10^{-3}$ ◇

Definition 11.3 – *If* $(X_1, ..., X_n)$ *i.i.d.,* $\mathcal{R} = \{T(X) > c_\alpha\}$*. For a realization* $x = (x_1, ..., x_n)$ *of* $X = (X_1, ..., X_n)$*, we call it the p-value of the test with region* $\mathcal{R}$*:*

$$\begin{aligned}\text{pval} &= \inf\{\alpha \in [0,1], T(X) > c_\alpha\} \\ &= \inf\{\alpha, H_0 \text{ is rejected on level } \alpha\}\end{aligned}$$

p-value $(\varphi(X) = \mathbb{1}_{\mathcal{R}}(X))$ *- significance level, critical probability*

Example 11.4 –

$$\begin{aligned}\text{pval} &= 1 - \Phi\left(\sqrt{n}\hat{\theta}^{\text{obs}}\right) \\ &= 1 - \Phi\left(T^{\text{obs}}\right) \\ &= 1 - P\left(\mathcal{N}(0,1) \leq T^{\text{obs}}\right)\end{aligned}$$

$$\begin{aligned}\text{pval} &= P\left(\underbrace{\mathcal{N}(0,1)}_{\text{loi de } T} > T^{\text{obs}}\right) \\ &= P\left(T > \underbrace{T^{\text{obs}}}_{\in\mathbb{R}}\right)\end{aligned}$$

◇

11.2.1 Generalization (formula for calculating a p-value).

$T(X)$ test statistic

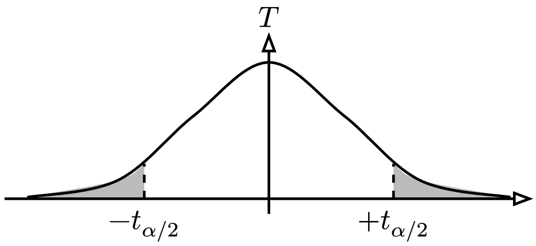
- $\mathcal{R} = \{T(X) > c\}$, then the p-value $= P_{H_0}\left(T(X) > T^{\text{obs}}\right)$
- $\mathcal{R} = \{T(X) < c\}$, then the p-value $= P_{H_0}\left(T(X) < T^{\text{obs}}\right)$
- $\mathcal{R} = \{|T(X)| > c\}$, the p-value $= P_{H_0}\left(|T(X)| > |T^{\text{obs}}|\right)$

11.2.2 Remarks

Remark 11.5 – *In the example*

$$\begin{aligned}P_{H_0}\left(|T| > T^{\text{obs}}\right) &= P_{H_0}\left(T > T^{\text{obs}} \text{ or } T < -T^{\text{obs}}\right) \\ &= P_{H_0}\left(T > T^{\text{obs}}\right) + P\left(T < -T^{\text{obs}}\right) \\ &= 1 - \Phi\left(T^{\text{obs}}\right) + \Phi\left(-T^{\text{obs}}\right) \\ \text{by symmetry} &= 2\left(1 - \Phi\left(T^{\text{obs}}\right)\right)\end{aligned}$$

the p-value of the two-tailed test is double the p-value of the one-tailed test.

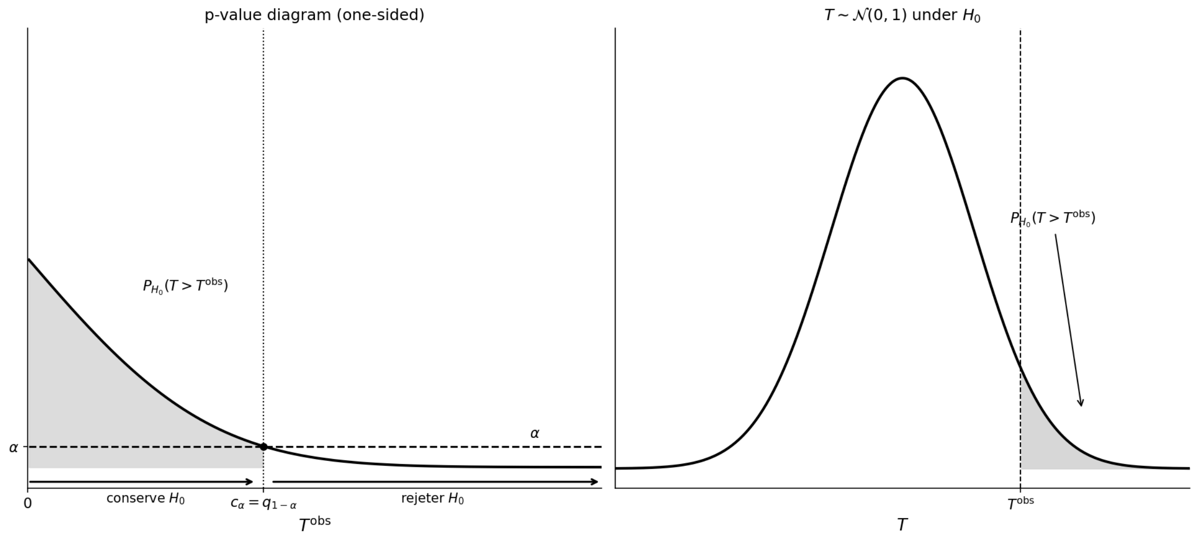


◇

Remark 11.6 – *If the distribution of T under H_0 is* <u>*discrete*</u>*.* ◇

11.2.3 Decision rule using the p-value

Example 11.7 – $\theta = 0,$ *against* $\theta > 0,$ $T = \sqrt{n}\hat{\theta} \sim \mathcal{N}(0,1)$



◇

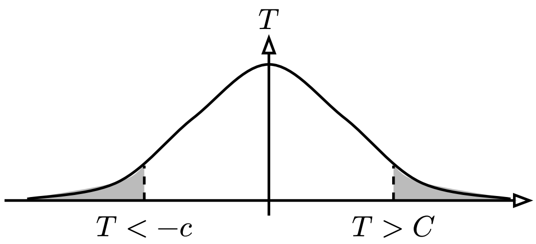
1. H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$, $\hat{\mu} = \overline{X}$

$$T = \frac{\overline{X} - \mu_0}{\frac{S_n}{\sqrt{n}}} \underset{H_0}{\overset{\text{loi exacte}}{\sim}} \text{Student}(n-1)$$

$H_1 \underset{\text{bilatère}}{\Rightarrow} \mathcal{R} = \{|T| > c\}$, what is the decision rule?

↳ calculation of $c = c_\alpha$ with the level condition $\Rightarrow c_\alpha = \text{qt}_{1-\frac{\alpha}{2}}(n-1)$ quantile Student$(n-1)$

$$\begin{aligned} \text{pvaleur} &= P_{H_0}\big(|T| > |T^{\text{obs}}|\big) \\ \text{(symmetry of Student's law)} &= 2P\big(T > |T^{\text{obs}}|\big) \\ &= 2\big(1 - F_{\text{Student}}\big(|T^{\text{obs}}|\big)\big) \end{aligned}$$



- If the p-value $< \alpha$, we reject H_0- If the p-value $\geq \alpha$ - we retain H_0

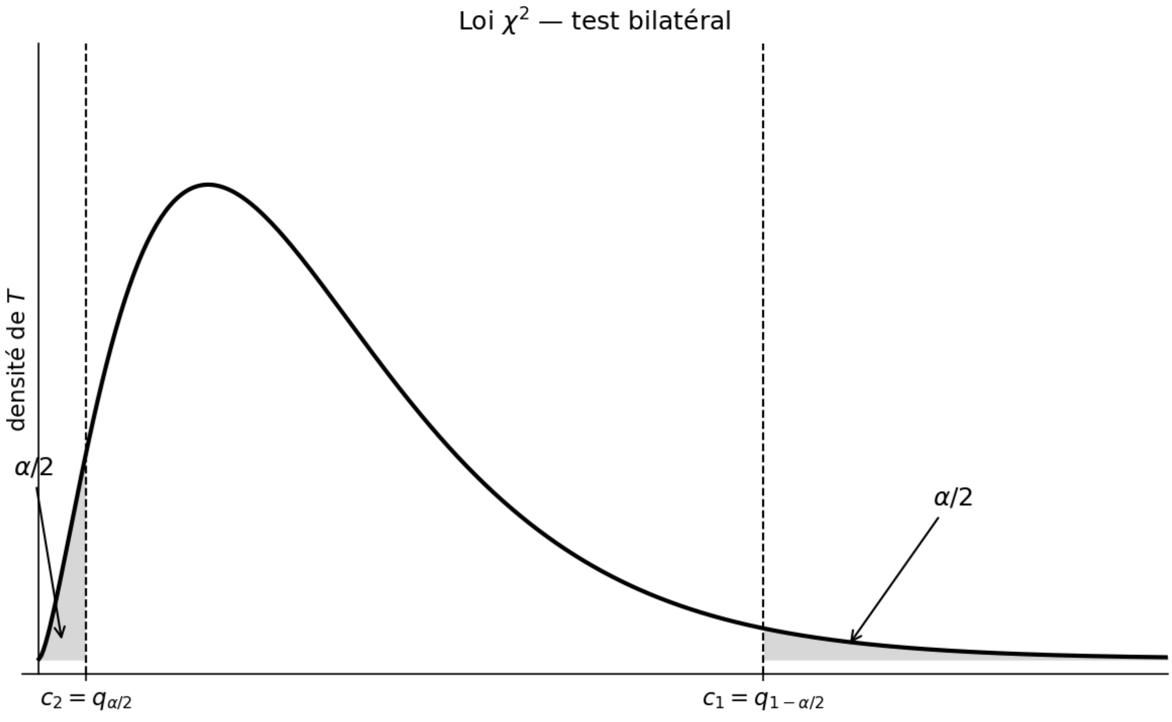
2. H_0: $\sigma^2 = \sigma_0^2$ against H_1: $\sigma^2 \neq \sigma_0^2$. Since σ^2 is unknown, we estimate it:

- S_n^2 unbiased
- $\widehat{\sigma^2}$ MLE

by the theorem of the estimators' distribution in the Gaussian model

$$0 \leq T = \frac{n\widehat{\sigma^2}}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_i^n \left(X_i - \overline{X}\right)^2}{\sigma^2} \underset{H_0}{\sim} \chi^2(n-1)$$

$$\mathcal{R} = \left\{T > q_{1-\frac{\alpha}{2}}\chi^2(n-1) \ \text{ or } \ T < q_{\frac{\alpha}{2}}\chi^2(n-1)\right\}$$



We calculate T^{obs}: we reject H_0 iff $T_{\text{obs}} > q_{1-\frac{\alpha}{2}}\chi^2$ or $T_{\text{obs}} < q\chi^2_{\frac{\alpha}{2}}$

What is the p-value?

$$\text{pvaleur} \underbrace{=}_{\text{par convention}} 2P\big(T > T^{\text{obs}}\big) \ \text{ or } \ 2P\big(T < T^{\text{obs}}\big)$$

p-value $< 5\%$