



* Evaluation: 0.4CC + 0.6 examen
 ↑ 80% partiel, 20% Interro 26/01

1. Modèle Statistique

DEF: un modèle statistique est un espace de proba $(\Omega, \mathcal{F}, \mathbb{P})$ où \mathbb{P} est une famille de loi de proba $\{P_\theta; \theta \in \Theta\}$
 si $\exists p \in \mathbb{N}^*$, $\Theta \subset \mathbb{R}^p$: modèle paramétrique

Sinon: modèle non paramétrique

[exemples de familles: $\mathcal{P} = \{IP(\lambda); \lambda > 0\}$ les lois de poisson

$\mathcal{P} = \{IP \text{ loi de probas}\} = \{IP; IP \text{ dont la densité admet une dérivée seconde bornée}\}$

DEF: une observation est une v.a. dont la loi $\in \{P_\theta, \theta \in \Theta\}$

Notre observation aura une structure de n-échantillons X_1, \dots, X_n des v.a. indep (réelles) identiquement distribuées, de loi commune $\in \{P_\theta, \theta \in \Theta\}$

Rq: (X_1, \dots, X_n) de loi $P_\theta^{\otimes n}$

↳ l'échantillon contient toute l'info sur P_θ , donc sur θ .

DEF: un modèle est identifiable ssi $\theta \rightarrow P_\theta$ est injective

Rq: observation $\begin{cases} \rightarrow X_i \text{ v.a.} \\ \rightarrow x_i = \text{donnée} \rightsquigarrow \text{réalisation de } X_i \end{cases}$

2. Estimateurs

Hyp: On observe X_1, \dots, X_n i.i.d. de loi commune $\in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ modèle paramétrique identifiable
 ($\exists \theta^*$ inconnu tq $P_{X_i} = P_{\theta^*}$)

DEF: [estimateur] un estimateur de θ est une fonction de l'échantillon (X_1, \dots, X_n) (mesurable) indep de θ
 (calculable à partir des données)

Notation: $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$. ($\hat{\theta}_n$ est une v.a.)

E.g: $\hat{\theta} = 0, \hat{\theta} = X_1 - X_3 \dots$

↳ 1) Comment définir un bon estimateur?

2) Comment construire un bon estimateur?

3. Risque quadratique

Idée: En moyenne, $\hat{\theta}$ doit être proche de $\theta \rightarrow \mathbb{E}(\hat{\theta} - \theta)$

DEF: [biais de $\hat{\theta}$]

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

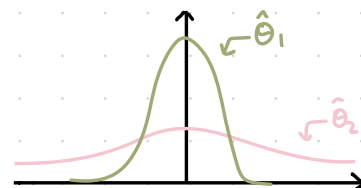
On dit que $\hat{\theta}$ est sans biais si $B(\hat{\theta}, \theta) = 0$

DEF: [Risque quadratique]

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

MSE (mean squared error)

On dit que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ ssi $R(\hat{\theta}_1, \theta) < R(\hat{\theta}_2, \theta)$



Exemple: [modèle de Poisson]

X_1, \dots, X_n i.i.d de loi P_θ de Poisson, $\theta > 0$

Estimateur de θ ? $\theta = \mathbb{E}[X_i]$
 $\rightarrow \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$

$$B(\hat{\theta}, \theta) = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] - \theta$$

linéarité

$$\downarrow = \frac{1}{n} \sum_i \mathbb{E}[X_i] - \theta$$

$$\stackrel{\text{i.i.d}}{=} \frac{1}{n} \cdot n \mathbb{E}[X_1] - \theta$$

$$= \mathbb{E}[X_1] - \theta = \theta - \theta = 0$$

$$\leadsto \mathbb{E}[\bar{X}] = \theta$$

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

$$= \text{Var}(\bar{X})$$

$$= \text{Var}\left(\frac{1}{n} \sum_i X_i\right)$$

$$= \frac{1}{n^2} \text{Var}\left(\sum_i X_i\right)$$

$$\stackrel{\text{indep}}{=} \frac{1}{n^2} \sum_i \text{Var}(X_i) \stackrel{\text{i.i.d}}{=} \frac{1}{n} \cdot \text{Var}(X_1)$$

PROP: [décomposition biais-variance durisque]

$$R(\hat{\theta}, \theta) = (\text{Biais}(\hat{\theta}, \theta))^2 + \text{Var}(\hat{\theta})$$

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

$$\stackrel{\text{def}}{=} \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2]$$

$$= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \theta)^2] + 2\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)]$$

$$= \text{Var}(\hat{\theta}) + (\text{B}(\hat{\theta}, \theta))^2 + 2\text{B}(\hat{\theta}, \theta)(\mathbb{E}[\hat{\theta}] - \mathbb{E}[\hat{\theta}])$$

4. Consistance

↳ propriété asymptotique

on ne considère que des constants.

DEF: (X_1, \dots, X_n) i.i.d de loi P_θ , $\theta \in \Theta \subset \mathbb{R}^p$

Soit $\hat{\theta}_n = h(X_1, \dots, X_n)$. On considère la suite de v.a. $(\hat{\theta}_n)_{n \geq 1}$.

$\hat{\theta}_n$ est un estimateur consistant (ou convergent) de θ ssi : $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta$

Rq: $(\hat{\theta}_n)$ est fortement consistant ssi : $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \theta$

Exemple [modèle de poisson]

$$\theta = \mathbb{E}[X_i]; \hat{\theta}_n = \bar{X}$$

* On peut invoquer la LGN : $\bar{X} \xrightarrow{\mathbb{P}} \theta = \mathbb{E}[X_i]$

$$* R(\hat{\theta}, \theta) = \text{Var}(\bar{X}) = \frac{\theta}{n} \xrightarrow[n \rightarrow +\infty]{} 0$$

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

$$\text{Ineq. BT: } \forall \varepsilon > 0, \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{\mathbb{E}|\hat{\theta} - \theta|^2}{\varepsilon^2} = \frac{R(\hat{\theta}, \theta)}{\varepsilon^2} \xrightarrow{} 0$$

• (X_1, \dots, X_n) i.i.d. Poisson(θ).

On veut estimer $\beta = \mathbb{P}(X_i = 0) = e^{-\theta}$

↳ $\hat{\beta} = e^{-\hat{\theta}} = e^{-\bar{X}}$ [méthode "plug-in"]

$\hat{\beta}$ est consistant pour estimer β .

LEMME [de l'application continue]

Si $Z_n \xrightarrow{\mathbb{P}} Z$, alors $h(Z_n) \xrightarrow{\mathbb{P}} h(Z)$ $\forall h$ fonction continue

Resume Cours 1:

Cadre paramétrique

Modèle statistique paramétrique:

On dispose d'une observation (X_1, \dots, X_n) , un échantillon de v.a. i.i.d. de loi commune P appartenant à une famille de lois de probabilités paramétrée $\{P_\theta, \theta \in \mathbb{M} \subset \mathbb{R}^p\}$

Rq: si $\mathbb{M} \subset$ espace de dim infinie \rightarrow modèle non-paramétrique

Question: estimer P , c'est estimer $\theta \in \mathbb{R}^p$

Exemples: Bernoulli (θ) , Exponentielle (θ) , gaussienne (μ, σ^2) , lois de densité $f_\theta(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$

Modèle identifiable: $\theta \rightarrow P_\theta$ injective

Notation: $\mathbb{E}_\theta[h(X_1, \dots, X_n)]$, $\text{Var}_\theta[h(X_1, \dots, X_n)]$.

loi de $(X_1, \dots, X_n) \rightarrow P_\theta^{\otimes n}$

Estimateur: $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$

Qualité: • risque: $R(\hat{\theta}, \theta) = \mathbb{E}((\hat{\theta} - \theta)^2)$

• consistance: $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta$

1 Méthode des moments

a) Définitions

On appelle moment théorique de la loi des X_i d'ordre k : $\mu_k = \mathbb{E}(X_i^k)$, $k \geq 1$

// moment empirique de la loi des X_i d'ordre k : $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$

Par la LGN, $\hat{\mu}_k \xrightarrow[n \rightarrow +\infty]{P} \mu_k$

La méthode des moments: si on peut écrire θ ou $g(\theta)$ paramètre d'intérêt comme une fonction des K premiers moments théorique: $\theta = \psi(\mu_1, \mu_2, \dots, \mu_K)$. Alors l'estimateur $\hat{\theta} = \psi(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_K)$ est obtenu par la méthode des moments.

b) Exemples:

• $X_i \sim$ Bernoulli (θ) à valeurs $0-1$, $\theta = P(X_i=1) = \mathbb{E}[X_i]$

$$\hookrightarrow \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

• $X_i \sim$ Exp (θ) , $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}_{x \geq 0}$. $\mathbb{E}[X_i] = 1/\theta \Leftrightarrow \theta = \frac{1}{\mu_1}$
par la méthode des moments, $\hat{\theta}^{MM} = \frac{1}{\hat{\mu}_1} = \frac{1}{\bar{X}}$

$$\text{rq: } \text{Var}(X_i) = \frac{1}{\theta^2} \Leftrightarrow \theta^2 = \frac{1}{\mathbb{E}X_i^2 - (\mathbb{E}X_i)^2}$$

$$\Leftrightarrow \theta = \frac{1}{\sqrt{\mu_2 - \mu_1^2}} \Rightarrow \hat{\theta}_2^{MM} = \frac{1}{\sqrt{\frac{1}{n} \sum X_i^2 - (\bar{X})^2}}$$

• X_1, \dots, X_n i.i.d. de loi f_θ de densité $f_\theta(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$, $\theta > 0$

$$\mathbb{E}_\theta[X_i] = \theta \int_0^1 x^\theta dx = \frac{\theta}{\theta+1}$$

$$\text{Méthode des moments } (\theta+1)\mu_1 = \theta \Leftrightarrow \theta(1-\mu_1) = \mu_1 \Leftrightarrow \theta = \frac{\mu_1}{1-\mu_1} = \frac{\mathbb{E}(X_i)}{1-\mathbb{E}(X_i)}$$

$$\Rightarrow \hat{\theta}_{MM} = \frac{\bar{X}}{1-\bar{X}}, \quad P_\theta(\bar{X}=1) = 0$$

↑ pas nécessairement consistant (+LGN) ^{hyp: moments finis + indépendance}

c) retour sur le LAC:

$(X_n)_{n \geq 1}$ suite de v.a. r. Si X_n converge vers X , que peut-on dire de $(g(X_n))_{n \geq 1}$?

si g continue si $X_n \xrightarrow{P} X$, alors $g(X_n) \xrightarrow{P} g(X)$

si $X_n \xrightarrow{d} X$, alors $g(X_n) \xrightarrow{d} g(X)$

rg: [condition suffisante]

Dg = { points de discontinuité de g }. si $P(X \in Dg) = 0$, le LAC est vrai

ex: $g(x) = \frac{x}{1-x}$ LGN: $\bar{X} \xrightarrow{\mathbb{P}} \mathbb{E}[X]$ étape à faire avant LAC
 LAC: $g(\bar{X}) \xrightarrow{\mathbb{P}} g(\mathbb{E}[X]) = \theta$
 $= \hat{\theta}_n$

LAC pour des couples de suites de v.a.:
 si $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$,
 • si $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ ou \mathbb{R}^2 continue, alors $g(X_n, Y_n) \xrightarrow{\mathbb{P}} g(X, Y)$
 • si $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, Y)$

(← exemple) $\hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n} \sum X_i^2 - (\bar{X})^2}}$ constant. 1?

LG: $\bar{X} \xrightarrow{\mathbb{P}} \mu_1$, $\frac{1}{n} \sum X_i^2 \xrightarrow{\mathbb{P}} \mu_2$] donc $\begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum X_i^2 \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $g(x, y) = \frac{1}{\sqrt{y-x^2}}$
 faux pour la convergence en loi continue sauf en $\{(x, y) \in \mathbb{R}^2, y=x^2\}$ de mesure nulle
 LAC $\Rightarrow \hat{\theta}_2 \xrightarrow{MM}$ constant de θ

PROP. $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} X \\ Y \end{pmatrix}$ ssi $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$

pv: (\Rightarrow) alors LAC $g(x, y) = x$ continue donc $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$
 (\Leftarrow) $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| + |Y_n - Y| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) + \mathbb{P}(|Y_n - Y| > \frac{\varepsilon}{2}) \rightarrow 0$
 (sauf cas indépendance)

Cette réciproque est fautive sur la convergence en loi !!

d) Variance Empirique

Si les X_i admettent une espérance μ et une variance σ^2 ,

on appelle variance empirique: $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ pas hypothèse d'indépendance ~ on ne peut pas appliquer LGN
 $= \frac{1}{n} \sum X_i^2 + \frac{1}{n} \sum X_i \bar{X}^2 - \frac{2}{n} \sum X_i \bar{X}$
 $= \frac{1}{n} \sum X_i^2 + \bar{X}^2 - 2\bar{X} \bar{X} = \tilde{\sigma}^2$

(Estimateur des moments: $\sigma^2 = \mathbb{E}X_i^2 - (\mathbb{E}X_i)^2$).

On remplace les moments théoriques par les moments empiriques

$\hookrightarrow \hat{\sigma}_n^{MM} = \frac{1}{n} \sum X_i^2 - (\bar{X})^2$

• Constant: $\hat{\sigma}^2 = \frac{1}{n} \sum X_i^2 - (\bar{X})^2$ LAC, $g(x, y) = y - x^2$ continue $\mathbb{R}^2 \rightarrow \mathbb{R}$
 $\begin{pmatrix} \bar{X} \xrightarrow{\mathbb{P}} \mathbb{E}X \\ \frac{1}{n} \sum X_i^2 \xrightarrow{\mathbb{P}} \mathbb{E}X^2 \end{pmatrix} \Rightarrow \begin{pmatrix} \bar{X} \\ \frac{1}{n} \sum X_i^2 \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} \mathbb{E}X \\ \mathbb{E}X^2 \end{pmatrix} \rightarrow \hat{\sigma}^2$ constant de $\text{Var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

(exercice: calculer le biais de $\hat{\sigma}_n^2$ + le risque de $\hat{\sigma}_n^2$)

2. Méthode du maximum de vraisemblance

a) Modèle dominé

$(P_\theta)_{\theta \in \Theta}$ est dominé s'il existe une mesure μ (positive σ -finie)

$\hookrightarrow X_i$ à valeurs dans E , $E = \cup E_n$ avec $\mu(E_n)$ fini

telle que $\forall \theta, P_\theta$ admet une densité par rapport à μ

En pratique:

• soit E au plus dénombrable: $\mu =$ mesure de comptage

si $\exists \{a_1, a_2, a_3, \dots\}$ tq $\sum_{k \geq 1} P_\theta(X_i = a_k) = 1$, alors $\mu = \sum_{k \geq 1} \delta_{a_k}$ avec $\delta_a(\{a\}) = 1$ mesure de Dirac.

ex: Bernoulli (θ) , $X_i = 1$ proba $\theta \rightarrow \mu = \delta_0 + \delta_1$
 $0 \text{ // } 1 - \theta$

On écrit $f_\theta(x) = P_\theta(\{x\}) = P_\theta(X_i = x)$

($x \in \{a_1, a_2, \dots\}$)

Soit $E = \mathbb{R}^p$, alors f_θ est la densité usuelle

b) Définitions

$\hookrightarrow f_\theta$ densité de P_θ

On appelle vraisemblance de l'échantillon (X_1, \dots, X_n) la fonction $\theta \rightarrow L_n(\theta) = \prod_{i=1}^n f_\theta(X_i)$
(variable aléatoire)

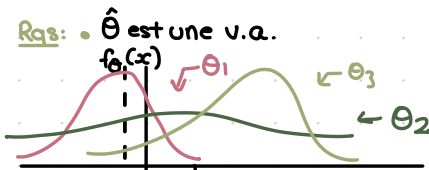
$\sqrt{\text{Likelihood}}$

Un estimateur du max de vraisemblance $\hat{\theta}_{MV}$ est défini par: $\forall \theta \in \Theta, L_n(\theta) \leq L_n(\hat{\theta})$

On travaille souvent avec la log-vraisemblance $\log L_n(\theta) = \sum_{i=1}^n \ln f_\theta(X_i)$ somme de v.a.

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

Rqs: • $\hat{\theta}$ est une v.a.



$X \rightarrow \hat{\theta} = \theta_1$, $\bar{X} \rightarrow \hat{\theta} = \theta_3$

Exemples:

• Bernoulli (θ)

X_i à valeurs 0-1, $f_\theta(x) = \theta^x (1-\theta)^{1-x}$

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

continue (C^∞) en θ

$\log L_n(\theta) = (\sum X_i) \ln \theta + (n - \sum X_i) \ln(1-\theta)$

$\rightarrow (\log L_n)'(\theta) = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta}$ (*)

équation de vraisemblance: $(\log L_n)'(\theta) = 0 \Leftrightarrow (1-\theta) \sum X_i = (n - \sum X_i) \theta$
 $\Leftrightarrow \sum X_i = n\theta \Rightarrow \theta = \frac{\sum X_i}{n}$

pt critique est-il un maximum?

(*) = $\frac{\sum X_i - n\theta}{\theta(1-\theta)}$ la dérivée s'annule et change de signe en \bar{X}

= $\frac{n}{\theta(1-\theta)} (\bar{X} - \theta)$

• La dérivée change de signe en $\bar{X} \rightarrow$ on a bien un max $\rightarrow \hat{\theta}^{MV} = \bar{X}$

• condition du second ordre:

si $(\log L_n)''(\theta) < 0$ pour tout $\theta \Rightarrow \log L_n$ est concave \Rightarrow max global

$$(\log L_n)''(\theta) = \frac{-\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2} < 0 \quad \forall \theta$$

• condition du 2nd ordre "locale"

si $(\log L_n)''(\hat{\theta}) < 0$, $\hat{\theta}$ est un max local. mais par unicite' du point critique, c'est L'EMV

Ex: $f_{\theta}(x) = \theta x^{\theta-1} \mathbb{1}_{x \in [0,1]}$ EMV?

Information de Fisher, efficacité

Soit $(P_\theta)_{\theta \in \Theta}$, $\Theta \subset \mathbb{R}^p$ (identifiable, dominé)
↳ toutes les lois admettent une densité

On note f_θ densité de P_θ

$\text{Supp } f_\theta = \{x \in E; f_\theta(x) > 0\}$

Étant donné (X_1, \dots, X_n) iid de loi P_θ

$\theta \rightarrow L_n(\theta)$ vraisemblance de l'échantillon
 $= \prod_{i=1}^n f_\theta(X_i) \rightarrow$ on peut calculer $\log L_n(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$

$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \log L_n(\theta)$

Propriété d'invariance de l'EMV

si $\hat{\theta}$ EMV de θ , $g(\hat{\theta})$ est un EMV de $g(\theta)$

Objectif: que peut on avoir de "mieux" comme estimateur?

↳ modèle régulier

1) Modèle régulier

DEF 1: Le modèle $(P_\theta)_{\theta \in \Theta}$ est dit régulier si :

a) Θ est un ouvert et pour tout $x \in E$ $\theta \rightarrow f_\theta(x)$ est C^1

b) $\text{Supp } f_\theta$ ne dépend pas de θ : $S = \{x; f_\theta(x) > 0\}$

c) Pour tout θ , l'application $x \mapsto \frac{(\frac{\partial f_\theta}{\partial \theta}(x))^2}{f_\theta(x)} \mathbb{1}_{f_\theta(x) > 0}$

est intégrable ($L_1(\mu)$) et l'intégrale $I(\theta) = \int_S \frac{(\frac{\partial f_\theta}{\partial \theta}(x))^2}{f_\theta(x)} dx$ est continue sur Θ

Notation: on note la dérivée de $f_\theta(x)$ par rapport à θ $\frac{\partial f_\theta}{\partial \theta}(x)$

La quantité $\theta \mapsto I(\theta)$ est appelée information de Fisher du modèle

Exemple:

• $f_\theta(x) = \theta e^{-x\theta}$ densité par rapport à $\mu(dx) = \mathbb{1}_{x \geq 0} dx$

$\theta \mapsto \theta e^{-x\theta}$ est C^∞ sur $\Theta =]0, +\infty[$

$\text{Supp } f_\theta = \mathbb{R}_+$

(si $X \sim \mathcal{E}(\theta)$, $\mathbb{E}[X^k] = \frac{k!}{\theta^k}$)

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

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

$I(\theta) = \int_0^{+\infty} \frac{(1-x\theta)^2}{\theta} \theta e^{-x\theta} dx = \frac{1}{\theta^2} E_\theta (1-X\theta)^2$
 $= \frac{1}{\theta^2} (1 - 2\theta \mathbb{E}[X] + \theta^2 \mathbb{E}[X^2])$

$= \frac{1}{\theta^2}$ continue sur $]0, +\infty[$

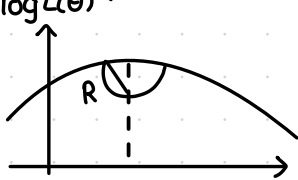
Ex2: (X_1, \dots, X_n) i.i.d. Poisson(θ) $f_\theta(x) = \frac{e^{-\theta} \theta^x}{x!}$

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

3. Info de Fisher et dérivée seconde

PROP3: En ajoutant que $\theta \mapsto f_\theta(x)$ est C^2 et que (H) vraie pour $\frac{\partial^2}{\partial \theta^2}$, alors l'info de Fisher s'écrit encore
 $I_n(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \log L_n(\theta) \right]$ (si $\hat{\theta} \in \text{EMV}$, $I_n(\hat{\theta}) > 0$)

$\log L(\theta)$ (si $\hat{\theta} \in \text{EMV}$, $I_n(\hat{\theta}) > 0$)



$\frac{\partial}{\partial \theta^2} \log(\theta) \propto \frac{1}{R}$

si courbe très "piquée" en l'EMV (i.e. info Fisher est grande) alors l'EMV est localisé de façon précise.

$n=1: \frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{\frac{\partial^2 f_\theta(x)}{\partial \theta^2}}{f_\theta(x)} - \left(\frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} \right)^2$
 $\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1)\right) = \int_S \frac{\frac{\partial^2 f_\theta(x)}{\partial \theta^2}}{f_\theta(x)} f_\theta(x) dx - \int_S \frac{\frac{\partial f_\theta(x)}{\partial \theta}}{f_\theta(x)} f_\theta(x) dx$
 $\underbrace{\hspace{10em}}_{I(\theta)}$

4) Inégalité de Gramer-Rao

Soit $g(\theta)$ le paramètre d'intérêt où $g: \Theta \rightarrow \mathbb{R}$, $\theta \in \mathbb{R}$

PROP4: Sous les hypothèses d'un modèle régulier, si pour tout θ , $I(\theta) > 0$
 alors pour tout estimateur $T = T(X_1, \dots, X_n)$ sans biais, $\mathbb{E}_\theta T^2 < +\infty$
 On a $\forall \theta \in \Theta$, $\text{Var}_\theta(T) \geq \frac{(g'(\theta))^2}{I_n(\theta)}$

DEF: si T réalise l'égalité, alors T est dit efficace

preuve: $\forall \theta$, $\mathbb{E}_\theta(T) = g(\theta)$

$\Rightarrow \frac{\partial}{\partial \theta} \mathbb{E}_\theta(T) = g'(\theta)$

$\stackrel{n=1}{T=T(X_1)} \Leftrightarrow \frac{\partial}{\partial \theta} \int_S T(x) f_\theta(x) dx = g'(\theta)$

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

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

Inég de Cauchy-Schwarz pour $\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) f_\theta(x) dx$ avec $h_1(x)$ et $h_2(x)$ centrées

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

Étude asymptotique des estimateurs

Dans un modèle paramétrique régulier, si $\hat{\theta}_n$ estimateur de θ ,

alors $\text{Var}(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{n I(\theta)}$

si $\text{Var}(\hat{\theta}_n) = \frac{1}{n I(\theta)}$ sans biais, $\hat{\theta}_n$ est efficace

Asymptotique: $n \rightarrow +\infty$

$\lim_{n \rightarrow +\infty} n \text{Var} \hat{\theta}_n \xrightarrow{n \rightarrow +\infty} \frac{1}{I(\theta)}$?

1) Convergences

$(X_n)_{n \geq 0}$ suite de v.a. réelles (\mathbb{R}^d)

→ convergence en loi $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ ssi $\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x)$
 en tout point de continuité de $x \rightarrow \mathbb{P}(X \leq x)$

→ lemme de Portmanteau:

Caractérisations équivalentes:

- Pour toute fonction continue bornée h , $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$

⇒ la convergence en loi est stable par passage aux fonctions continues (LAC)

MAIS il est en général faux que si $X_n \xrightarrow{\mathcal{L}} X$ alors $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$

Cela est vrai dans 3 cas:

① • si $\forall n$, X_n et Y_n sont indépendantes alors convergence en loi de X_n et Y_n
 X et Y sont indépendantes alors cv en loi du couple $\begin{pmatrix} X_n \\ Y_n \end{pmatrix}$

② • si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$) $\Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} X \\ Y \end{pmatrix} \Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ Y \end{pmatrix}$

③. Lemme de Slutsky

Si $\begin{cases} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} c \end{cases}$, alors $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ c \end{pmatrix}$

en appliquant le LAC:

$h(x,y) = \begin{cases} x+y & X_n + Y_n \xrightarrow{\mathcal{L}} X + c \\ xy & X_n Y_n \xrightarrow{\mathcal{L}} cX \\ x/y & \text{si } c \neq 0 \quad X_n/Y_n \xrightarrow{\mathcal{L}} X/c \end{cases}$

2) Consistance des estimateurs

DEF1: $\hat{\theta}_n$ asymptotiquement sans biais ssi $\text{Biais}(\hat{\theta}_n, \theta) = \mathbb{E}[\hat{\theta}_n] - \theta \xrightarrow[n \rightarrow +\infty]{} 0$

Rq: la convergence en proba n'implique pas la convergence des espérances!

si $X_n \xrightarrow{\mathbb{P}} X$

$|X_n| \leq Y \in L^1$, alors par convergence dominée $X_n \rightarrow X$ dans L^1 .

exemple: $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$ estimateur des moments de $\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Biais: $(\hat{\sigma}_n^2, \sigma^2) = \frac{1}{n} \sigma^2$ as. sans biais (as. = asymptotiquement)

outils pour montrer la consistance

- * LGN
- * si $R(\hat{\theta}_n, \theta) \rightarrow 0$, alors $\hat{\theta}_n$ consistant car $L^2 \Rightarrow$ en proba
- * revenir à la définition de la convergence en proba

• si les (X_i) i.i.d, alors (X_i^2) est i.i.d.

$$\mathbb{E}[X_i^2] < +\infty$$

$$\text{LGN: } \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X^2] = \sigma^2 + \mu^2$$

• $\bar{X} \xrightarrow{\mathbb{P}} \mu$ (LGN)

LAC avec $h(x) = x^2$; $(\bar{X})^2 \xrightarrow{\mathbb{P}} \mu^2$

$$\text{Donc } \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \frac{1}{n} \sum_{i=1}^n X_i \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} \sigma^2 + \mu^2 \\ \mu \end{pmatrix}$$

• LAC $h(x, y) = x - y^2$

Donc

$$\frac{1}{n} \sum X_i^2 - \bar{X}^2 \xrightarrow{\mathbb{P}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

3) Normalité asymptotique:

$\hat{\theta}_n$ consistant pour θ

↳ Question: quelle est la vitesse de convergence de $\hat{\theta}_n$ vers θ ?

(X_1, \dots, X_n) i.i.d, d'espérance $\sigma^2(\theta)$ $\hat{\theta} = \bar{X}$

TLC: $\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \sigma^2(\theta))$ quelle que soit la loi des X_i
↳ vitesse en \sqrt{n}

DEF: $(\hat{\theta}_n)$ est un estimateur asymptotiquement normal ssi

- vitesse de convergence en \sqrt{n}
- convergence en loi
- loi limite est la loi normale

exemple: $\hat{\sigma}_n^2$ est elle as. normale?

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

(X_1, \dots, X_n) i.i.d, d'espérance μ , de variance σ^2 , $\mu_4 = \mathbb{E}((X_i - \mu)^4) < +\infty$

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

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

• TLC: (X_i) i.i.d, alors les $(X_i - \mu)$ sont i.i.d, d'espérance σ^2

$$\sqrt{n} \cdot \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \mu_4 - \sigma^4)$$

$$\neq \text{Var}(X_i - \mu)^2 = \mathbb{E}((X_i - \mu)^4) - \sigma^4 = \mu_4 - \sigma^4$$

• TLC: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \underbrace{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \sigma^2 \right)}_{Z \sim \mathcal{N}(0, \mu_4 - \sigma^4)} - \underbrace{\sqrt{n}(\bar{X} - \mu)^2}_{\sqrt{n}(\bar{X} - \mu) \times (\bar{X} - \mu)}$$

$$\downarrow \mathcal{L} \quad \mathbb{P} \downarrow \mathcal{L}$$

$$\mathcal{N}(0, \sigma^2) \quad 0 \text{ constante!}$$

$$\bar{X} - \mu \xrightarrow{\mathcal{L}} 0, \sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{L}} U \sim \mathcal{N}(0, 1) \Rightarrow \sqrt{n}(\bar{X} - \mu)^2 \xrightarrow{\mathcal{L}} 0 \quad (\mathbb{P})$$

nouvelle applic du lemme de Slutsky:

lemme de Slutsky

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} Z + 0$$

Donc $\hat{\sigma}_n^2$ est un estimateur asymptotiquement normal.

Rq: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$

$\Leftrightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$
 propriété des lois normales

Application du lemme de Slutsky: si $\hat{\sigma}^2$ est un estimateur consistant de σ^2 , alors on a encore

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (\hat{\sigma} = \hat{\sigma}(X_1, \dots, X_n))$$

preuve (de la remarque):

$$\begin{aligned} \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}} &= \underbrace{\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)} \times \underbrace{\frac{\sigma}{\hat{\sigma}}}_{\xrightarrow{\mathbb{P}} 1} \\ &\quad (\hat{\theta} \text{ as. normal}) \\ &\quad (\text{par consistence } \sigma^2 \xrightarrow{\mathbb{P}} \sigma^2 \\ &\quad \xrightarrow{\text{LAC}} \sigma^2 \xrightarrow{\mathbb{P}} \sigma) \\ &\xrightarrow{\text{par Slutsky}} \mathcal{L} \rightarrow 1. Z \sim \mathcal{N}(0, 1) \end{aligned}$$

4) δ -méthode:

$\hat{\theta}$ estimateur as. normal: quelle est la loi as. de $g(\theta)$?

LEMME: [méthode δ]

Soit Z_n une suite de v.a. réelles tq $\sqrt{n}(Z_n - \mu) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \sigma^2)$

Soit g une fonction dérivable, $g'(\mu) \neq 0$

Sous ces hypothèses, on a $\sqrt{n}[g(Z_n) - g(\mu)] \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \tilde{Z} \sim \mathcal{N}(0, (g'(\mu))^2 \sigma^2)$

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + (x - \mu)R(x - \mu) \text{ où } R(y) \xrightarrow[y \rightarrow 0]{} 0$$

pu:

$$\sqrt{n}(g(Z_n) - g(\mu)) = \underbrace{g'(\mu) \sqrt{n}(Z_n - \mu)}_{\xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, \sigma^2)} + \underbrace{\sqrt{n}(Z_n - \mu)R(Z_n - \mu)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \text{ et } \xrightarrow{\mathbb{P}} 0}$$

A-t-ous $Z_n \xrightarrow{\mathbb{P}} \mu$?

$$\begin{aligned} \mathbb{P}(|Z_n - \mu| > \varepsilon) &= \mathbb{P}\left(\frac{\sqrt{n}|Z_n - \mu|}{\sigma} > \frac{\sqrt{n}\varepsilon}{\sigma}\right) = \mathbb{P}\left(\frac{\sqrt{n}(Z_n - \mu)}{\sigma} > \frac{\sqrt{n}\varepsilon}{\sigma}\right) + \mathbb{P}\left(\frac{\sqrt{n}(Z_n - \mu)}{\sigma} < -\frac{\sqrt{n}\varepsilon}{\sigma}\right) \\ &\stackrel{(\text{erreur})}{\approx} 1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) + \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right) = 2\left(1 - \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)\right) \end{aligned}$$

Φ f.n. $\frac{\sqrt{n}(Z_n - \mu)}{\sigma}$

Fonction de répartition empirique

(X_1, \dots, X_n) échantillon i.i.d. à valeurs réelles de loi F inconnue

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

DEF: La fonction de répartition empirique, associée à (X_1, \dots, X_n) est définie par:

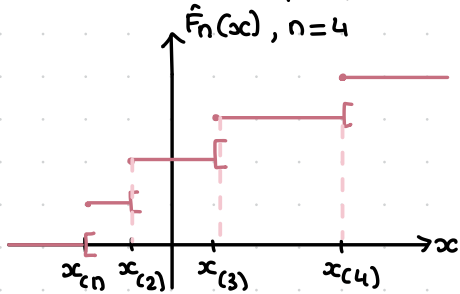
$\hat{F}_n : \mathbb{R} \rightarrow [0, 1]$
 $x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$

$\forall x \in \mathbb{R}, \hat{F}_n(x)$ est une v.a., estimateur de $F(x)$

DEF: Loi empirique $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ = loi discrète uniforme sur $\{X_1, \dots, X_n\}$

Représentation graphique:

Conditionnellement $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ valeurs ordonnées



si les $x_i \neq$; saut de $1/n$

a) Propriétés immédiates:

$n \hat{F}_n(x) = \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$ suit la loi binomiale $(n, F(x))$

$R(\hat{F}_n(x), F(x)) = 0 + \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n \mathbb{1}_{X_i \leq x} \right) = \frac{1}{n} F(x) (1 - F(x)) \xrightarrow[n \rightarrow +\infty]{} 0$
indep

ecdf
 → empirical cumulative distribution function

donc $\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow{\mathbb{P}} F(x)$

ou bien LGIV $\hat{F}_n(x)$ estimateur consistant de $F(x)$

On a un résultat de convergence uniforme: $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$ (Thm de Glivenko - Cantelli)

$\hat{F}_n(x)$ est-il asymptotiquement normal?

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

TCL: les X_i sont iid, donc les $\mathbb{1}_{X_i \leq x}$ sont iid

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

$\left(\frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \right) \stackrel{\text{var Bin}}{\sim} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$

b) Estimation empirique:

"plug-in" ou méthode de substitution

paramètre d'intérêt $\theta = c(F)$ (caractéristique de la fonction inconnue)

la méthode empirique définit $\hat{\theta}$, estimateur empirique en remplaçant F par \hat{F}_n

$\hookrightarrow \hat{\theta}_n = c(\hat{F}_n)$

ex: $\theta = \mathbb{E}_F(X_1) \xrightarrow{= \int x dF(x)} \hat{\theta}_n = \mathbb{E}_{\hat{F}_n}(X_1) = \sum_{i=1}^n X_i \cdot \frac{1}{n}$ (si les X_i distincts)

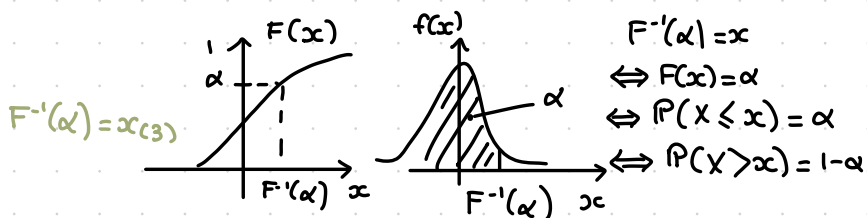
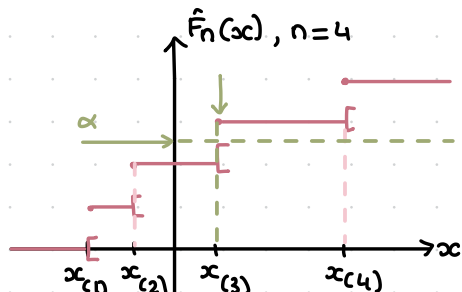
$= \bar{X}$

$$\theta = \text{Var}_F(X) \longrightarrow \hat{\theta}_n = \text{Var}_{\hat{F}_n}(X) = \frac{1}{n} \sum (X_i - \bar{X})^2 \quad (\text{estimateurs des moments})$$

c) Inverse généralisé

DEF: On définit l'inverse généralisé de F par $F^{-1}: [0,1] \longrightarrow \mathbb{R}$
 $\forall \alpha \in [0,1], F^{-1}(\alpha) = \inf \{ x \in \mathbb{R}, F(x) \geq \alpha \}$

- si F strict \uparrow , $\inf x \text{ tq } F(x) \geq \alpha \Leftrightarrow x \geq F^{-1}(\alpha) \rightsquigarrow x = F^{-1}(\alpha)$
- si F est la fr d'une loi discrète



VOCAB: F^{-1} s'appelle aussi la fonction quantile

$F^{-1}(\alpha)$ = quantile d'ordre α de la loi F

$F^{-1}(1/4)$ = 1^{er} quantile, $F^{-1}(1/2)$ = médiane, $F^{-1}(3/4)$ = 3^{er} quantile

LEMME: U v.a. uniforme sur $[0,1]$

F une f.c.

alors $F^{-1}(U)$ est une v.a. de loi F .

si F est bijective, $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$
 \uparrow F bijective car $P(U \leq x) = x$ sur $[0,1]$

si F est discrète, F^{-1} inverse généralisé, $F^{-1}(y) \leq x \Leftrightarrow y \leq F(x)$ (suite en TD)

d) Quantile empirique

DEF: On définit le quantile empirique (sample quantile) d'ordre α comme étant le quantile de \hat{F}_n :

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

Propriétés:

- On peut montrer que $\hat{q}_{n,\alpha} = X_{(\lceil n\alpha \rceil)}$ où $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ est l'ech. ordonné des $(X_i)_{1 \leq i \leq n}$
 $\lceil u \rceil = + \text{petit entier } \geq u$

ex: $\alpha = \frac{1}{2} \left[\frac{n}{2} \right]$ si $n = 2k$ médiane $\hat{q}_{n,1/2} = X_{(k)}$
 si $n = 2k+1$ médiane $\hat{q}_{n,1/2} = X_{(k+1)}$

Dans les logiciels, plusieurs façons de calculer les quantiles empiriques

$n=4$ médiane = $X_{(2)}$

$n=5$ médiane = $X_{(3)} \rightsquigarrow \frac{X_{(2)} + X_{(3)}}{2}$

$n=6$ médiane = $X_{(3)}$

- **consistance** si $\alpha \in]0,1[$, si F est strictement croissante au voisinage de α , alors $\hat{q}_{n,\alpha}$ consistant de q_α .

Intervalle de confiance (pas dans le partiel)

(X_1, \dots, X_n) i.i.d. de loi $P \in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$.

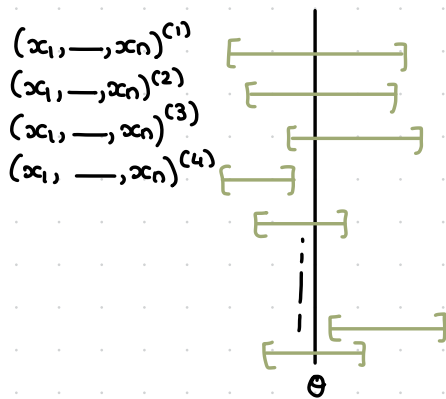
On s'intéresse à $\theta \in \mathbb{R}$ ou $g(\theta): \mathbb{R}^p \rightarrow \mathbb{R}$

1) DEF:

Un intervalle de confiance pour θ , de niveau de confiance $1 - \alpha$; $\alpha \in]0, 1[$, est un intervalle dont les bornes sont aléatoires, fonctions de l'échantillon et ne dépendent PAS des paramètres inconnus du modèle, et tel que $P([B_{\text{inf}}(X_1, \dots, X_n); B_{\text{sup}}(X_1, \dots, X_n)] \ni \theta) \geq 1 - \alpha$

- Un IC est calculable à partir des données
- Si l'inégalité est une égalité = niveau de confiance est exact
- si on a $P(\theta \in [B_{\text{inf}}, B_{\text{sup}}]) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$, niveau est asymptotique
- en général, $\alpha = 1\%$, 5%

Interprétation:



$IC = [B_{\text{inf}}(X_1, \dots, X_n), B_{\text{sup}}(X_1, \dots, X_n)]$
formule mathématique qui garantit le niveau $1 - \alpha$

On observe $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, une réalisation de l'échantillon aléatoire
↳ on calcule $IC = [2.3; 5.1]$ de niveau de confiance 95%.

En moyenne, sur 100 intervalles calculées 5% 'râte' θ .

2) Méthode pivotale

(X_1, \dots, X_n) i.i.d. d'espérance $\theta \in \mathbb{R}$, de variance $\sigma^2(\theta)$

Soit $\hat{\theta}$, asymptotiquement normal: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, \sigma^2(\theta))$

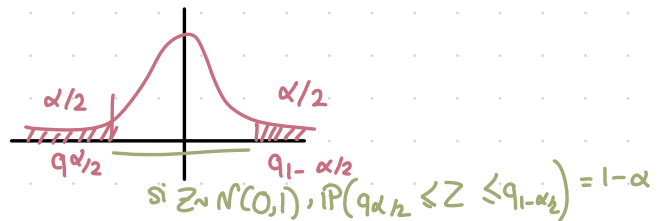
$$\Leftrightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1)$$

propriété des lois Gaussiennes

Par définition des quantiles gaussiens, $q_\alpha = \Phi^{-1}(\alpha)$ où Φ f.r. de $\mathcal{N}(0, 1)$

$$P\left(q_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)} \leq q_{1-\alpha/2}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$$

↳ inconnu



• pivot ou statistique pivotale = $\frac{\sqrt{n}(\hat{\theta} - \theta)}{\hat{\sigma}}$ statistique centrée réduite issue de $\hat{\theta}$, où $\sigma^2(\theta)$ estimé par $\hat{\sigma}^2$, consistant pour estimer $\sigma^2(\theta)$.

$$\text{Si c'est le cas, } \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma(\theta)} \times \frac{\sigma^2(\theta)}{\hat{\sigma}^2} \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1)$$

($\hat{\theta}$ as. normal) $\downarrow \mathcal{N}(0, 1)$ $\downarrow P(\text{estimateur consistant})$ lemme de Slutsky,

• on en déduit $P\left(q_{\alpha/2} \leq \sqrt{n} \frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq q_{1-\alpha/2}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$

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

Compléments: (avant partiel)

- 1) Retour sur normalité a.s.
- 2) Exemple
- 3) Pivot asymptotique
- 4) Exemple 2

1. Propriétés asymptotiques d'une suite d'estimateurs $(\hat{\theta}_n)_{n \geq 1}$

→ Consistance $\hat{\theta}_n \xrightarrow{P} \theta$

→ Normalité asymptotique: si $\exists \sigma^2 > 0$ $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{L} \mathcal{N}(0, \sigma^2)$

De façon générale, si il existe $U_n \xrightarrow[n \rightarrow +\infty]{L} 0$

$V_n(\hat{\theta}_n - \theta) \xrightarrow{L} Y$

on dit que $\hat{\theta}_n$ converge à la vitesse $\frac{1}{V_n}$

Rq: si $\hat{\theta}_n$ as. normal $\Rightarrow \hat{\theta}_n$ consistant

$$\hat{\theta}_n - \theta = \frac{1}{\sqrt{n}} \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{L} 0 \quad \text{Slutsky}$$

\downarrow $\mathcal{N}(0, \sigma^2)$
 \downarrow 0

$$\mu_n = \frac{1}{\sqrt{n}} \rightarrow 0$$

→ Δ-méthode

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

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

$$\hookrightarrow X_n \approx 1 + \frac{1}{\sqrt{n}} Z$$

g dérivable en 1

* si g est dérivable en 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-méthode

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

g dérivable en θ $g(x) = g(\theta) + (x-\theta)[g'(\theta) + r(x)]$; $r(x) \xrightarrow[x \rightarrow \theta]{} 0$

$\hat{\theta}_n \xrightarrow{P} \theta$ donc (LAC) $r(\hat{\theta}_n) \rightarrow r(\theta) = 0$

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

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = \underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}_Z \underbrace{[g'(\theta) + r(\hat{\theta}_n)]}_{g'(\theta)} \xrightarrow{L} g'(\theta) Z \sim \mathcal{N}(0, (g'(\theta))^2)$$

Slutsky

2) Exemple

X_1, \dots, X_n de loi de densité $f(x) = \frac{1}{\mu} e^{-x/\mu}$; $x \geq 0, \mu = E[X_1] > 0$

μ est estimé par $\hat{\mu} = \bar{X}$

efficace? $\log L_n(\mu) = -n \log \mu - \frac{1}{\mu} \sum_{i=1}^n X_i$

↓
sans biais qui atteint la borne de CR

$$\text{Var}(\hat{\mu}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right)$$

$$\stackrel{\text{indep}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \stackrel{\text{i.d.}}{=} \frac{1}{n} \text{Var}(X_1) = \frac{\mu^2}{n}$$

$$\text{Var}(X_1) = \mu^2$$

méthode pivotale pour IC:

On estime $\sqrt{\theta(1-\theta)}$ par $\sqrt{\hat{\theta}(1-\hat{\theta})}$ "plug-in"

par le LAC, $g(x) = \sqrt{x(1-x)}$; $x \in]0,1[$, $\sqrt{\hat{\theta}(1-\hat{\theta})}$ est un estimateur consistant de $\sqrt{\theta(1-\theta)}$

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{\hat{\theta}(1-\hat{\theta})}} = \underbrace{\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)}}}_{\substack{\text{(TCL)} \\ \downarrow \\ \mathcal{L} \\ \mathcal{N}(0,1)}} \times \underbrace{\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\hat{\theta}(1-\hat{\theta})}}}_{\substack{\downarrow \text{P} \\ \text{consistant}}}$$

donc, par Slutsky,

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

4) Exemple 2:

(X_1, \dots, X_n) de densité $\theta > 0$: $f_{\theta}(x) = \frac{3}{\theta} x^2 \exp(-\frac{x^3}{\theta}) \mathbb{1}_{x \geq 0}$ par rapport à mesure $\mathbb{1}_{x \geq 0} dx$
 pour gérer l'indicatrice ↪ changement de mesure dominante
 (ou dire partyp, $X_i \geq 0 \forall i$)

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

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

$(\log L_n)''(\theta) = \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum X_i^3 \Rightarrow (\log L_n)''(\hat{\theta}) = \frac{n}{\hat{\theta}^2} - \frac{2}{\hat{\theta}^3} \cdot n\hat{\theta} = \frac{-n}{\hat{\theta}^2} < 0$
 ⇒ unicite' ⇒ max global

TCL: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2)$
 ⇔ $\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\theta} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$
 pivot asymptotique

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

si on note $q_{\alpha/2}$ et $q_{1-\alpha/2}$ quantiles de $\mathcal{N}(0, 1)$

$\mathbb{P}\left(q_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\theta}} \leq q_{1-\alpha/2}\right) \xrightarrow{n \rightarrow +\infty} 1-\alpha$

$\mathbb{P}\left(q_{\alpha/2} \frac{\hat{\theta}}{\sqrt{n}} \leq \hat{\theta}_n - \theta \leq q_{1-\alpha/2} \frac{\hat{\theta}}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 1-\alpha$

$\mathbb{P}\left(\hat{\theta}_n - q_{1-\alpha/2} \frac{\hat{\theta}}{\sqrt{n}} \leq \theta \leq \hat{\theta}_n + q_{\alpha/2} \frac{\hat{\theta}}{\sqrt{n}}\right) \xrightarrow{n \rightarrow +\infty} 1-\alpha$

⇒ IC(θ) de niveau asymptotique $(1-\alpha)$

Estimation dans les échantillons gaussiens:

- 1) Loi normale et lois dérivées
- 2) Loi des estimateurs empiriques
- 3) IC des paramètres
- 4) Exercice

DEF: Z est dite gaussienne (normale) centrée réduite si sa loi admet pour densité:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; x \in \mathbb{R}. \text{ On note } Z \sim \mathcal{N}(0,1)$$

X est dite de loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ si $X = \mu + \sigma Z$, notée $X \sim \mathcal{N}(\mu, \sigma^2)$

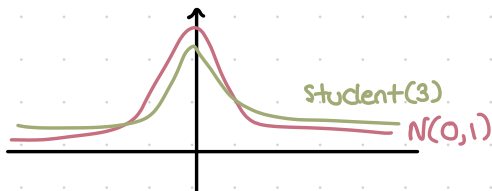
Autres caractérisations de la loi normale:

• par densité: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

• par la fonction génératrice des moments: $M(t) = \mathbb{E}[e^{tx}] = e^{t\mu + \frac{1}{2}\sigma^2 t^2}; \forall t \in \mathbb{R}$

Remarque:

- $\sigma^2 = 0 \rightarrow X = \mu$ p.s.
- si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, et $\lambda \in \mathbb{R}$, alors $\lambda X_1 + X_2 \sim \mathcal{N}(\lambda\mu_1 + \mu_2, \lambda^2\sigma_1^2 + \sigma_2^2)$
- moments centrés: densité symétrique par rapport à μ . $\mathbb{E}[(X-\mu)^k]$



• tous les moments centrés d'ordre impaire sont nuls.

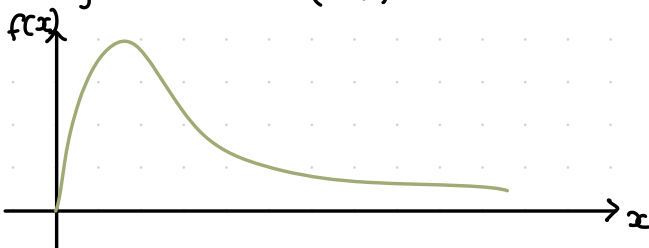
• $\mu_{2k} = \frac{(2k)!}{2^k k!} \sigma^{2k}$

$\rightarrow \mathbb{E}[(X-\mu)^4] = 3\sigma^4$

$\rightarrow \text{Var}(X) = \mathbb{E}[(X-\mu)^2] = \sigma^2$

DEF: (X_1, \dots, X_d) échantillon i.i.d. $\mathcal{N}(0,1)$

La loi de $X_1^2 + X_2^2 + \dots + X_d^2$ est appelée loi du χ^2 (chi 2) à d degrés de liberté (ddl) / degrees of freedom (df)



Corr:

- si Y de loi $\chi^2(d)$, $E[Y] = d$, $\text{Var}(Y) = 2d$
- $\text{Var}(X_i^2 + \dots + X_d^2) \underset{\text{indep}}{=} d \text{Var}(X_i^2)$
 $E X_i^4 = E[X_i^2]^2 = 3 - 1 = 2$
- support \mathbb{R}_+
- $M(t) = (1-t)^{-d/2}$; $(t < 1/2)$

DEF: Si $X \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(d)$ indépendants, la loi de $Z = \frac{X}{\sqrt{Y/d}}$ est appelée loi de Student à d ddl

Rq: si $d \rightarrow +\infty$, la loi de Student converge vers la loi $\mathcal{N}(0,1)$

PV: $\frac{Y}{d} \stackrel{\text{def}}{=} \frac{1}{d} \sum_{i=1}^d U_i^2$, où $U_i \sim \mathcal{N}(0,1)$ indep entre elles de X
 $\xrightarrow[\text{LGN}]{\mathbb{P}} E(U_i^2) = 1$

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

Par le lemme de Slutsky, $Z \xrightarrow{\mathcal{L}} 1 \cdot X \sim \mathcal{N}(0,1)$

* On introduit (X_1, \dots, X_n) i.i.d $\sim \mathcal{N}(\mu, \sigma^2)$ où μ et σ^2 param inconnus

$$\rightarrow \mu = E[X_i] \rightsquigarrow \hat{\mu} = \bar{X}$$

$$\rightarrow \sigma^2 = \text{Var}(X_i) \rightsquigarrow \hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

Soit $S_n^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ non biaisé

2) Loi des estimateurs empiriques:

Thm: [loi de $\hat{\mu}$ et $\hat{\sigma}^2$]

• \bar{X} et $\sum_{i=1}^n (X_i - \bar{X})^2$ sont des v.a. indépendantes

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

$$\bullet \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1) \Rightarrow \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1) \text{ et } \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

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

• \bar{X} et $(\underbrace{X_1 - \bar{X}, \dots, X_n - \bar{X}}_T)$ sont indépendants

+ preuve

preuve:

$$\begin{aligned}
 M(u, t_1, \dots, t_n) &= \mathbb{E} \left[e^{u\bar{X} + t_1(X_1 - \bar{X}) + \dots + t_n(X_n - \bar{X})} \right] \\
 &= \mathbb{E} \left[\exp \left(\left(\frac{u}{n} + \frac{t_1 + t_2 + \dots + t_n}{n} \right) X_1 \right) \dots \exp \left(\frac{u}{n} + \bar{t} \right) X_n \right] \\
 &= \mathbb{E} \left[\prod_{i=1}^n e^{(\frac{u}{n} + t_i - \bar{t}) X_i} \right] \\
 &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n \mathbb{E} \left[e^{(\mu_n + t_i - \bar{t}) X_i} \right] \\
 &= \prod_{i=1}^n e^{M(\mu_n + t_i - \bar{t})} \\
 &= \prod_{i=1}^n e^{M \left(\frac{u}{n} + t_i - \bar{t} \right) + \frac{\sigma^2}{2} (\mu_n + t_i - \bar{t})^2} \\
 &= e^{\sum_{i=1}^n M \sum_i (t_i - \bar{t}) + \frac{\sigma^2}{2} \sum_i \left(\frac{u^2}{n^2} + (t_i - \bar{t})^2 + 2 \frac{u}{n} (t_i - \bar{t}) \right)} \\
 &= e^{M u + \frac{\sigma^2}{2} \left(\frac{u^2}{n} + \sum_i (t_i - \bar{t})^2 \right)} \\
 &= e^{M u + \frac{\sigma^2 u^2}{2n}} e^{\frac{\sigma^2}{2} \sum_i (t_i - \bar{t})^2} \\
 &\stackrel{M_{\bar{X}}(u)}{=} e^{M u} \stackrel{M_T(t_1, \dots, t_n)}{=} e^{\frac{\sigma^2}{2} \sum_i (t_i - \bar{t})^2}
 \end{aligned}$$

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

par indep $\Rightarrow M_{\chi^2(n)}(t) = M_{\chi^2(1)}(t)^n \Rightarrow M_{\chi^2(1)}(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{\frac{(n-1)}{2}}$

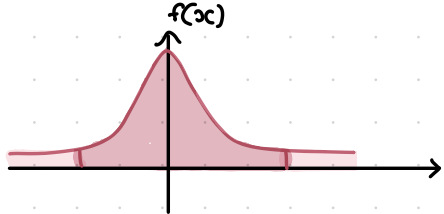
$$\frac{\bar{X} - \mu}{\frac{S_n}{\sqrt{n}}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\frac{S_n}{\sqrt{n}} \times \frac{\sqrt{n}}{\sigma}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{S_n^2/\sigma^2}}$$

qui caractérise la loi $\chi^2(n-1)$

donc $\bar{X} + S_n^2$ indep \Rightarrow def Student $\text{Student}(n-1)$ \square

3) IC des paramètres

Pivot: $\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \sim \text{Student}(n-1)$
loi exacte



$$\mathbb{P} \left(q_{\frac{\alpha}{2}} t(n-1) \leq \frac{\bar{X} - \mu}{S_n/\sqrt{n}} \leq q_{1-\frac{\alpha}{2}} t(n-1) \right)$$

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

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

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

$$\Rightarrow \text{IC} = \left[\frac{n\hat{\sigma}^2}{q_{1-\frac{\alpha}{2}} \chi^2(n-1)}, \frac{n\hat{\sigma}^2}{q_{\frac{\alpha}{2}} \chi^2(n-1)} \right] = \mathbb{P} \left(\frac{n\hat{\sigma}^2}{q_{1-\frac{\alpha}{2}} \chi^2(n-1)} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{q_{\frac{\alpha}{2}} \chi^2(n-1)} \right) = 1 - \alpha$$

q: $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$ et $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$; $\frac{\bar{X} - \mu}{S_n/\sqrt{n}} \sim \text{Student}(n-1)$

Exercice:

$M_{\mu}(\hat{\mu}_n, \hat{\sigma}^2)$ sont les EMV de μ et σ^2
 $R(S_n^2, \sigma^2) > R(\hat{\sigma}_n^2, \sigma^2)$ où R représente un risque

(Δ! exam)

Intro. aux tests statistiques

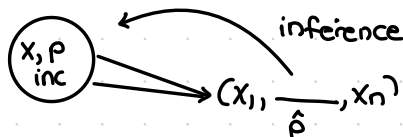
1) Exemple:

Contrôle de qualité: industriel produit des "pièces" $\begin{cases} \rightarrow \text{de bonne qualité} \\ \rightarrow \text{défectueuses} \end{cases}$

↳ Pour l'industriel, on suppose acceptable une proportion de 20% de pièces défectueuses.
 ↳ prélever "au hasard" n pièces, vérifiées ... ($p \leq 20\%$)

modélisation:

i-ème pièce $X_i = \begin{cases} 0 & \text{si bonne qualité} \\ 1 & \text{si défectueuse} \end{cases}$



$p = P(X_i = 1)$

↳ on prélève n pièces et on observe un échantillon (X_1, \dots, X_n) dont les valeurs observées sont (x_1, \dots, x_n) .

Que vaut p?

→ on estime → proportion empirique
 → $X_i \sim \text{Bernouilli}(p)$ indep → $\hat{p} = \bar{X}$

On observe $\bar{x} = 0.22$, $n = 100$

IC(p)?

$\hat{p} = \bar{X}$
TLC: $\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$

on estime l'écart type par $\frac{\hat{p}(1-\hat{p})}{n}$ (consistant)

lemme de Slutsky

$\frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \times \frac{\sqrt{\frac{p(1-p)}{n}}}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}}$

↑ IP (LGN): $\bar{X} \xrightarrow{IP} p$
 LAC: $g(x) = \sqrt{\frac{x(1-x)}{n}}$

$\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$

application méthode pivotale:

$P\left(q_{\text{norm}}_{\alpha/2} \leq \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq q_{\text{norm}}_{1-\frac{\alpha}{2}}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$

$\Leftrightarrow P\left(\bar{X} - q_{\text{norm}}_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} + q_{\text{norm}}_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha$

$IC(p) = \bar{X} \pm q_{\text{norm}}_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$ de niveau asymptotique $1 - \alpha$

ex: $\bar{x} = 0.22$, $\alpha = 5\%$, $n = 100$

$IC = [0.14, 0.30]$

Q: Est-ce que $p \leq 0.2$ ou bien $p > 0.2$?

2) Principe d'un test

$$\Theta \subset]0, 1[$$

On veut tester si $p \leq 0.2$ ou $p > 0.2$

$$\Theta = \Theta_0 \cup \Theta_1 \text{ sous ensembles disjoints}$$

" "

$$]0, 0.2] \quad]0.2, 1[$$

On teste $H_0: p \in \Theta_0$ contre $H_1: p \in \Theta_1$
 $p \leq 0.2$ $p > 0.2$

CCL: Soit on conserve H_0 ($p \leq 0.2$)
Soit on rejette H_0 (on conclut $p > 0.2$)

DEF: un test de H_0 contre H_1 est défini par la construction d'une région de rejet de H_0 , \mathcal{R} :

- si $(X_1, \dots, X_n) \in \mathcal{R}$, on rejette H_0 (au profit de H_1)
- si $(X_1, \dots, X_n) \notin \mathcal{R}$, on conserve H_0

Souvent, $\mathcal{R} = \{(X_1, \dots, X_n), T(X_1, \dots, X_n) > c\}$

T : statistique de test (à valeur réelle)

c : seuil du test

Rq: la décision d'un test est aléatoire (dépend de T aléatoire)

Comment relier \mathcal{R} aux hypothèses testées?

3) Risques d'erreur:

DEF: | erreur de 1^{ère} espèce \mathcal{R} région de rejet du test de H_0 contre H_1
risque de type I (rejeter H_0 à tort)

est la fonction définie sur $\Theta_0 \rightarrow [0, 1]$

$$\alpha: p \rightarrow \mathbb{P}_p((X_1, \dots, X_n) \in \mathcal{R}_0) \\ = \mathbb{P}_p(\text{on rejette } H_0)$$

Le test est dit de niveau α si $\sup_{p \in \Theta_0} \mathbb{P}_p(\text{rejet de } H_0) \leq \alpha$

Rq: erreur de 1^{ère} espèce = $\mathbb{P}(\text{rejet de } H_0 \text{ à tort})$

décision \ réalité	H_0 vraie	H_1 vraie
H_0 vraie	✓	erreur de 1 ^{ère} espèce
H_1 vraie	erreur de 2 ^{ème} espèce	✓

DEF: l'erreur de 2^{ème} espèce est la fonction définie sur $\Theta_1 \rightarrow [0, 1]$

$$\beta: p \mapsto \mathbb{P}_p((X_1, \dots, X_n) \notin \mathcal{R}) \\ = \mathbb{P}_p(\text{on conserve } H_0)$$

Rq: erreur de 2^{ème} espèce est $\mathbb{P}(\text{conserver } H_0 \text{ à tort})$

puissance d'un test: = 1 - erreur de 2nd espèce

$$\Pi: p \in \Theta_1 \longrightarrow \mathbb{P}_p((X_1, \dots, X_n) \in \mathcal{R})$$

choix: les 2 erreurs ne peuvent pas être minimiser simultanément.

En général, $\alpha \nearrow$ quand $\beta \searrow$

test: on choisit de contrôler l'erreur de 1^{ère} espèce
(\Rightarrow l'erreur de 2nde espèce est inconnu en général).

4) Construction d'un test:

Principe: Déterminer \mathcal{R} tq erreur de 1^{ère} espèce $\leq \alpha$

(si on a plusieurs tests, on choisira (point de vue théorique) celui dont l'erreur de 2nde espèce est la plus petite (ou de puissance la plus grande))

\rightarrow di symétrie de H_0 et H_1 dans la construction.

ex: $H_0: p \leq 0.2$ contre $H_1: p > 0.2$
(\bar{m} test que $p = 0.2$ dans ce cas)

(retour sur l'exemple),

- p inconnu donc on l'estime $\hat{p} = \bar{X}$
- idée: sous H_1 , \hat{p} prend de + grandes valeurs que sous H_0
 $\rightarrow \mathcal{R}$ du type $\hat{p} > c$

$$\mathbb{P}_p(\hat{p} > c) \leq \alpha?$$

(pb: trouver c tq $\mathbb{P}_p(\hat{p} > c) \leq \alpha$)

calcul? \rightarrow loi libre du paramètre p ?
(dépend de p)

$$\hat{p} = \bar{X} \rightarrow \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \text{ a pour loi approchée } \mathcal{N}(0, 1)$$
$$\mathbb{P}(\hat{p} > c) = \mathbb{P}\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > c'\right) \quad \text{'' } \frac{c-p}{\sqrt{\frac{p(1-p)}{n}}}$$

On veut que:

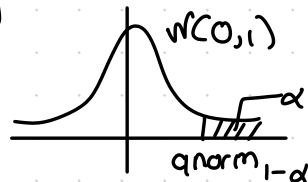
$$\sup_{p \in \Theta_0} \mathbb{P}\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > c'\right) \leq \alpha$$

\rightarrow le sup est atteint en $p = 0.2$

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

Trouver c tq:

$$\mathbb{P}\left((X_1, \dots, X_n) \in \mathcal{R}\right) \xrightarrow[n \rightarrow +\infty]{} \alpha$$



$$\mathbb{P}\left(\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2 \times 0.8}{n}}} > c\right) \xrightarrow[n \rightarrow +\infty]{} \alpha \text{ ssi } c = q_{1-\alpha}^{\text{norm}}$$

- rejet de H_0 ssi $\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} > q_{1-\alpha}^{\text{norm}}$
statistique de test

A.N: $\alpha = 5\%$, $q_{norm} = 1.645$, $n = 100$

$$\hat{p} = \bar{x} = 0.22$$
$$\hookrightarrow T_{obs} = \frac{\bar{x} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{100}}} = \frac{0.02}{0.04} = 1/2 < 1.645 \quad \text{rejet } H_0 \Leftrightarrow \bar{x} > 0.2 + 1.645 \cdot \sqrt{\frac{0.2(1-0.2)}{100}}$$
$$\Leftrightarrow \bar{x} > 0.266$$

CCL: $1/2 < 1.645$

On conserve H_0 . (on ne connaît pas le risque associé)

II. Tests d'hypothèses sur un paramètre

1) Formalisme d'un test

DEF: Un test d'hypothèse est une fonction ψ (mesurable) de l'échantillon (X_1, \dots, X_n) à valeurs dans $\{0, 1\}$

- H_0 est acceptée si $\psi(X_1, \dots, X_n) = 0$
- H_0 est rejetée si $\psi(X_1, \dots, X_n) = 1$

Le domaine $\{(X_1, \dots, X_n), \psi(X_1, \dots, X_n) = 1\} = \mathcal{R}$ est la région de rejet du test, \mathcal{R}^c est la région d'acceptation. On peut écrire: $\psi(X_1, \dots, X_n) = \mathbb{1}_{\mathcal{R}}(X_1, \dots, X_n)$

Très souvent, \mathcal{R} est construite à partir de $T = T(X_1, \dots, X_n)$ statistique de test, elle-même basée sur un estimateur $\hat{\theta}_n$ de θ , paramètre d'intérêt.

Q: Comment construire \mathcal{R} ?

De manière générale, on testera $H_0: \theta = a$ contre $H_1: \theta \neq a$
 $a \in \Theta$

$H_0: \theta \leq a$ contre $H_1: \theta > a$ (ex. contrôle de qualité')

Si on considère une partition $\Theta_0 \cup \Theta_1 = \Theta$ (espace des paramètres), $\Theta_0 \cap \Theta_1 = \emptyset$

$H_0: \theta \in \Theta_0$ contre $H_1: \theta \in \Theta_1$

Vocabulaire:

• $\Theta_0 = \{a\}$ H_0 hyp simple
 $\Theta_1 = \Theta \setminus \{a\}$, H_1 est une hyp bilatère } → test bilatère

• si $\Theta_0 =]-\infty, a]$ et $\Theta_1 =]a, +\infty[$
 H_0 et H_1 sont unilatère. } → test unilatère

• $H_0 = \theta = a$ contre $H_1: \theta > a$ → test unilatère

Risques d'erreur d'un test:

(on privilégie H_0 . Les 2 erreurs évoluent en sens inverse; on ne peut pas contrôler les deux à la fois)

erreur de 1^{ère} espèce: celle que l'on veut contrôler

$$\alpha: \Theta_0 \longrightarrow [0, 1]$$

$$\theta \longrightarrow \mathbb{P}_\theta \left((X_1, \dots, X_n) \in \mathcal{R} \right) = \mathbb{E}_\theta[\psi(X)]$$

$$= \mathbb{P}_{H_0 \text{ vraie}}(\text{rejet de } H_0) \quad \downarrow \text{ sous } H_0$$

niveaux: ssi $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left((X_1, \dots, X_n) \in \mathcal{R}_0 \right) \leq \alpha$

↑ pour les lois discrètes
 = pour les lois continues exactes
 → lois asymptotiques

erreur de 2^{ème} espèce: $\beta: \Theta_1 \longrightarrow [0, 1]$

$$\theta \longrightarrow \mathbb{P}_\theta \left((X_1, \dots, X_n) \in \mathcal{R}^c \right)$$

$$= \mathbb{P}_{H_1 \text{ vraie}}(\text{conserver } H_0) = 1 - \mathbb{E}_\theta[\psi(X)]$$

↓ sous H_1

fonction de puissance: $\pi: \Theta \longrightarrow [0, 1]$

$$\theta \longmapsto \mathbb{P}_\theta \left((X_1, \dots, X_n) \in \mathcal{R}_0 \right)$$

si $\theta \in \Theta_0$ $\pi(\theta) = \alpha(\theta)$

si $\theta \in \Theta_1$: $\pi(\theta) = \mathbb{P}_\theta \left((X_1, \dots, X_n) \in \mathcal{R}_0 \right) = 1 - \mathbb{P}_{H_1} \left((X_1, \dots, X_n) \in \mathcal{R}^c \right)$
 $= 1 - \beta(\theta)$

2) Exemple.

(X_1, \dots, X_n) i.i.d. de loi $\mathcal{N}(\theta, 1)$

Hypothèses à tester:

$H_0: \theta \leq 0$ contre $H_1: \theta > 0$

$\theta = \mathbb{E}[X_i]$ inconnu donc on l'estime: $\hat{\theta} = \bar{X}$

Première idée: rejet de H_0 si $\hat{\theta} > 0$

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

soit $\theta \leq 0$

$$\mathbb{P}_\theta(\hat{\theta} > 0) = \mathbb{P}_\theta(\bar{X} > 0)$$

loi de \bar{X} ? $\bar{X} \sim \mathcal{N}(\theta, 1/n)$ (toute C.L. de gaussiennes est une gaussienne)

loi exacte $\mathbb{E}[\bar{X}] = \mathbb{E}[X_i] = \theta$

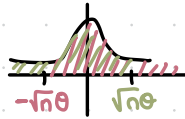
reflexe: normaliser la gaussienne $\text{Var}(\bar{X}) = \frac{1}{n} \text{Var} X_i = 1/n$

$$\alpha(\theta) = \mathbb{P}_\theta\left(\frac{\bar{X} - \theta}{\sqrt{1/n}} > -\sqrt{n}\theta\right) = \mathbb{P}\left(\mathcal{N}(0, 1) > -\sqrt{n}\theta\right)$$

si ϕ : f.r. de la loi $\mathcal{N}(0, 1)$ $= 1 - \phi(-\sqrt{n}\theta)$

$= \phi(\sqrt{n}\theta)$

symétrie de $\mathcal{N}(0, 1)$



$$= \sup_{\theta \leq 0} \phi(\sqrt{n}\theta) = \phi(0) = 1/2 = 50\%$$

i.e. une chance sur 2 de se tromper: pas acceptable

→ On souhaite α petit: $\alpha = 5\%$

$$\mathcal{R} = \{ \hat{\theta} > 0 \} \rightsquigarrow \mathcal{R} = \{ \hat{\theta} > c \}; c > 0 \text{ seuil du test}$$

valeur de $c = c(\alpha)$ tq $\sup_{\theta \leq 0} \alpha(\theta) \leq \alpha$

$$\alpha(\theta) = \mathbb{P}_{\theta \leq 0}(\bar{X} > c) = \mathbb{P}_\theta\left(\frac{\bar{X} - \theta}{\sqrt{1/n}} > \frac{c - \theta}{\sqrt{1/n}}\right) = \mathbb{P}(\mathcal{N}(0, 1) > \sqrt{n}(c - \theta))$$

condition de niveau:

Trouver c tq $\sup_{\theta \leq 0} \mathbb{P}(\mathcal{N}(0, 1) > \sqrt{n}(c - \theta)) = \alpha$ ↙ loi continue

$$\Leftrightarrow \mathbb{P}(\mathcal{N}(0, 1) > \sqrt{n}c) = \alpha$$

$$\Leftrightarrow 1 - \phi(\sqrt{n}c) = \alpha$$

$$\Leftrightarrow \phi(\sqrt{n}c) = 1 - \alpha$$

$$\Leftrightarrow \sqrt{n}c = \phi^{-1}(1 - \alpha) \Rightarrow c_\alpha = \frac{1}{\sqrt{n}} q_{1-\alpha}$$

(= quantile $1-\alpha$)

On a construit un test de niveau α avec $\mathcal{R} = \{ (X_1, \dots, X_n); \bar{X} > q_{1-\alpha}^{\text{norm}} / \sqrt{n} \}$

A.N: $\alpha = 5\% \Rightarrow q_{1-\alpha}^{\text{norm}} = 1.645$ $n = 100$

→ $c_\alpha = 0.1645$

expérience: → \bar{X}_{obs} = réalisation de \bar{X} sur mes données

si $\bar{X}_{\text{obs}} = 0.1 < c_\alpha$, on ne rejette pas H_0

si $\bar{X}_{\text{obs}} = 0.3 > c_\alpha \Rightarrow$ rejet de H_0

3) Construction d'un test:

(exemple)

α fixe:

- ① identifier le paramètre d'intérêt θ
définir les hypothèses H_0 et H_1 : $H_0: \theta \leq 0$ contre $H_1: \theta > 0$
- ② définir la forme de R $R = \{T > c\}$
forme de $H_1 \Rightarrow$ forme de $R = \{T > c\}$ ou bien $\{T < c\}$ (plus rarement)
trouver une statistique de test $T =$ version normalisée de $\hat{\theta} = \frac{\hat{\theta} - \theta}{\sqrt{v(\hat{\theta})}}$
- ③ Trouver le seuil c pour avoir un test de niveau α

Tests d'un paramètre gaussien

1) Résumé de la construction d'un test

$X = X_1, \dots, X_n$ i.i.d de loi \mathbb{P}_θ

1) Préciser les hypothèses testées

même test dans les deux cas

$$H_0: \theta = \theta_0 \text{ contre } H_1: \theta > \theta_0$$

↑ sur les paramètres du modèle

$$H_0: \theta \leq \theta_0$$

2) Statistique de test: $T(X)$: sous H_0 $T(X)$ calculable

version normalisée de l'estimateur de θ la loi de T sous H_0 permet de distinguer H_0 et H_1

→ $\mathcal{R} = \{T(X) > c\}$ (sous H_1 , si la loi de T s'écarte de H_0 vers la droite)

(si $H_1: \theta < \theta_0$ → $\mathcal{R} = \{T(X) < c\}$)

si test bilatère $H_1: \theta \neq \theta_0$ → $\mathcal{R} = \{|T(X)| > c\}$
 $= \{T(X) > c \text{ ou } T(X) < -c\}$

3) Règle de décision:

α niveau fixé.

condition de niveau $\sup_{\theta \leq \theta_0} \mathbb{P}_{H_0}(T(X) > c_\alpha) = \alpha$ (si loi de T sous H_0 est continue)

$\leq \alpha$ (si loi de T discrète)

engeneral, $\mathbb{P}_{\theta = \theta_0}(T(X) > c_\alpha) \xrightarrow{n \rightarrow +\infty} \alpha$ (si loi de T est asymptotique)

4) Application numérique

↑ par monotonie de la f.r. de la loi de T

calcul du seuil

de la réalisation de $T = T_{\text{obs}} = T(x)$ si $x = (x_1, \dots, x_n)$ réalisation de (X_1, \dots, X_n) dans notre expérience

- Si $T_{\text{obs}} > c_\alpha$ alors on rejette H_0 , avec un risque de se tromper de α
- si $T_{\text{obs}} \leq c_\alpha$, alors on conserve H_0 , avec un risque de se tromper inconnu (en général)

rq: le test de $H_0: \theta \geq \theta_0$ contre $H_1: \theta < \theta_0$ est le même que le test de $H_0: \theta = \theta_0$ contre $H_1: \theta < \theta_0$

$\mathcal{R} = \{T < c\}$

2) P-valeur

ex: X_1, \dots, X_n i.i.d. $\mathcal{N}(\theta, 1)$; test $H_0: \theta = 0$ contre $H_1: \theta > 0$

→ $T = \frac{\hat{\theta} - 0}{1/\sqrt{n}} = \sqrt{n}\hat{\theta} \underset{H_0}{\sim} \mathcal{N}(0, 1)$, Φ f.r. de $\mathcal{N}(0, 1)$

$\hat{\theta} = \bar{x}$

$\mathcal{R} = \{T > c\}$

condition de niveau: $\mathbb{P}_{\theta=0}(T > c) = \alpha \Leftrightarrow c = q_{1-\alpha}^{\text{norm}} = \Phi^{-1}(1-\alpha)$

↓
 $\mathcal{N}(0, 1)$

⇒ $\mathcal{R} = \{\sqrt{n}\hat{\theta} > \Phi^{-1}(1-\alpha)\}$

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

si $\alpha = 5\%$, rejette-t-on à 5% ? 10% ? 1% ?

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

DEF: si (X_1, \dots, X_n) i.i.d., $\mathcal{R} = \{T > c_\alpha\}$

Pour une réalisation $x = (x_1, \dots, x_n)$ de $X = (X_1, \dots, X_n)$, on appelle p-valeur du test du région de rejet \mathcal{R}

$(\psi(x) = 1_{\mathcal{R}}(x))$: $p\text{-val} = \inf\{\alpha \in [0, 1] \text{ tq } T(x) > c_\alpha\}$

$= \inf\{\alpha, H_0 \text{ est rejetée au niveau } \alpha\}$

↓ p-valeur / niveau de signifiativité / probabilité critique

exemple: $p\text{-val} = 1 - \Phi(\sqrt{n}\hat{\theta}^{\text{obs}})$
 $= 1 - \Phi(T^{\text{obs}})$
 $= 1 - \mathbb{P}(N(0,1) \leq T^{\text{obs}})$
 $p\text{-val} = \mathbb{P}(N(0,1) > T^{\text{obs}})$
 $= \mathbb{P}_{H_0}(T > T^{\text{obs}})$
 $N(0,1) \in \mathbb{R}$ loi de T

Généralisation (formule de calcul d'une p-valeur)

$T(x)$ statistique de test

- $R = \{T(x) > c\}$, alors p-valeur = $\mathbb{P}_{H_0}(T(x) > T^{\text{obs}})$
- $R = \{T(x) < c\}$ alors p-valeur = $\mathbb{P}_{H_0}(T(x) < T^{\text{obs}})$
- $R = \{|T(x)| > c\}$ alors p-valeur = $\mathbb{P}_{H_0}(|T(x)| > |T^{\text{obs}}|)$

Remarques:

① on considère le test bilatère $\theta = 0$ contre $\theta \neq 0$

$$\begin{aligned} \mathbb{P}_{H_0}(|T| > |T^{\text{obs}}|) &= \mathbb{P}_{H_0}(T > |T^{\text{obs}}| \text{ ou } T < -|T^{\text{obs}}|) \\ &= \mathbb{P}_{H_0}(T > |T^{\text{obs}}|) + \mathbb{P}(T < -|T^{\text{obs}}|) \\ &\stackrel{\sim N(0,1)}{=} 1 - \Phi(|T^{\text{obs}}|) + \Phi(-|T^{\text{obs}}|) \\ &= 2(1 - \Phi(|T^{\text{obs}}|)) \end{aligned}$$

symétrie

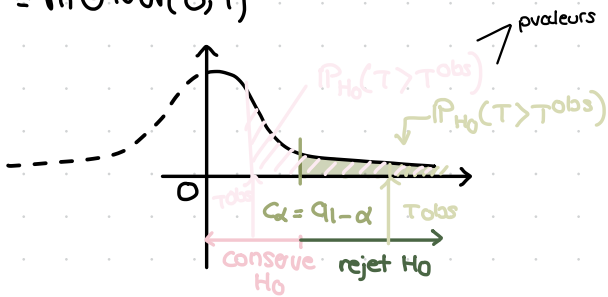
La p-valeur du test bilatère est le double de la p-valeur du test unitaire

② si la loi de T sous H_0 est discrète

Règle de décision avec la p-valeur

ex: $\theta = 0$ contre $\theta > 0$

$$T = \sqrt{n}\hat{\theta} \sim N(0,1)$$



\Rightarrow rejet $H_0 \Leftrightarrow p\text{-val} < \alpha$

3) Test sur un paramètre gaussien

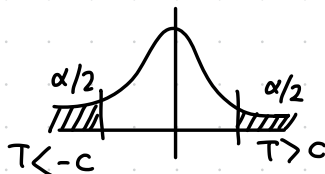
X_1, \dots, X_n i.i.d. $N(\mu, \sigma^2)$

- test sur μ
- test sur σ^2

a) $H_0: \mu = \mu_0$ contre $H_1: \mu \neq \mu_0$. $\bar{\mu} = \bar{X}$

$$T = \frac{\bar{X} - \mu_0}{S_n / \sqrt{n}} \underset{H_0}{\sim} \text{Student } (n-1)$$

loi exacte



H_1 bilatère $\Rightarrow R = \{|T| > c\}$

règle de décision? \rightarrow calcul de $c = c_\alpha$ avec la condition de niveau $\Rightarrow c_\alpha = q_{1-\alpha/2}^{(n-1)}$ quantile Student $(n-1)$

$$p\text{valeur} = P_{H_0}(|T| > |T^{\text{obs}}|)$$

$$= 2 P(T > |T^{\text{obs}}|)$$

symétrie
loi de Student
↓
Student(n-1)

si p valeur < α , rejet de H_0
> α , on conserve H_0

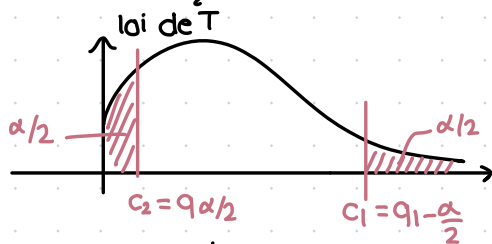
b) $H_0: \sigma^2 = \sigma_0^2$ contre $H_1: \sigma^2 \neq \sigma_0^2$

σ^2 inconnu donc on l'estime $\rightarrow s_n^2$ sans biais

$\hat{\sigma}_n^2$ EMV par le thm de la loi des estimateurs dans le modèle gaussien

$$0 \leq T = \frac{n \hat{\sigma}_n^2}{\sigma^2} = \frac{(n-1) s_n^2}{\sigma^2} = \frac{\sum_i (X_i - \bar{X})^2}{\sigma^2} \underset{H_0}{\sim} \chi^2(n-1)$$

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



On calcule T^{obs} : on rejette H_0 ssi $T^{\text{obs}} > q_{1-\frac{\alpha}{2}} \chi^2$ ou $T^{\text{obs}} < q_{\frac{\alpha}{2}} \chi^2$

p valeur ?

$$p\text{valeur} = 2 P(T > T^{\text{obs}}) \text{ ou } 2 P(T < T^{\text{obs}})$$

↑
par convention
p valeur < 5%

Test de Student (+ test)

(X_1, \dots, X_{n_1}) i. i. d. $\mathcal{N}(\mu_1, \sigma_1^2)$
 (Y_1, \dots, Y_{n_2}) i. i. d. $\mathcal{N}(\mu_2, \sigma_2^2)$
 Les 2 échantillons sont indépendants



hyp sup: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

On veut tester $H_0: \mu_1 = \mu_2$ contre $H_1: \mu_1 \neq \mu_2$
ou < ou > selon l'expérience
 $\Leftrightarrow \mu_1 - \mu_2 = 0$

exemple: efficacité d'un traitement qui diminue le taux de cholestérol
 $\mu_1 = \mu_2$ contre $\mu_2 < \mu_1$

① Statistique de test

$\mu_1 - \mu_2 = 0$?

idée: on estime $\hat{\mu}_1 - \hat{\mu}_2$ par $\bar{X} - \bar{Y}$

loi de $\bar{X} - \bar{Y}$? CL de gaussiennes indépendantes = $\mathcal{N}(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$

$E[\bar{X} - \bar{Y}] = \mu_1 - \mu_2$ par linéarité de E sp + (X_i) et (Y_i) i. d.

$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y})$ $\Delta!$

indep des 2 échantillons

$$= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

si σ^2 connue: $\frac{\bar{X} - \bar{Y} - \underbrace{(\mu_1 - \mu_2)}_{= 0 \text{ sous } H_0}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

en pratique σ^2 inconnue

→ on l'estime $\left| \begin{matrix} S_n^2 \\ \hat{\sigma}^2 \end{matrix} \right. *$

PROP: Sous les hypothèses de notre modèle.

* 2 échantillons gaussiens indépendants

* $\sigma_1^2 = \sigma_2^2$

alors

$S_n^2 = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right)$ est un estimateur sans biais de σ^2
2 relations d'indépendance

et $T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_n \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ a pour loi exacte la loi Student $(n_1 + n_2 - 2)$

pv: admise

* sous H_0 : $T = \frac{\bar{X} - \bar{Y}}{S_n \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ a pour loi Student $(n_1 + n_2 - 2)$

② région de rejet

$$R = \{ |T| > c \}$$

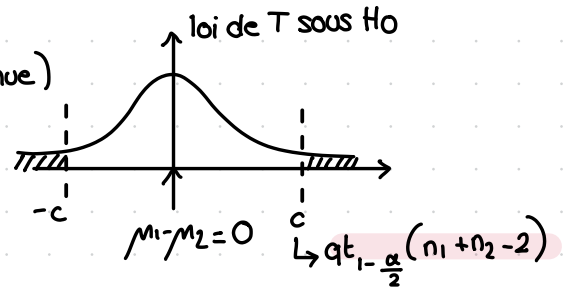
③ règle de décision

2 façons équivalentes:

a) calcul du seuil : α fixe'

b) condition de niveau: $\mathbb{P}_{\mu_1 = \mu_2} (|T| > c_\alpha) = \alpha$ (T de loi continue)

$$\Rightarrow c_\alpha = qt_{1-\frac{\alpha}{2}}(n_1+n_2-2)$$



c) calcul de la p-valeur:

$$p\text{-valeur} = \mathbb{P}_{H_0} (|T| > |T_{\text{obs}}|)$$

car loi continue

$$= 2\mathbb{P}(T > |T_{\text{obs}}|) = 2(1 - F(|T_{\text{obs}}|))$$

Student(n1+n2-2)

loi de T
symétrique

Application numérique:

$$n_1 = 12, \bar{x} = 1.50, S_x^{\text{obs}} = 0.95 \quad \left(= \sqrt{\frac{1}{n_1-1} \sum (x_i - \bar{x})^2} \right)$$

carte type observée

$$n_2 = 8, \bar{y} = 2.35, S_y^{\text{obs}} = 1.35$$

$$T^{\text{obs}} = \frac{\bar{x} - \bar{y}}{S_{(x,y)}^{\text{obs}} \sqrt{\frac{1}{12} + \frac{1}{8}}}$$

$$S_{(x,y)}^2 = \frac{1}{n_1+n_2-2} \cdot \left(\underbrace{\sum (x_i - \bar{x})^2}_{(n_1-1)(S_x^{\text{obs}})^2} + \underbrace{\sum (y_i - \bar{y})^2}_{(n_2-1)(S_y^{\text{obs}})^2} \right) \quad \leadsto T^{\text{obs}} = 1.801$$

$$\alpha = 5\% \rightarrow c_\alpha = qt_{0.975}(12+8-2) = 2.101$$

$$p\text{-valeur} = 2(1 - F(1.801))$$

Student(12+8-2)

$$= 2 \times (0.05) = 0.10 > \alpha, \text{ on ne rejette pas } H_0$$

a) $T^{\text{obs}} = 1.801 < c_\alpha = 2.101$ on ne rejette pas H_0 , les 2 éch. n'ont pas des moyennes différentes

exam:

loi de test d'un modèle gaussien)

2-3 question de cours (résultat / exercice traité en cours.

exercice de test

exercice de synthèse