Statistiques Inférentielles

Programme :

0 - Introduction → Définition

I - Echantillonnage → Cadre probabiliste.

II - Estimation

III - Comportement asymptotique.

IV - Généralités sur les tests.

V - Un peu de stats non paramétriques.

Chapitre 0 : Introduction

Définition : On dit qu'on se trouve face à un problème statistique si :

- On est confronté à des éventualités, en nb fini ou infini, dont on sait qu'au moins une est vraie
- On doit choisir une de ces éventualités
- En se basant sur les résultats d'une expérience aléatoire (à définir après)

Exemple : On veut étudier la distribution du montant de sinistres auto.

On suppose que la distribution est une Gamma.

Les éventualités sont l'ensemble des paramètres possibles pour la loi Gamma i.e. $\mathbb{R}_+^* \times \mathbb{R}_+^*$

On va choisir les bons paramètres en se basant sur les montants des sinistres observés.

Définition : On appelle modèle statistique le triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ où

- $\mathcal{X}$ est l'ensemble fondamental ou l'espace des résultats i.e l'endroit où vivent les observat°.
- $\mathcal{A}$ est la tribu associée à $\mathcal{X}$ (a priori on ne s'en sert pas).
- $\mathcal{P}$ est une famille de probabilités définies sur $(\mathcal{X}, \mathcal{A})$ qui contient la distribution qui a généré les observations.

On dira que $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ est un modèle statistique paramétrique si :

$$\exists p \in \mathbb{N} \text{ tq } \mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$$

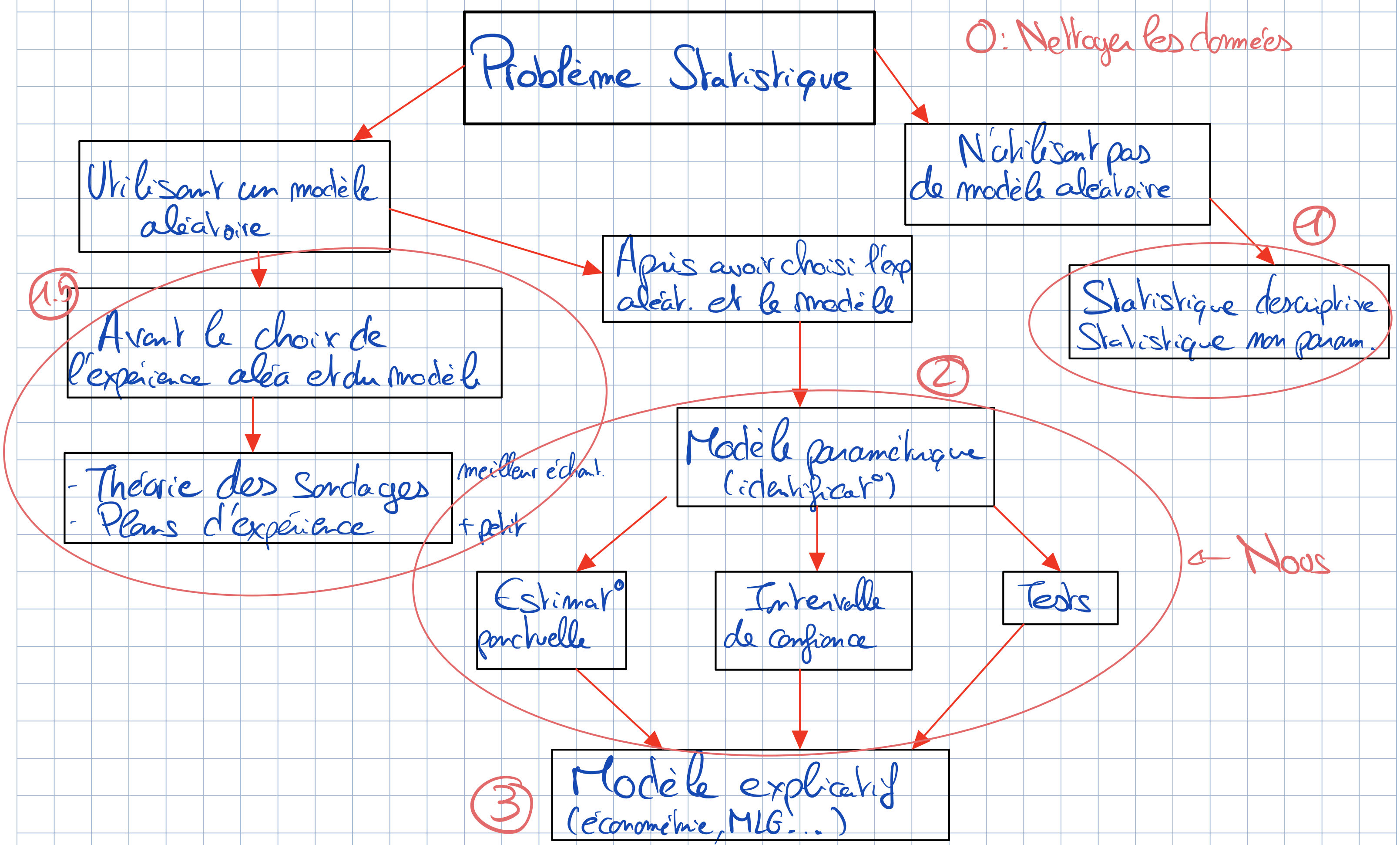
distribut° indexée par un paramètre

espace des paramètres i.e là où vit le paramètre

Exemple : $\mathcal{P} = \{ \underbrace{\Gamma(a,b)}_{P_\theta} : \underbrace{(a,b)}_{\theta} \in \underbrace{\mathbb{R}_+^* \times \mathbb{R}_+^*}_{\Theta} \subseteq \mathbb{R}^2 \}$

Remarque : On va distinguer 3 types de problèmes :

- estimation ponctuelle : on se place dans le modèle paramétrique, les "éventualités" sont représentées par θ et on veut choisir θ.
- estimation ensembliste ou par intervalle de confiance : on va choisir un ensemble ou un intervalle qui va contenir le "vrai" paramètre avec une proba élevée.
- Tests : On fait une hypothèse sur la distribution et on compare au résultat obtenu sur les observations. On doit "juste" choisir entre vrai ou faux (pour l'hypothèse).



Chapitre I : Echantillonage.

I - Modèle statistique

Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire de n VA.

$X_i : (\Omega, \mathcal{B}, \mathbb{P}_X) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ $\quad X : (\Omega, \mathcal{A}, \mathbb{Q}) \longrightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_X)$

$\mathbb{P}_X$ est la loi de proba image du vecteur X.

En général, on ne connait pas $\mathbb{P}_X$ et on veut l'approximer le mieux possible.

Hypothèse fondamentale : La proba $\mathbb{P}_X$ appartient à une famille de lois de probas (déf sur $(\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$ notée $\mathcal{P}_X$

Modèle d'échantillonage : C'est le triplet $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}_X)$

espace des résultats i.e l'espace d'arrivée du vecteur ⚠ peut être $\mathbb{N}^*$

Tribu

Famille des lois de proba sur $(\mathbb{R}^*, \mathcal{B}(\mathbb{R}^*))$ qui contient $\mathbb{P}_X$

Remarque : Le choix de la famille $\mathcal{P}_X$ est crucial (risque de modèle)

On se base sur des études précédentes, sur la stat. non paramétrique ou sur un expert.

Définition : Une série statistique $(x_1, \ldots, x_n)$ est une réalisat° du vecteur aléatoire i.e $\exists \omega \in \Omega \quad (x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$.

Le modèle d'échantillonage indépendant est un cas particulier où on considère $X_1, \ldots, X_n$ indep et identiquement distribuées.

On a les propriétés suivantes :

- les va sont identiquement distrib., elles ont m̂ loi de $\mathbb{P}_{X_i}$ ne dépend pas de i et de $\forall i \in [\![1, n]\!] \quad \mathbb{P}_{X_i} = \mathbb{P}$
- les va sont indep, la loi du vecteur est la loi produit $\mathbb{P}_X = \mathbb{P}^{\otimes n}$

De manière abusive, on note le modèle d'échantillonage indep.

$$\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}\right)^n$$

à l'extérieur pour mq les va sont i.i.d et c'est la taille de l'échantillon.

espace des résultats — Tribu — famille de lois de probas qui contient la "vraie" proba $\mathbb{P}$

Le modèle d'échantillonage (indep) paramétrique est le cas où la famille $\mathcal{P}$ est indicée par un paramètre θ

$$\mathcal{P} = \left\{ \mathbb{P}_\theta,\ \theta \in \Theta \subseteq \mathbb{R}^p \right\}$$

dimens° du paramètre

loi de proba qui correspond à la valeur θ du paramètre — param. — espace des param. i.e les valeurs possibles des param.

Remarque : Sinon on a un modèle non paramétrique

Exemple : On peut prendre un modèle $\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}\right)^n$

où $\mathcal{P} = \left\{ f\lambda : f \geq 0 \text{ tq } \int f \, d\lambda = 1 \right\}$

= ensemble des distributions qui admettent une densité p/r à Lebesgue

Exemple : Si on considère des va $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

Le modèle stat est

$$\Big(\mathbb{R}, \mathcal{B}(\mathbb{R}), \underbrace{\mathcal{N}(\mu,\sigma^2)}_{\mathbb{P}_\theta},\ \underbrace{(\mu,\sigma^2)}_{\theta} \in \underbrace{\mathbb{R} \times \mathbb{R}_*^+}_{\Theta} \subseteq \mathbb{R}^{\overset{p}{2}}\Big)^n$$

Hypothèse fondamentale : Le modèle est dominé par une mesure μ

Rappel (ou pas) : Soit $\mathbb{P}$ et μ deux mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on dit que $\mathbb{P}$ est absolum^t continue p/r à μ ou μ domine $\mathbb{P}$ (noté $\mathbb{P} \ll \mu$)

si $\forall A \subset \mathcal{B}(\mathbb{R}^*)$, $\mu(A) = 0 \Longrightarrow \mathbb{P}(A) = 0$

Rappel: Radon-Nikodym : Si $\mathbb{P} \ll \mu$, alors $\exists$ une va X tq

$$\mathbb{E}^{\mathbb{P}}(X) = 1 \text{ et } \mathbb{P}(A) = \mathbb{E}[X \mathbb{1}_A].$$

Ici X est juste une fonct° qui intègre à 1 p/r à μ

On dit que le modèle est dominé par une mesure μ si ttes les probas $\mathbb{P}_X$ dans $\mathcal{P}_X$ sont dominées par μ i.e $\forall \mathbb{P}_X \in \mathcal{P}_X$, $\mathbb{P}_X \ll \mu$.

Dans ce cas, $\exists$ une fonct° f_X (Radon-Nikodym) tq $\mathbb{P}_X = f_X \cdot \mu$ (i.e $\mathbb{P}(A) = \int_A f_X \, d\mu$).

La fonct° f_X est la vraisemblance du modèle (likehood en anglais)

notée $L_X(x) = f_X(x)$

Traduct° - Exemple: Si le modèle est dominé par λ alors la loi $\mathbb{P}_X$ admet une densité p/r à λ et $L_X(\underset{\in \mathbb{R}^n}{x}) = f_X(\underset{\in \mathbb{R}^n}{x})$ densité du vecteur X.

Dans le cas d'un modèle d'échantillonage indep. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta)^n$ est dominé ssi $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\theta, \theta \in \Theta)$ est dominé

Dans ce cas là, on a

$$\underbrace{L_X(x_1, \ldots, x_n, \theta)}_{\text{densité de } X} = \underbrace{f_{X_1}(x_1) \ldots f_{X_n}(x_n)}_{\text{les } f_{X_i} \text{ sont égaux}} = \underbrace{f_\theta(x_1) \ldots f_\theta(x_n)}_{\text{produit des densités de } X_i} = \prod_{i=1}^{n} f_\theta(x_i)$$

Dans le cas de lois discrètes, on peut construire une mesure μ (mesure de comptage) tq

$$\mu = \sum_{n \in \mathbb{N}} \delta_n$$

mesure de Dirac en n i.e $\delta_n(A) = \begin{cases} 1 & \text{si } n \in A \\ 0 & \text{sinon} \end{cases}$

et tq $\mathbb{P}_\theta \ll \mu$ (ex: Poisson, ...)

la densité est la fonct° de masse et on a $L_X(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i)$

Remarque: La vraisemblance du modèle $L_X(x_1, \ldots, x_n, \theta)$ est l'outil de base

Dans le cas discret $L(x_1, \ldots, x_n, \theta)$ est la proba d'observer $x_1, \ldots, x_n$ sous le paramètre θ

Si on fixe $x_1, \dots, x_n$, + $L(x_1, \dots, x_n, \theta)$ est élevée, + la proba est élevée et + le paramètre θ est "vraisemblable"

On va souvent utiliser la log-vraisemblance i.e. $\ell(x_1, \dots, x_n, \theta) = \log L(x_1, \dots, x_n, \theta)$

II - Définit° d'une statistique

On se place dans le modèle général $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}_x)$ une statistique est une VA définie sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

Idée : Dans un modèle paramétrique, on veut que la statistique apporte de l'informat° pour approcher θ. Elle ne doit pas dépendre (son écriture) de la famille $\mathcal{P}_\theta$

D'habitude on note la statistique T_n.

Définit° : Une statistique est une fonct° de $X_1, \dots, X_n$

$$T_n = h(X_1, \dots, X_n)$$

Pas la "vraie" définit° d'une stat

Exemple : On prend $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$

Le modèle stat associé est $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(\mu, 1), \mu \in \mathbb{R} \subseteq \mathbb{R}^1)^n$

On peut prendre $T_n = \underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i}$

L'écriture ne dépend pas de μ

On a :
- la loi des grands nombres dit que $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_i] = \mu$
- On peut calculer la loi de T_n.

T_n suit une loi Gaussienne comme CL de Gaussiennes indep

On a $\mathbb{E}[T_n] = \mathbb{E}[\frac{1}{n} \sum_{i=1}^{n} X_i] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu$

et $V(T_n) = V(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n}$

$\Rightarrow T_n \sim \mathcal{N}(\mu, \frac{1}{n})$

Remarque : On distingue la statistique $T_n = \underbrace{h(X_1, -, X_n)}_{VA}$ et la statistique observée $\underbrace{h(x_1, -, x_n)}_{\in \mathbb{R}}$

Rappel : On a $x_1, \dots, x_n$ données

→ Statistiques descriptives : On a des observat° et on décrit les données

→ Statistiques Inférentielles : On considère les observat° comme des réalisat° de v.a. et on veut des infos sur ces v.a. ⇒ Modèle proba.

Notre modèle : On considère $X_1, -, X_n$ comme des v.a. iid avec $X_i \sim \mathbb{P}$

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})^n$$

espace des résultats = l'ensemble des résultats possibles pour X_i

Tribu

Famille de proba qui contient la "vraie" proba $\mathbb{P}$

Dans le cours, on va supposer :

- $\mathcal{P} = \mathcal{P}_\theta = \{ \mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p \}$ et qui est de dimens° p.

La proba est déterminée par un paramètre θ

⇒ qui prend ses valeurs dans Θ

- Le modèle est dominé par μ i.e $\forall \mathbb{P} \in \mathcal{P}, \mathbb{P} \ll \mu$

Critère : Si les distribut° $\mathbb{P}$ admettent une densité ou une fonct° de masse, le modèle est dominé

On appelle vraisemblance, la densité (p/r à la mesure qui domine) du vecteur $(X_1, -, X_n)$

Dans notre cas, comme $X_1, -, X_n$ iid, on a :

$$\underbrace{L(x_1, -, x_n, \theta)}_{\substack{\text{vraisemblance du modèle} \\ = \text{densité de } (X_1, -, X_n) \text{ sous} \\ \text{l'hypothèse } \theta}} = \prod_{i=1}^{n} \underbrace{f_\theta(x_i)}_{\substack{\text{vraisemblance individuelle} \\ = \text{densité de } X_i \text{ sous} \\ \text{l'hypothèse } \theta}}$$

Définition (pas bonne) : Une statistique est une fonct° de $X_1, -, X_n$ dont l'écriture ne dépend pas de θ.

Un estimateur de $g(\theta)$ est une statistique à valeurs dans Θ. On le note souvent T_n

Remarques :

* Un estimateur ne dépend pas de θ (son écriture ne dépend pas de θ) mais sa distribut° dépend de θ.

* On peut parfois estimer une fonct° de paramètre θ (par exemple la $VAR_{0,05}$ i.e le quantile d'ordre 0,995), mais pour nous, souvent, $g(\theta)=\theta$

II - Notions de base sur les estimateurs

On se place dans le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p)^n$

i.e $X_1, \dots, X_n$ iid tq $X_i \sim \mathbb{P}_\theta$ et $\mathbb{P}_\theta \in \mathcal{P}_\theta$.

On prend un estimateur T_n, on veut qu'il apporte des infos sur $g(\theta) \longrightarrow$ on veut certaines notions sur T_n

Définition : On appelle biais de l'estimateur T_n de $g(\theta)$ la quantité

$$B_{g(\theta)}(T_n) = \mathbb{E}[T_n] - g(\theta)$$ c'est une fonct° de θ.

On dit que T_n est sans biais pour $g(\theta)$ si $B_{g(\theta)}(T_n) = 0$

Traduct° : Si T_n est sans biais alors il donne "en moyenne" la vraie valeur de $g(\theta)$

On dit que T_n est asymptotiquement sans biais pour $g(\theta)$ si $B_{g(\theta)}(T_n) \xrightarrow[n\to\infty]{} 0$

Définit° : On dit que T_n est convergent pour $g(\theta)$ si T_n converge en proba vers $g(\theta)$

i.e. $T_n \xrightarrow{\mathbb{P}} g(\theta)$

i.e $\forall \varepsilon > 0, \mathbb{P}(\| T_n - g(\theta) \| > \varepsilon) \longrightarrow 0$

ou

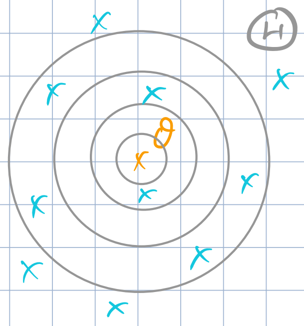
$\forall \varepsilon > 0, \mathbb{P}(\| T_n - g(\theta) \| < \varepsilon) \longrightarrow 1$

Critère de convergence :

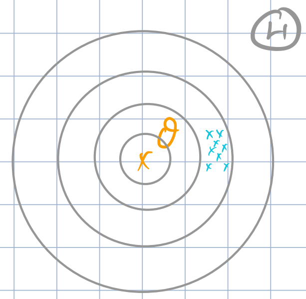
Si T_n est asymptotiquem^t sans biais pour $g(\theta)$ (i.e $\mathbb{E}[T_n] \xrightarrow[n\to\infty]{} g(\theta)$ et $V(T_n) \xrightarrow[n\to\infty]{} 0$)

alors T_n est convergent pour $g(\theta)$.

Comparaison d'estimateurs :



$T_n^{(1)}$ est sans biais mais grande variance



$T_n^{(2)}$ a du biais mais une petite variance

Pour comparer deux estimateurs du m̂ paramètre $g(\theta)$, on utilise le risque quadratique.
Il est défini par $\mathcal{R}(T_n, g(\theta)) = \mathbb{E}[(T_n - g(\theta))^2]$

Propriété : $\mathcal{R}(T_n, g(\theta)) = B_{g(\theta)}(T_n)^2 + V(T_n)$

Remarques :

* Si on a deux estimateurs, on choisira, en théorie, celui qui a le plus petit risque quadratique
* De manière générale, on essaye de minimiser le risque quadratique (en minimisant le biais car la variance est bornée inf. (Borne de Cramer-Rao)

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{U}[0;\theta]$.

2 façons d'estimer θ.

- On a la loi des grands nombres $\overline{X}_n \longrightarrow \mathbb{E}[X]$
  ici on a $\overline{X}_n \longrightarrow \frac{\theta}{2}$
  donc $2\overline{X}_n$ doit être un estimateur correct de θ.

- θ est la valeur max théorique de X_i, donc $\max\limits_{i \in [\![1,n]\!]} X_i$ doit être un estimateur pas dégueulasse de θ.

Quel estimateur est préférable au sens du risque quadratique ?

On pose $T_n = 2\overline{X}_n$ et $T_{max} = \max\limits_{i \in [\![1,n]\!]} X_i$

Etude de T_n

- Biais de T_n :

$$\mathbb{E}[T_n] = \mathbb{E}[2\bar{X}_n] = 2\,\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{2}{n}\sum_{i=1}^{n}\underbrace{\mathbb{E}[X_i]}_{\frac{\theta}{2}} = \theta$$

$$B_\theta(T_n) = \mathbb{E}[T_n] - \theta = 0$$

- Variance de T_n :

$$V(T_n) = V(2\times\frac{1}{n}\sum_{i=1}^{n} X_i) \overset{iid}{=} \frac{4}{n^2}\sum_{i=1}^{n}\underbrace{V(X_i)}_{=\frac{\theta^2}{12}} = \frac{\theta^2}{3n^2} \xrightarrow[n\to\infty]{} 0$$

cl : T_n est convergent pour θ ou $T_n \xrightarrow{\mathbb{P}} \theta$

Risque Quadratique de T_n :

$$R(T_n,\theta) = \mathbb{E}[(T_n-\theta)^2] = B_\theta(T_n)^2 + V(T_n) = V(T_n) = \frac{\theta^2}{3n}$$

Etude de T_{max} :

$T_{max} = \max_{i\in[\![1,n]\!]} X_i$ $\qquad$ $\mathbb{E}(\max X_i) \neq \max \mathbb{E}[X_i]$.

Loi de T_{max} :

$$\mathbb{P}(\max_{i\in[\![1,n]\!]} X_i \leq x) = \mathbb{P}(X_1\leq x, \dots, X_n \leq x)$$

$$\overset{indep}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \leq x)$$

$$F_{max}(x) \overset{i.d}{=} \mathbb{P}(X_1\leq x)^n = \left(\frac{x}{\theta}\right)^n \quad si\ x\in[0;\theta]$$

Densité de T_{max} :

$$f_{max}(x) = \frac{nx^{n-1}}{\theta^n}\,\mathbb{1}_{[0;\theta]}(x)$$

Biais de T_{max}

$$\mathbb{E}[T_{max}] = \int x f_{max}(x)dx = \int_0^\theta x\frac{nx^{n-1}}{\theta^n}dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta$$

T_{max} est asympt. sans biais car $\mathbb{E}[T_{max}] \xrightarrow[n\to\infty]{} \theta$

$$B_\theta(T_{max}) = \mathbb{E}[T_{max}] - \theta = -\frac{1}{n+1}\theta$$

Variance de T_{max} :

$$\mathbb{E}[T_{max}^2] = \int x^2 f_{max}(x)dx = \int_0^\theta \frac{nx^{n+1}}{\theta^n}dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2}$$

$$V[T_{max}] = \mathbb{E}[T_{max}^2] - \mathbb{E}[T_{max}]^2$$
$$= \frac{n\theta^2}{n+2} - \frac{(n\theta)^2}{(n+1)^2} = \frac{n\theta^2}{(n+1)^2(n+2)}$$

Risque Quadratique de T_{max} :

$$R(T_{max}, \theta) = B_\theta(T_{max})^2 + V(T_{max}) = \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+1)^2(n+2)}$$
$$= \frac{2\theta^2}{(n+1)(n+2)}$$

On avait $R(T_n, \theta) = \dfrac{\theta^2}{3n}$

Pour $n=1$ ou $n=2$, $R(T_n,\theta) = R(T_{max},\theta)$

si $n \geq 3$, $R(T_{max},\theta) \leq R(T_n,\theta)$ $\Rightarrow$ On préfère T_{max} au sens du risque quadratique.

On a $\mathbb{E}[T_{max}] = \dfrac{n}{n+1}\theta$

<u>Remarque</u> : Si on a un estimateur T_n de θ tq $\mathbb{E}[T_n] = a\theta + b$

alors $T_n^* = \dfrac{T_n - b}{a}$ est sans biais pour θ.

On pose $T^*_{max} = \dfrac{n+1}{n} T_{max} = \dfrac{n+1}{n} \max\limits_{i \in [\![1,n]\!]} X_i$

On a T^*_{max} sans biais pour θ car $\mathbb{E}[T^*_{max}] = \dfrac{n+1}{n}\mathbb{E}[T_{max}] = \theta$

On a

$$V(T^*_{max}) = V\left(\frac{n+1}{n}T_{max}\right) = \frac{(n+1)^2}{n^2}V(T_{max}) = \frac{\theta^2}{n(n+2)} \xrightarrow[n\to\infty]{} 0$$

et on obtient $R(T^*_{max},\theta) = \underbrace{B_\theta(T_{max})^2}_{=0} + V(T^*_{max}) = \dfrac{\theta^2}{n(n+2)}$

Est-ce qu'on a $R(T^*_{max},\theta) \leq R(T_{max},\theta)$?

On calcule
$$\frac{R(T^*_{max},\theta)}{R(T_{max},\theta)} = \frac{\frac{\theta^2}{n(n+2)}}{\frac{2\theta^2}{(n+2)(n+1)}} = \frac{n+1}{2n} \leq 1.$$

$\Rightarrow$ <u>T^*_{max} est préférable au sens du risque quadratique</u>

Cas particulier : moyenne et variance aléatoire

On se place dans le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p)^n$

On suppose que $\mathbb{P}_\theta$ admet une espérance μ et une variance $\sigma^2 < \infty$

On a des estimateurs "naturels" de μ et σ^2 (qui viennent de la stat. descriptive).

Si $(X_1, \dots, X_n)$ est le vecteur associé au modèle, on appelle moyenne aléatoire la v.a

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

On appelle variance aléatoire la stat

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

On va utiliser la variance corrigée définie par $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$

Théorème :

* $\overline{X}_n$ est sans biais et convergent pour μ
* V_n est asympt. sans biais et convergent par fois pour σ^2
* S_n^2 est sans biais et convergent pour σ^2.

III - Statistiques Gaussienne

Définition : Un vecteur $X = (X_1, \dots, X_n)$ est Gaussien d'espérance μ et de matrice de variance Σ si sa densité est :

$$f(x) = \frac{1}{(\det \Sigma)^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu)\right)$$

Caractérisat° : $X = (X_1, \dots, X_n)$ est Gaussien $\Longleftrightarrow$ $\forall \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n, \sum_{i=1}^{n} \lambda_i X_i$ Gaussien

Propriété : Soit $X = (X_1, \dots, X_n)$ Gaussien,

alors, $\operatorname{Cov}(X_i, X_j) = 0 \Longleftrightarrow X_i \perp\!\!\!\perp X_j$

Définit° : Une var suit une loi Gamma $\Gamma(a,\lambda)$ si sa densité vaut :

$$f_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbb{1}_{\{x \geq 0\}}$$

⚠ pas la m̂ paramétrisat° que sur Wikipedia

Propriété :

* $\mathcal{E}(\lambda) = \Gamma(1,\lambda)$
* Si $X \sim \Gamma(a,\lambda)$ et $Y \sim \Gamma(b,\lambda)$ et $X \perp\!\!\!\perp Y$
 alors $\Gamma(a,\lambda) + \Gamma(b,\lambda) = \Gamma(a+b,\lambda)$

Définit° : Si $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{N}(0,1)$ alors,

$$Z = \sum_{i=1}^{n} X_i^2 \sim \chi^2_n \quad \text{(Khi deux à } n \text{ degrés de libertés)}$$

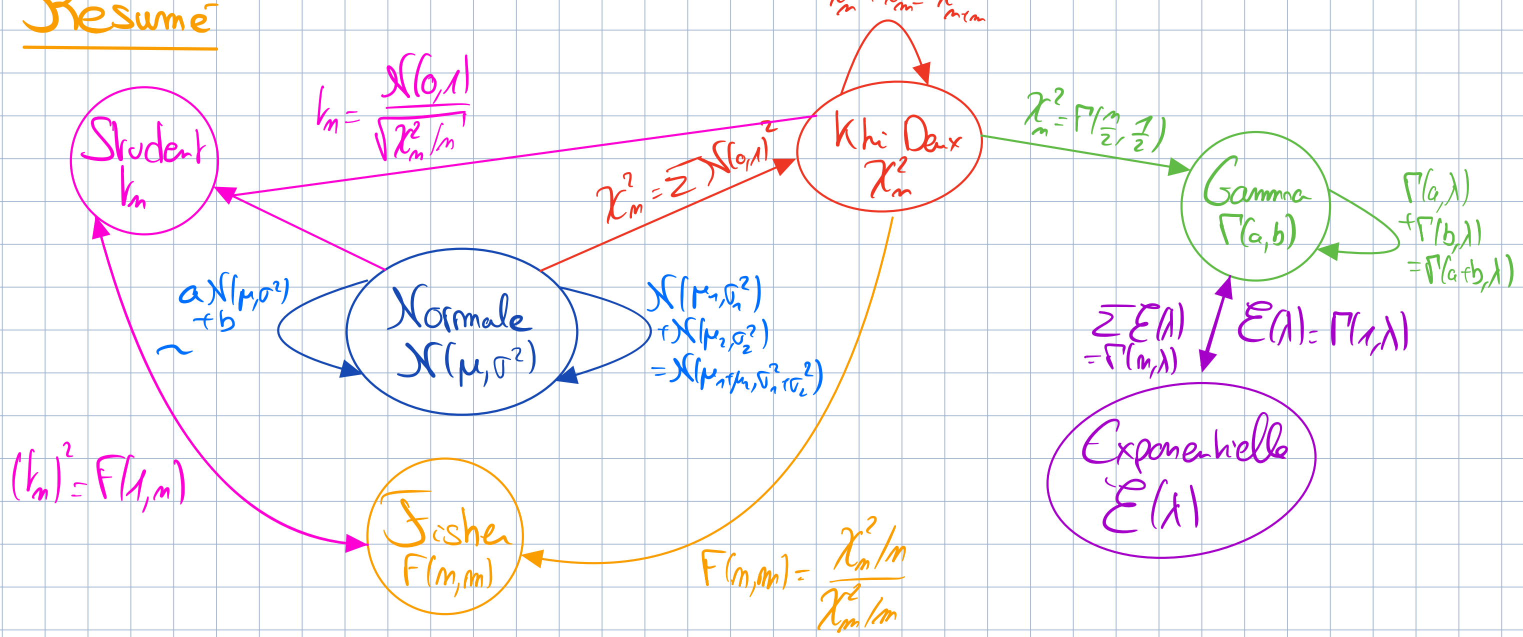
Propriété : $\chi^2_n = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

Définit° :

* Soit $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_n$ et $X \perp\!\!\!\perp Y$,
 alors, $Z = \dfrac{X}{\sqrt{Y/n}} \sim t_n$ (Student à n degrés de liberté)

* Soit $X \sim \chi^2_n$, $Y \sim \chi^2_m$ et $X \perp\!\!\!\perp Y$,
 alors $F = \dfrac{X/n}{Y/m} \sim F(n,m)$ (Fisher à n et m degrés de libertés)

Résumé



Rappel : $X_1, \dots, X_n$ iid tq $X_i \sim \mathbb{P}_\theta$

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p)^n$$

Statistique = T_n = fonct° de $X_1, \dots, X_n$ (sans θ)
Estimateur = Statistique à valeurs dans Θ

Critères de qualité d'un estimateur T_n de $g(\theta)$

Est-ce qu'en moyenne on vise la bonne valeur ?

⊛ **Biais** $B_{g(\theta)}(T_n) = \mathbb{E}[T_n] - g(\theta)$

T_n est (asympt) sans biais si $B_{g(\theta)}(T_n) = 0$ $\left(\xrightarrow[n\to\infty]{} 0\right)$

Est-ce qu'on gagne en précis° qd $n \to \infty$?

⊛ **Convergence :** $T_n \xrightarrow{\mathbb{P}} g(\theta)$

ou $B_{g(\theta)}(T_n) \longrightarrow 0$ et $V(T_n) \longrightarrow 0$

⊛ **Risque Quadratique moyen**

$$\mathcal{R}(T_n, g(\theta)) = \mathbb{E}[(T_n - g(\theta))^2] = B_{g(\theta)}(T_n)^2 + V(T_n)$$

mélange des 2 autres critères

Ça sert à :

- Comparer 2 estimateurs (on prend celui qui a le + petit RQM)
- On essaye de minimiser le RQM d'un estimateur

Propriété :

* Si $\theta = \mathbb{E}[X]$ alors $\overline{X}_n$ est un estimateur sans biais convergent pour θ.
* Si $\theta = V(X)$ alors $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ est un estimateur sans biais convergent pour θ

Statistiques Gaussiennes

Théorème (Cochran) : Soit $X_1, \dots, X_n$ des v.a. iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

* $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
* $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$ avec $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$
* $\overline{X}_n \perp\!\!\!\perp S_n^2$

Remarque : Si on connait la valeur de μ, on peut estimer σ^2 par $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$

Si on pose $Y_i=(X_i-\mu)^2$, on a $\mathbb{E}[Y_i]=\mathbb{E}[(X_i-\mu)^2]=\sigma^2$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i$ est convergent et sans biais pour σ^2

$V_n^* = \frac{1}{n}\sum (X_i-\mu)^2$

Rappel : $W=\sum_{i=1}^{n} Z_i^2$ avec $Z_i \overset{iid}{\sim} \mathcal{N}(0,1)$ alors $W\sim \chi_n^2$

Ici : $X_i\sim\mathcal{N}(\mu,\sigma^2) \Rightarrow \frac{X_i-\mu}{\sigma}\sim\mathcal{N}(0,1)$

$$\Rightarrow \underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{n}{\sigma^2}\times\frac{1}{n}\sum(X_i-\mu)^2=\frac{n}{\sigma^2}V_n^*}\sim\chi_n^2$$

Cl : $\frac{n}{\sigma^2}V_n^*\sim\chi_n^2$

Début de preuve :

On part du fait que $(X_1-\bar{X}_n, \dots, X_n-\bar{X}_n, \bar{X}_n)$ forme un vecteur Gaussien car toute CL des composantes est Gaussienne

On a $\mathrm{Cov}(X_1-\bar{X}_n,\bar{X}_n)=\mathrm{Cov}(X_1,\bar{X}_n)-\mathrm{Cov}(\bar{X}_n,\bar{X}_n)$

$$=\frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathrm{Cov}(X_1,X_i)}_{=\begin{cases}V(X_1) \text{ si } i=1\\ 0 \text{ sinon}\end{cases}}-\underbrace{V(\bar{X}_n)}_{=\frac{\sigma^2}{n}}=0.$$

$\Rightarrow \forall j\in[\![1,n]\!],\ X_j-\bar{X}_n \perp\!\!\!\perp \bar{X}_n$

Chapitre II : Estimation

On reste dans le cadre du modèle d'échantillonage paramétrique

$$\left(\underbrace{\mathbb{R}, \mathcal{B}(\mathbb{R})}_{\text{ou } (\mathbb{N}, \mathcal{P}\text{arties}(\mathbb{N}))}, P_\theta, \theta \in \Theta \subseteq \mathbb{R}^{1}\right)^n$$

On travaille (au début) avec un paramètre unidimensionnel

et on a $(X_1, -, X_n)$ le vecteur associé au modèle.

- Soit $f(x,\theta)$ la densité (density) de la distribut° des X_i (ou la fonct° de masse (mass funct°) si la loi des X_i est discrète).
- Soit $L(x_1, -, x_n, \theta)$ la vraisemblance (likelihood) du modèle

On a $\underbrace{L(x_1, -, x_n, \theta)}_{\text{densité de } (X_1,-,X_n)}$ $= \prod_{i=1}^{n} f(x_i, \theta)$ car le modèle est indep. ($\perp\!\!\!\perp$)

densité de $(X_1,-,X_n)$ p/r à la mesure dominante

Soit $\mathcal{X} \subset \mathbb{R}$ est le support de $f(x,\theta)$

$$\mathcal{X} = \{x \in \mathbb{R} \mid f(x,\theta) > 0\}$$

__Exemple__: - Si $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mathcal{X} = \mathbb{R}$

- Si $X_i \sim \mathcal{U}[0;\theta]$, $\mathcal{X} = [0,\theta]$
- Si $X_i \sim \mathcal{P}(\theta)$, $\mathcal{X} = \mathbb{N}$

Le support de L est $\mathcal{X}^n$

I - Hypothèses fondamentales sur f.

On va faire des hypothèses de régularité sur f.

⚠ Si les hypothèses ne sont pas vérifiés, on peut quand m̂ faire de l'estimat° mais avec d'autres techniques (moments, moindres carrés, ...).

Hypothèses de régularité :

(H1) : $\mathcal{X} \perp\!\!\!\perp \theta$) C'est l'hypothèse qui va foirer le + souvent

(H2) : Θ est un ouvert (le fait d'avoir $f(x,\theta) > 0$, $\forall (x,\theta) \in \mathcal{X} \times \Theta$ permet de ne pas avoir de pbs de différentiabilité à la frontière.)) Si Θ est fermé, on l'ouvre

(H3): $\frac{\partial}{\partial\theta} f(x,\theta)$ et $\frac{\partial^2}{\partial\theta^2} f(x,\theta)$ sont définies $\forall (x,\theta) \in \mathcal{X}\times\Theta$

Ça permet de dériver f deux fois p/r à θ sous le signe $\int$

(H4): Les dérivés premières et secondes de f p/r à θ sont dominés par des fonct° μ-intégrables sur tout compact K inclus dans Θ ($K \subset \Theta$).
Pour tout compact $K \subset \Theta$, il existe deux fonct° μ-intégrables $\varphi(x)$ et $\psi(x)$ tq $\forall \theta \in K$ et pour presque tout $x \in \mathcal{X}$

$$\left|\frac{\partial}{\partial\theta} f(x,\theta)\right| < \varphi(x) \text{ et } \left|\frac{\partial^2}{\partial\theta^2} f(x,\theta)\right| < \psi(x).$$

II - Informat°

Informat° : important
= tout ce qu'on sait sur θ (avec $X_1, -, X_n$)

Définit° (Score) :

Score d'une V.A. X_i

On appelle fonct° Score ou Score la fonct° définie par :

$$S : \mathcal{X}\times\Theta \longrightarrow \mathbb{R}$$
$$(x,\theta) \longmapsto \frac{\partial}{\partial\theta} \ln(f(x,\theta))$$

Remarques :

* Cette fonct° Score est bien définie si (H1), (H2) et (H3) sont vérifiées
* Elle intervient souvent en estimat°
* On peut définir le Score du modèle i.e. de $(X_1, -, X_n)$

$$S_n(x_1, -, x_n, \theta) = \frac{\partial}{\partial\theta} \ln(L(x_1, -, x_n, \theta))$$

Si on est dans le cas du modèle d'échantillonage i.e $X_1, -, X_n$ iid

$$S_n(x_1, -, x_n, \theta) = \frac{\partial}{\partial\theta} \ln(L(x_1, -, x_n, \theta)) = \frac{\partial}{\partial\theta} \ln\left(\prod_{i=1}^{n} f(x_i, \theta)\right)$$
$$= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln(f(x_i, \theta))$$
$$= \sum_{i=1}^{n} S(x_i, \theta)$$

Propriété : Si (H4) est vérifiée alors on a :

$$\mathbb{E}[S(X,\theta)] = 0 \quad \underline{\text{ou}} \quad \mathbb{E}[S_n(X_1, -, X_n, \theta)] = 0$$

Preuve : On part du fait que f est une densité donc $\int f(x,\theta)\, dx = 1$

$d\mu(x)$

On dérive p/r à θ et comme on a (H4)

$$\frac{\partial}{\partial\theta} 1 = 0 = \frac{\partial}{\partial\theta}\int f(x,\theta)\,dx = \int \frac{\partial}{\partial\theta} f(x,\theta)\,dx$$

$$= \int \frac{\partial}{\partial\theta} \ln(f(x,\theta)) \cdot f(x,\theta)\,dx.$$

$(\ln(u))' = \frac{u'}{u}$

$u' = (\ln(u))' \times u$

$$= \mathbb{E}\left[\frac{\partial}{\partial\theta} \ln(f(X,\theta))\right] = \mathbb{E}[S(X,\theta)].$$

Définit°: On appelle Informat° de Fisher la fonct° I définie par:

$$I : \Theta \longrightarrow \mathbb{R}^+$$
$$\theta \longmapsto I(\theta) = \mathbb{E}[(S(X,\theta))^2]$$

Informat° contenue dans un X_i

Remarques:

* L'informat° est définie si (H1), (H2) et (H3) sont vérifiées
* L'informat° dépend uniquem^t de θ et du modèle choisi. Elle représente l'informat° contenue dans X_i pour θ.
* On peut définir l'informat° pour le modèle entier i.e $X_1, \dots, X_n$ et on a

$$I_n(\theta) = \mathbb{E}[(S_n(X_1,\dots,X_n,\theta))^2].$$

Propriété: Si (H4) est vérifiée, on a:

$$I(\theta) = V(S(X,\theta)) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(f(X,\theta))\right].$$

ou

$$I_n(\theta) = V(S_n(X_1,\dots,X_n,\theta)) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(L(X_1,\dots,X_n,\theta))\right]$$

Preuve:

On reprend la preuve précédente:

$$\int \frac{\partial}{\partial\theta}\ln(f(x,\theta)) \cdot f(x,\theta)\,dx = 0$$

On dérive encore p/r à θ.

$$\int \frac{\partial^2}{\partial\theta^2}\ln(f(x,\theta)) \cdot f(x,\theta)\,dx + \int \frac{\partial}{\partial\theta}\ln(f(x,\theta)) \cdot \overbrace{\frac{\partial}{\partial\theta} f(x,\theta)}^{\left(\frac{\partial}{\partial\theta}\ln(f(x,\theta))\right)\cdot f(x,\theta)}\,dx = 0$$

$$\Rightarrow \int \frac{\partial^2}{\partial\theta^2}\ln(f(x,\theta)) f(x,\theta)\,dx + \int \left(\frac{\partial}{\partial\theta}\ln(f(x,\theta))\right)^2 f(x,\theta)\,dx = 0$$

$$\Rightarrow \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(f(X,\theta))\right] + \mathbb{E}[(S(X,\theta))^2] = 0$$

$$\Rightarrow \underbrace{\mathbb{E}[(S(X,\theta))^2]}_{=I(\theta)} = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(f(X,\theta))]$$

$$\mathbb{E}[(S(X,\theta))^2] - \underbrace{\mathbb{E}[S(X,\theta)]^2}_{=0 \text{ par prop sur le score}}$$
$$= V(S(X,\theta))$$

Propriété : Si on est dans le modèle d'échantillonage i.e $X_1, -, X_n$ iid et on a (H4) vraie, on a :

$$I_n(\theta) = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(\underbrace{L(X_1,-,X_n,\theta)}_{\prod_{i=1}^n f(X_i,\theta)}))] = -\sum_{i=1}^n \underbrace{\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(f(X,\theta))]}_{=I(\theta)}$$

$$= n\underbrace{I(\theta)}_{}$$

Informat° contenue dans un X_i. Chaque X_i (indep des autres) amène $I(\theta)$ comme info.

Exemple :

On a $X_1,-,X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$

Le modèle est :
$$(\mathbb{R}_+^*, \mathcal{B}(\mathbb{R}_+^*), \mathcal{E}(\theta), \theta\in\mathbb{R}_+^*)^n$$

On veut l'informat° du modèle $\Leftarrow$ Informat° dans $X_1,-,X_n$ i.e $I_n(\theta)$

On a $L(x_1,-,x_n,\theta) = \prod_{i=1}^n f(x_i,\theta) = \prod_{i=1}^n \theta e^{-\theta x} = \theta^n \exp(-\theta\sum_{i=1}^n x_i)$

$\ln(L(x_1,-,x_n,\theta)) = n\ln(\theta) - \theta\sum_{i=1}^n x_i$

$S_n(x_1,-,x_n,\theta) = \frac{\partial}{\partial\theta}\ln(L(x_1,-,x_n,\theta)) = \frac{n}{\theta} - \sum_{i=1}^n x_i$

On peut calculer l'info par 3 formules (dans ce cas)

$$I_n(\theta) = \underbrace{\mathbb{E}[(S_n(X_1,-,X_n,\theta))^2]}_{\text{d'habitude galère}} = \underbrace{V(S_n(X_1,-,X_n,\theta))}_{\text{un peu moins difficile}} = -\underbrace{\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(L(X_1,-,X_n,\theta))]}_{\text{normalement le + facile.}}$$

* $I_n(\theta) = V(S_n(X_1,-,X_n,\theta))$

Rappels :
* Si $X_1,-,X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$ alors $\sum X_i \sim \Gamma(n,\theta)$
* $V(\Gamma(n,\theta)) = \frac{n}{\theta^2}$

Ici, on a : $I_n(\theta) = V(\frac{n}{\theta} - \sum_{i=1}^n X_i)$
$$= V(\sum_{i=1}^n X_i)$$
$$= \frac{n}{\theta^2} \quad \text{et donc } I(\theta) = \frac{1}{\theta^2}$$

* $I_n(\theta) = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(L(X_1,-,X_n,\theta))]$

$$\frac{\partial^2}{\partial \theta^2} \ln(L(X_1, \dots, X_n, \theta)) = -\frac{n}{\theta^2}$$

$$\Rightarrow I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \ln(L(X_1, \dots, X_n, \theta))\right] = \frac{n}{\theta^2}$$

Informat° contenue dans $X_1, \dots, X_n$ sur le paramètre θ.

On peut comparer l'informat° à l'intérieur d'un m̄ modèle.

Si $\begin{cases} \theta \text{ est petit} \longrightarrow \text{bcp d'informat°} \longrightarrow \text{"facile" à estimer} \\ \theta \text{ est grand} \longrightarrow \text{pas trop d'informat°} \longrightarrow + \text{difficile à estimer} \end{cases}$

Si $\theta = 1$ $\quad I_n(\theta) = n$

Si $\theta = 10$ $\quad I_n(\theta) = \frac{n}{100}$ $\quad$ (100x moins d'info).

__Exemple__: $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$

$$L(x_1, \dots, x_n, \mu) = \prod_{i=1}^{n} f(x_i, \mu) = \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

$$= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

$$\ln(L(x_1, \dots, x_n, \mu)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$S_n(x_1, \dots, x_n, \mu) = \frac{\partial}{\partial\mu} \ln(L(x_1, \dots, x_n, \mu)) = \sum_{i=1}^{n} x_i - n\mu$$

$$\frac{\partial^2}{\partial\mu^2} \ln(L(x_1, \dots, x_n, \mu)) = -n$$

$$I_n(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial\mu^2} \ln(L(X_1, \dots, X_n, \mu))\right] = n$$

L'informat° ne dépend pas de μ.
On a la m̄ qtt d'info quelque soit μ, on aura la m̄ "précision d'estimation" quelque soit μ.

III - Inégalité de Cramer-Rao.

__Hypothèses supplémentaires:__

(HS) $\quad 0 < I(\theta) < \infty \quad$ et $\quad 0 < I_n(\theta) < \infty$

il faut qu'on ait de l'info

mais pas à l'∞ quand m̄

On prend un estimateur T_n de θ tq $\mathbb{E}[T_n] = g(\theta)$

↳ Traduit le biais de T_n

Si $g = Id$, T_n est sans biais pour θ.

Théorème : Inégalité de (Fréchet-Darmois)-Cramer Rao

Si on a (H1), (H2), (H3), (H4) et (H5) vérifiées et si g dérivable alors on a

$$V(T_n) \geq \underbrace{\frac{(g'(\theta))^2}{I_n(\theta)}}_{\text{Borne inférieure de l'inégalité de Cramer-Rao}} = K_{T_n}(\theta)$$

Preuve :

On part de

$$Cov\left(T_n, \frac{\partial}{\partial\theta}\ln(L)\right) = \mathbb{E}\left[T_n \frac{\partial}{\partial\theta}\ln(L)\right] - \mathbb{E}[T_n]\,\underbrace{\mathbb{E}\left[\overbrace{\frac{\partial}{\partial\theta}\ln L}^{\text{Score}}\right]}_{=0}$$

$$= \int T_n \underbrace{\frac{\partial}{\partial\theta}\ln L \cdot L}_{\frac{\partial}{\partial\theta}L}\, d\mu^{\otimes n}$$

$$= \int \frac{\partial}{\partial\theta}(T_n \cdot L)\, d\mu^{\otimes n} = \frac{\partial}{\partial\theta}\underbrace{\int T_n \cdot L\, d\mu^{\otimes n}}_{=\mathbb{E}[T_n]=g(\theta)}$$

$$= g'(\theta).$$

On applique l'inégalité de Cauchy-Schwartz.

$$\left(Cov\left(T_n, \frac{\partial}{\partial\theta}\ln L\right)\right)^2 \leq V(T_n) \cdot \underbrace{V\left(\frac{\partial}{\partial\theta}\ln L\right)}_{I_n(\theta)}$$

$$\Rightarrow V(T_n) \geq \frac{(g'(\theta))^2}{I_n(\theta)}.$$

Remarques :

* $V(T_n) \geq \frac{(g'(\theta))^2}{I_n(\theta)}$

dépend de l'estimateur T_n

dépend uniquement du modèle

* Si T_n est sans biais alors $V(T_n) \geq I_n(\theta)^{-1}$

Définition : On dit que T_n est un estimateur efficace si :

$$V(T_n) = K_{T_n}(\theta) \quad \text{(Borne de Cramer-Rao)}$$

Si T_n n'est pas efficace mais $\dfrac{V(T_n)}{K_{T_n}(\theta)} \longrightarrow 1$, on dit que T_n est asympt. efficace.

Remarque : Si on a un estimateur T_n efficace pour θ alors parmi tous les estimateurs qui ont le m̂ biais, c'est celui qui minimise le risque quadratique

$$(R(T_n, \theta) = B_\theta(T_n)^2 + V(T_n))$$

Rappel : On a des données $x_1, \dots, x_n$. Si on considère les données comme telles, tout ce qu'on peut faire c'est les décrire $\Rightarrow$ Stat. descriptive

En stat. inf., on voit les données comme des réalisat° de v.a., on veut avoir des informat° ou à approcher la ou les v.a. qui ont généré les observat°.

Exemple : Sous R, on fait :

```
> X = rpois(100, 30)
> Y = rpois(100, 30)
> mean(X)
      [1] 29.573
> mean(Y)
      [1] 31.264
```

$(X \sim \mathcal{P}(30))$

Simulat° $\longrightarrow$ r
densité $\longrightarrow$ d
fdr $\longrightarrow$ p
quantile $\longrightarrow$ q

pois
exp
norm
⋮

En stat. descriptive les 2 séries sont différentes.

En stat. inf. elles sont "égales" i.e. les v.a. qui ont générées les données ont m̂ distribut°.

La différence est due à la variabilité naturelle de l'expérience aléatoire.

(Je réécris la def pour + de cohérence)

Définition : On dit que T_n est un estimateur efficace si :

$$V(T_n) = K_{T_n}(\theta) \quad \text{(Borne de Cramer-Rao)}$$

Si T_n n'est pas efficace mais $\dfrac{V(T_n)}{K_{T_n}(\theta)} \longrightarrow 1$, on dit que T_n est asympt. efficace.

Exemple : $X_1, \ldots, X_n$ iid telles que $X_i \sim \mathcal{P}(\theta)$

On prend $T_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ (est un estimateur sans biais et convergent pour θ car $\mathbb{E}[X] = \theta$)

On a $V(T_n) = V\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{\theta}{n}$ et $\mathbb{E}[T_n] = \theta$.

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$
$$= \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\theta}$$

$$\ln(L(x_1, \ldots, x_n, \theta)) = \left(\sum_{i=1}^{n} x_i\right) \ln(\theta) - n\theta - \ln\left(\prod_{i=1}^{n} x_i\right)$$

$$I_n(\theta) = V(S_n(X_1, \ldots, X_n, \theta)) = V\left(\frac{\sum_{i=1}^{n} X_i}{\theta} - n\right) = \frac{\sum_{i=1}^{n} V(X_i)}{\theta^2} = \frac{n}{\theta}$$

Borne de Cramer-Rao : Ici $\mathbb{E}[T_n] = \theta$ donc $h(\theta) = \theta$

$$K_{T_n} = \frac{(h'(\theta))^2}{I_n(\theta)} = \frac{1}{n/\theta} = \frac{\theta}{n} = V(T_n) \Rightarrow T_n \text{ est efficace pour } \theta$$

Relation entre estimateurs efficaces

Propriétés :

① Si T_n est un estimateur efficace de θ, alors $kT_n + b$ est aussi un estimateur efficace de θ, $\forall k \in \mathbb{R}^*, \forall b \in \mathbb{R}$.

② Soient T_{1n} et T_{2n} deux estimateurs sans biais du paramètre θ. S'ils sont tous les deux efficaces, alors $T_{1n} = T_{2n}$ presque sûrement.

Traduct° : $R(T_n, \theta) = B_\theta(T_n)^2 + V(T_n)$

Si on considère les estimateurs sans biais $R(T_n, \theta) = V(T_n)$, comme tous les estimateurs sans biais ont la même borne de Cramer-Rao et que T_n atteint cette borne alors T_n est le meilleur estimateur au sens du risque quadratique parmi les estimateurs sans biais de paramètre θ.

T_n est BLUE (Best Linear Unbiased Estimator) pour θ.

Preuve :

① Si T_n est tq $\mathbb{E}[T_n] = h(\theta)$ et $V(T_n) = \frac{(h'(\theta))^2}{I_n(\theta)}$

Si $T_n^* = aT_n + b$, on a $\mathbb{E}[T_n^*] = a\,\mathbb{E}[T_n] + b$
$$= a\,h(\theta) + b.$$

Donc $K_{T_m^*}(\theta) = \dfrac{(a\, h'(\theta))^2}{I_m(\theta)} = a^2 \dfrac{(h'(\theta))^2}{I_m(\theta)} = a^2 K_{T_m}(\theta) = a^2 V(T_m)$

et $V(T_m^*) = V(aT_m + b) = a^2 V(T_m)$

=

$\Rightarrow T_m^*$ est efficace

② Soit $T_{m,1}$ et $T_{m,2}$ deux estimateurs sans biais et efficace de θ.

$T_{m,3} = \dfrac{T_{m,1} + T_{m,2}}{2}$ est un estimateur sans biais de θ.

Donc $V(T_{m,3}) = \frac{1}{4}\left[V(T_{m,1}) + V(T_{m,2}) + 2\rho\sigma_1\sigma_2\right]$ où ρ est le coeff de corrélat° linéaire entre $T_{m,1}$ et $T_{m,2}$

On pose $V(T_{m,1}) = V(T_{m,2}) = V$ (car ils sont tous les deux sans biais et efficace).

$V(T_{m,3}) = \frac{1}{4}\left[2V + 2\rho V\right] = \frac{V}{2}(1+\rho)$

On ne peut avoir $\rho < 1$ car sinon $V(T_{m,3}) < V$
On a alors $\rho = 1$

Si $\rho = 1$, on a $T_{m,1} - \mathbb{E}[T_{m,1}] = \alpha(T_{m,2} - \mathbb{E}[T_{m,2}])$
mais comme $V(T_{m,1}) = V(T_{m,2})$ on a $\alpha = 1$ et donc $T_{m,1} = T_{m,2}$ p.s.

Remarque : Le ① dit que si on a un estimateur efficace, c'est sympa de chercher une transformat° linéaire qui minimise ou annule le biais (on garde l'efficacité).

Dégradation de l'information

On a $X_1, \ldots, X_m$ iid tq $X_i \sim \mathcal{P}_\theta$. $I_m(\theta)$ est toute la qtt d'info qu'on a sur θ (je peux pas en avoir +).
On prend $X_1, \ldots, X_m$ et on résume tout ça dans une stat. T_m.

Qu'est-ce qu'il s'est passé au niveau de l'informat°?

Si $h(t,\theta)$ est la vraissemblance de T, on note $I_T(\theta)$ l'informat° relative à T :

$$I_T(\theta) = \mathbb{E}\left[S(T,\theta)^2\right] \text{ où } S(T,\theta) = \frac{\partial \ln(h(t,\theta))}{\partial\theta}$$

$I_T(\theta)$ vérifie les m̂ prop (centrée, relat° avec $\frac{\partial^2}{\partial\theta^2} h$, inégalité de Cramer-Rao) que $I(\theta)$.

Proposit° :
Soit T une stat., on a $I_T(\theta) \leq I_m(\theta)$ avec égalité <u>ssi</u> la stat T est exhaustive pour le paramètre θ.

Exemple : On prend $X_1, \ldots, X_m$ iid tq $X_i \sim \mathcal{P}(\theta)$
On prend $T_m = \underbrace{\sum_{i=1}^{m} X_i}_{\Sigma \text{ de } \mathcal{P}(\theta) \text{ indep}} \sim \mathcal{P}_m(\theta)$.

Ici $h(t,\theta) = \mathbb{P}(T_n = t) = \frac{(n\theta)^t}{t!} e^{-n\theta}$

$\ln(h(t,\theta)) = t\ln(n\theta) - \ln(t!) - n\theta$

$\frac{\partial}{\partial\theta} \ln(h(t,\theta)) = \frac{t}{\theta} - n = S_1(t,\theta)$ $\qquad \frac{\partial^2}{\partial\theta^2} \ln(h(t,\theta)) = -\frac{t}{\theta^2}$

$$I_{T_n}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \ln(h(T,\theta))\right] = -\mathbb{E}\left(-\frac{\overset{\mathcal{P}(n\theta)}{T}}{\theta^2}\right) = \frac{\mathbb{E}[T]}{\theta^2} = \frac{n}{\theta} = I_n(\theta)$$

On a autant d'info. pour θ dans $X_1, \dots, X_n$ que dans $\sum X_i$ si on a des v.a.

Conséquence : On a un portefeuille avec n contrats d'assurance indep.

On s'intéresse au nbr de sinistres sur chaque contrat, on peut considérer que ce sont des réalisat° de v.a. de Poisson iid

On observe $0, 1, 0, 0, 0, 2, 1, 0, \dots, 2, 1$ et on a $\sum x_i = 148$

Vocabulaire :

Si on a $I_{T_n}(\theta) = I_n(\theta)$ on dit que T_n est exhaustive pour θ (en anglais sufficient statistic).

IV – Exhaustivité

Dans un problème stat avec un paramètre θ, l'échantillon $X_1, \dots, X_n$ apporte une certaine qtt d'info ($I_n(\theta)$)
On résume l'échantillon pour une statistique T_n et on veut essayer de ne pas perdre d'infos.

Une stat T_n qui conserve l'info est une stat exhaustive.
On peut traduire ça par le fait que la loi conditionnelle de $X_1, \dots, X_n$ sachant que T_n est indep de θ.

⚠ Spoiler : 99,9 % de stats que l'on va étudier sont exhaustives.
Formellement, on se place dans le modèle $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta \subseteq \mathbb{R})^n$

Définition : (Principe de factorisat°)

La stat T_n est ici une v.a. définie sur $(\mathcal{X}, \mathcal{B}_\mathcal{X}, \mathbb{P}_\theta, \theta \in \Theta \subset \mathbb{R})^n$ à valeurs dans $(\Theta, \mathcal{B}_\Theta)$

On considère :

- la vraisemblance (fonct° de densité ou de proba) de T_n que l'on va noter $h(t,\theta)$
- la densité conjointe de l'échantillon $(X_1, \dots, X_n)$ notée $L(x_1, \dots, x_n, \theta)$

On dira que la stat. T_n est exhaustive pour θ s'il existe une fonct° k sur $\mathcal{X}^n$ tq :

$$\underbrace{L(x_1, \dots, x_n, \theta)}_{\text{vraisemblance ou densité de } (X_1, \dots, X_n, \theta)} = \underbrace{h(t,\theta)}_{\text{'' de } T_n} \underbrace{k(x_1, \dots, x_n)}_{\text{'' conditionnelle de } X_1, \dots, X_n \text{ sachant que } T_n \text{ indep de } \theta}$$

Si T_n exhaustive pour θ.

On a :

$$L(x_1, \dots, x_n, \theta) = h(t, \theta)\, k(x_1, \dots, x_n)$$

$$\ln(L(x_1, \dots, x_n, \theta)) = \ln(h(t,\theta)) + \ln(k(x_1, \dots, x_n))$$

$$\frac{\partial}{\partial \theta} \ln(L(x_1, \dots, x_n, \theta)) = \frac{\partial}{\partial \theta} \ln(h(t,\theta))$$

$$\vdots$$

$$I_n(\theta) = I(\theta)$$

On peut aussi dire que T_n est exhaustive pour θ si :

$$\frac{L(x_1, \dots, x_n, \theta)}{h(t,\theta)} \text{ est indep de } \theta.$$

Théorème (de factorisat°) :

Pour qu'une stat T soit exhaustive, il suffit que la vraisemblance s'écrive :

$$L(x_1, \dots, x_n, \theta) = \underbrace{\phi(t,\theta)}_{\text{dépend indépendamment de } t \text{ et } \theta} \underbrace{\psi(x_1, \dots, x_n)}_{\text{dépend uniquem}^t \text{ de } x_1, \dots, x_n}$$

Remarque :

- C'est une bonne méthode pour mq T_n est exhaustive mais pas pour montrer qu'elle n'est pas exhaustive
- Une transformat° bijective d'une stat. exhaustive est exhaustive

Exemple :

$X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{P}(\theta)$ et $T_n = \sum_{i=1}^{n} X_i$

On a $$L(x_1, \dots, x_n, \theta) = \frac{\theta^{\sum x_i}}{\prod x_i!} e^{-n\theta} = \underbrace{\theta^{\sum x_i} e^{-n\theta}}_{\varphi(t,\theta)} \times \underbrace{\left(\prod^{n} x_i!\right)^{-1}}_{\varphi(x_1, \dots, x_n)}$$

$$t = \sum x_i$$

V - Exhaustivité et estimateurs efficaces : la famille exponentielle

Définition : On appelle famille exponentielle à paramètre unidimensionnel θ toute loi de proba (discrète ou continue) dont la vraisemblance peut se mettre sous la forme :

$$f(x,\theta) = \begin{cases} \exp[\alpha(\theta)\beta(x) + \gamma(\theta) + \delta(x)] & \text{si } x \in \mathcal{X} \\ 0 & \text{si } x \notin \mathcal{X} \end{cases}$$

avec α et γ des fonct° de 2 fois $\neq$

Si $\alpha(\theta)=\theta$, c'est la forme naturelle de la famille exponentielle

Si $\beta(\theta)=x$, c'est la forme canonique de la famille exponentielle

Rq: Les fonct° α, β, γ et δ ne sont pas uniques.

Proposition:

Dans la famille exponentielle, toute stat sous la forme $T_n = k\sum_{i=1}^{n}\beta(X_i)$ est exhaustive pour θ.

Preuve:

Si f est dans la famille exponentielle, on a:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \exp[\alpha(\theta)\beta(x_i) + \gamma(\theta) + \delta(x_i)]$$

important

$$= \exp[\alpha(\theta)\sum\beta(x_i) + n\gamma(\theta) + \sum\delta(x_i)]$$

$$= \underbrace{\exp[\alpha(\theta)\sum\beta(x_i) + n\gamma(\theta)]}_{\varphi(t=\sum\beta(x_i),\theta)} \cdot \underbrace{\exp[\sum\delta(x_i)]}_{\psi(x_1, \dots, x_n)}$$

Exemple:

Si $X \sim \mathcal{P}(\theta)$, on a $f(x) = \dfrac{\theta^x}{x!}e^{-\theta}$

On a $f(x) = \exp(\underbrace{x}_{\beta(x)}\underbrace{\ln(\theta)}_{\alpha(\theta)} \underbrace{-\theta}_{\gamma(\theta)} \underbrace{-\ln(x!)}_{\delta(x)})$ si $x \in \mathbb{N}$

$$\Rightarrow T_n = \frac{1}{n}\sum_{i=1}^{n}\beta(x_i) = \frac{1}{n}\sum_{i=1}^{n}X_i \text{ est une stat exhaustive.}$$

Modèle exponentiel:

P_θ appartient au modèle exponentiel si sa densité s'écrit

$$f(x,\theta) = \exp(\alpha(\theta)\beta(x) + \gamma(\theta) + \delta(x))$$

$$\Rightarrow a\sum_{i=1}^{n}\beta(x_i) \text{ est une stat exhaustive}$$

Proposit°: Si $\mathcal{X}$ est indep de θ alors le modèle admet une stat exhaustive ssi il est exponentiel

Théorème : Si $\mathcal{X}$ est indep de θ et que le modèle admet un estimateur efficace alors le modèle est exponentiel car l'estimateur est exhaustif.

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

On veut estimer μ. On sait que $\overline{X}_n$ est sans biais et convergent pour μ

On a :

$$L(x_1, \dots, x_n, \mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right)$$

$$\ln(L(x_1, \dots, x_n, \mu)) = \frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \ln(L(x_1, \dots, x_n, \mu)) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{\sum_{i=1}^{n} x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\frac{\partial^2}{\partial \mu^2} \ln(L(x_1, \dots, x_n, \mu)) = -\frac{n}{\sigma}$$

$$I_n(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial \mu^2} \ln L(X_1, \dots, X_n, \mu)\right] = -\mathbb{E}\left[-\frac{n}{\sigma^2}\right] = \frac{n}{\sigma^2}$$

On a $\mathbb{E}[\overline{X}_n] = \mu$ et $V(\overline{X}_n) = \frac{\sum V(X_i)}{n^2} = \frac{\sigma^2}{n}$ $=$

$\overline{X}_n$ est efficace pour θ si $V(\overline{X}_n) = \frac{1}{I_n(\mu)} = \frac{\sigma^2}{n}$

Ici : $\mathcal{X} = \mathbb{R}$ indép de θ.

$\Longrightarrow$ Le modèle est exponentiel

Vérificat°

$$f(x, \mu) = (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left(\frac{-1}{2\sigma^2} \overbrace{(x - \mu)^2}^{x^2 - 2x\mu + \mu^2}\right)$$

$$= \exp\Big(\underbrace{x}_{\beta(x)} \underbrace{\frac{\mu}{\sigma^2}}_{\alpha(\mu)} - \underbrace{\frac{\mu^2}{2\sigma^2}}_{\delta(\mu)} \underbrace{- \frac{x^2}{2\sigma^2} - \ln(\sigma) - \frac{1}{2}\ln(2\pi)}_{\gamma(x)}\Big)$$

$\Longrightarrow$ Le modèle est exponentiel.

$\Longrightarrow$ $\overline{X}_n$ est exhaustive pour μ

Vérificat° :

On a $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

$$L_{\overline{X}}(x, \mu) = \left(\frac{1}{2\pi}\right)\left(\frac{\sigma}{\sqrt{n}}\right)^{-1} \exp\left(-\frac{n}{2\sigma^2}(x - \mu)^2\right)$$

$$\ln(L_{\bar X}(x,\mu)) = -\ln(2\pi) - \ln\left(\frac{\sigma}{\sqrt{n}}\right) - \frac{n}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial\mu}\ln(L_{\bar X}(x,\mu)) = \frac{n}{\sigma^2}(x-\mu) = \frac{nx}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\frac{\partial^2}{\partial\mu^2}\ln(L_{\bar X}(x,\mu)) = -\frac{n}{\sigma^2} \Rightarrow I_{\bar X}(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial\mu^2}\ln(L_{\bar X}(x,\mu))\right] = \frac{n}{\sigma^2} = I_n(\mu)$$

$\Rightarrow \bar X_n$ exhaustive.

V - Quelques méthodes d'estimation.

Méthode empirique ou bricolage

Si le paramètre θ représente une quantité particulière pour le modèle (espérance, variance, borne sup...), on peut choisir "naturellement" la quantité empirique correspondante (issue de la stat descriptive) calculée sur l'échantillon $X_1, \ldots, X_n$

Par exemple, $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{P}(\theta)$

On a $\mathbb{E}[X] = \theta$ et $V(X) = \theta$

On peut estimer θ par $\frac{1}{n}\sum_{i=1}^{n} X_i$ ou par $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar X_n)^2$

Avantage: Facile à utiliser

Inconvénient: Limité

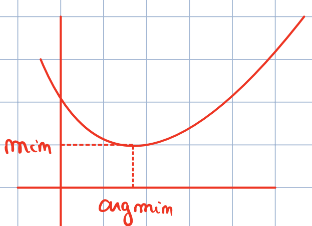
Moindres carrés:

Définition: On appelle estimateur des moindres carrés de θ la statistique

$$\hat\theta_{MC} = \underset{\theta\in\Theta}{\arg\min} \underbrace{\sum_{i=1}^{n}(X_i - h_i(\theta))^2} \quad \text{où } h_i(\theta) = \mathbb{E}[X_i]$$

le θ qui minimise

Idée: On cherche θ qui minimise la distance entre les observat° et leur valeur théorique.



Exemple: En économétrie, on prend

$$X_i = a u_i + b + \varepsilon_i$$

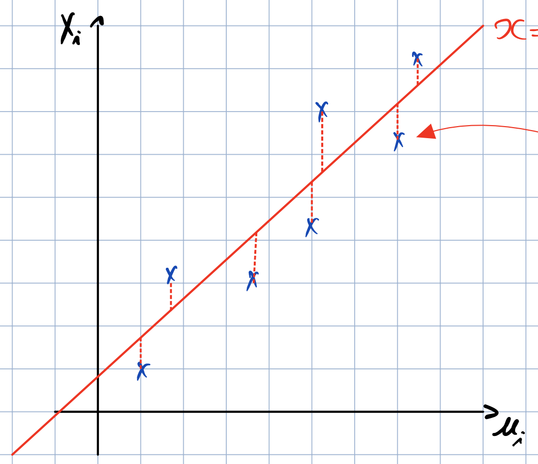
On a $\mathbb{E}[X_i] = a u_i + b$.

Les v.a. observées

connue

paramètres du modèle.

v.a centrée et tq $V(\varepsilon_i) = \sigma^2$ et $\varepsilon_1, \ldots, \varepsilon_n$ indep.



$\sum_{i=1}^{m} (X_i - au_i - b)^2$

On cherche a et b qui minimise cette distance

Ici ssi $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

$X_i \sim \mathcal{N}(au_i + b, \sigma^2)$ donc les X_i sont indep mais n'ont pas la même loi.

Théorème : Si on est dans le cas $X_1, \dots, X_m$ iid tq $\mathbb{E}[X_i] = h(\theta)$ avec h bijective

$$\hat{\theta}_{MC} = \underset{\theta \in \Theta}{\arg\min} \sum_{i=1}^{m} (X_i - h(\theta))^2$$

alors $\hat{\theta}_m = h^{-1}(\bar{X}_m)$

Exemple : $X_1, \dots, X_m$ iid tq $X_i \sim \mathcal{E}(\theta)$. On a $\mathbb{E}[X_i] = \frac{1}{\theta}$.

$$\hat{\theta}_{MC} = \underset{m}{\arg\min} \sum_{i=1}^{m} (X_i - \frac{1}{\theta})^2$$

$$\Rightarrow \hat{\theta}_{MC} = (\bar{X}_m)^{-1}$$

$$\begin{cases} h(\theta) = \sum (X_i - \frac{1}{\theta})^2 \\ h'(\theta) = \frac{\theta}{\theta} 2 \sum (X_i - \frac{1}{\theta}) \end{cases}$$

Truc

$\Rightarrow \frac{1}{\theta} = \frac{\sum X_i}{m}$

$\Rightarrow \hat{\theta} = \frac{m}{\sum X_i}$

Avantages :

- adaptée à des paramètres multidimensionnels
- adaptée au cas où les v.a sont indep mais pas identiquement distribuées (cf modèle linéaire).

Inconvénient :

- Truc d'économétrie
- Il faut que l'espérance s'exprime comme une fct° de (tous les) paramètres.

Méthode des moments

Rappel : Soit X une v.a.r

Le moment (théorique) d'ordre r de X est $M_r(\theta) = \mathbb{E}[X^r]$

Le moment centré (théorique) d'ordre r de X est $\bar{M}_r(\theta) = \mathbb{E}[(X - \mathbb{E}[X])^r]$

Si on a un échantillon d'observation $(x_1, \dots, x_n)$, le moment empirique d'ordre r est $m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$

Le moment empirique centré d'ordre r est $\bar{m}_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$

Principe : On suppose que le paramètre est de dimension p. On cherche θ qui égale les p premiers moments empiriques et théoriques. Autrement dit, on pose le système d'équat° :

$$\begin{cases} M_1(\theta) = m_1 \\ \quad \vdots \\ M_p(\theta) = m_p \end{cases} \quad \text{ou} \quad \begin{cases} M_1(\theta) = m_1 \\ \quad \vdots \\ \bar{M}_p(\theta) = \bar{m}_p \end{cases}$$

dépendent de θ

dépend pas de θ

$\Rightarrow$ Nb d'équat° = dimens° du paramètre.

Exemple : On prend $X_1, \dots, X_n$ iid tq $X_i \sim \text{Gamma}(a,b)$ de densité $\frac{x^{a-1}\theta^a}{\Gamma(a)} e^{-\theta x}$.

On a $\mathbb{E}[X_i] = \frac{a}{\theta}$ et $V(X) = \frac{a}{\theta^2}$

L'estimateur des moments $(\hat{a}_n, \hat{\theta}_n)$ est tq

$$\begin{cases} \bar{X}_n = \dfrac{\hat{a}_n}{\hat{\theta}_n} \\ \overline{X^2_n} - \bar{X}_n^2 = \dfrac{\hat{a}_n}{\hat{\theta}_n^2} \end{cases} \Rightarrow \quad \hat{a} = \frac{(\bar{X}_n)^2}{\overline{X^2_n} - \bar{X}^2}, \quad \hat{\theta}_n = \frac{\bar{X}_n}{\overline{X^2_n} - \bar{X}_n^2}$$

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

Avantages

- adaptée aux paramètres multidimensionnels
- Relativement simple à mettre en place
- On peut trouver des propriétés asymptotiques pour les estimateurs
- On n'a pas besoin de connaître le modèle, juste les moments.
- Numériquement léger.

Inconvénients:

- réservé à cas iid
- On a besoin des moments jusqu'à l'ordre p.

Méthode du maximum de vraisemblance

Définition: On appelle estimateur du maximum de vraisemblance (maximum likelihood estimator) de θ la statistique $\hat{\theta}_{MV}$ rendant maximale la vraisemblance i.e.

$$\hat{\theta}_{MV} = \underset{\theta \in \Theta}{\arg\max}\, L(X_1, \ldots, X_n, \theta)$$

Remarque:

- L'idée de maximiser la vraisemblance est assez naturelle :
  c'est la valeur qui maximise la probabilité d'obtenir les réalisations.
- C'est valable quelque soit le modèle mais si on a les bonnes hypothèses, c'est + facile.

Propriété: Si le modèle vérifie les hypothèses (H1) à (H4)

alors pour que $\hat{\theta}_{MV}$ soit un EMV (estimateur du maximum de vraisemblance) de θ, il est nécessaire que:

$$\begin{cases} \dfrac{\partial}{\partial \theta} \ln L(X_1, \ldots, X_n, \theta)\Big|_{\theta = \hat{\theta}_{MV}} = 0 \quad \text{i.e. } S_n(X_1, \ldots, X_n, \hat{\theta}_n) = 0 \\ \dfrac{\partial^2}{\partial \theta^2} \ln L(X_1, \ldots, X_n, \theta)\Big|_{\theta = \hat{\theta}_{MV}} < 0 \end{cases}$$

Exemple: $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$

On a bien toutes les hypothèses.

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

$$\ln L(x_1, \ldots, x_n, \theta) = n \ln(\theta) - \theta \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, \ldots, x_n, \theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

$\hat{\theta}_{MV}$ est tq $\dfrac{n}{\hat{\theta}_{MV}} - \sum_{i=1}^{n} x_i = 0$

$$\Rightarrow \hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$\frac{\partial^2}{\partial \theta^2} \ln L(x_1, \ldots, x_n, \theta) = -\frac{n}{\theta^2}$$

$$\Rightarrow \hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^{n} X_i} \text{ est l'EMV de } \theta.$$

Exemple:

On prend $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{U}[0;\theta]$

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0;\theta]}(x_i).$$

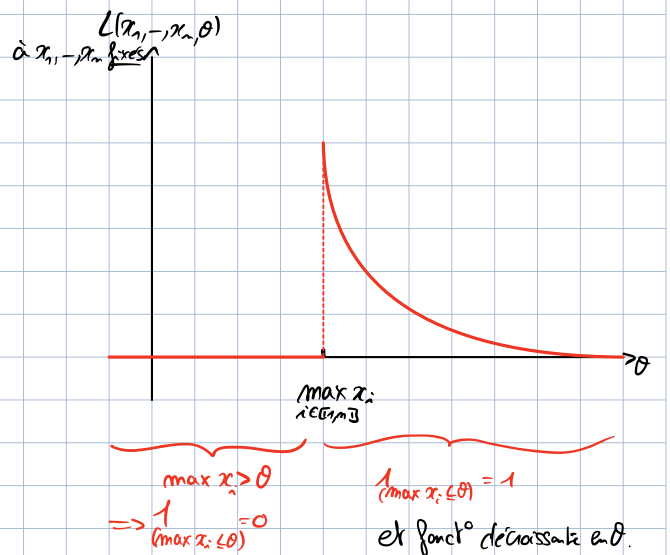
$$= \frac{1}{\theta^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}_{[0;\theta]}(x_i)}_{= \begin{cases} 1 & \text{si tous les } x_i \text{ sont } [0;\theta] \\ 0 & \text{s'il y en a qui se branle} \end{cases}} \rightarrow = \mathbb{1}_{(\min x_i \geq 0)} \times \mathbb{1}_{(\max x_i \leq \theta)}$$

On n'a pas (H1) donc on ne peut pas passer par le ln

On maximise directement la vraisemblance

$$L(x_1, \ldots, x_n, \theta) = \frac{1}{\theta^n} \mathbb{1}_{(\min x_i \geq 0)} \times \mathbb{1}_{(\max x_i \leq \theta)}$$

$$\Rightarrow \hat{\theta}_{MV} = \max_{i \in [\![1,n]\!]} X_i$$



Avantages:

- C'est celui qu'on utilise le + en pratique
- S'il existe une stat exhaustive T_n, $\hat{\theta}_{MV}$ dépend que de T_n
- Il a beaucoup de propriétés asymptotiques.

Inconvénients:

- Il faut très bien connaître le modèle.
- Numériquement ça peut être lourd.

Rappel : $X_1, \ldots, X_n$ iid tq $X_i \sim \mathbb{P}_\theta$

T_n estimateur de θ

Exhaustivité

1- $I_{T_n}(\theta) = I_n(\theta)$ autant d'informat° dans $X_1, \ldots, X_n$ que dans T_n

2- Si h est la densité de T_n.

$$L(x_1, \ldots, x_n, \theta) = h(t, \theta)\, k(x_1, \ldots, x_n)$$

3- $L(x_1, \ldots, x_n, \theta) = \varphi(t, \theta)\, \psi(x_1, \ldots, x_n)$

4- Si on a le modèle exponentiel i.e. $f(x) = \exp\{\alpha(\theta)\beta(x) + \delta(x) + \gamma(\theta)\}$

$$\Rightarrow \sum \beta(x_i) \text{ exhaustive}$$

5- Si $\mathcal{X}$ ne dépend pas de θ (H1) alors efficace $\Rightarrow$ exhaustif.

Méthodes d'estimation :

- **Moindres carrés** : $h_i(\theta) = \mathbb{E}[X_i]$

$$\hat{\theta}_{MC} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{n} (X_i - h_i(\theta))^2$$

- **Méthode des moments** :

Si θ est de dimens° p, $\mathbb{E}[X_i^k] = M_k(\theta)$

et $\frac{1}{n}\sum_{i=1}^{n} x_i^k = m_k$

$\hat{\theta}_{MM}$ est tq $\begin{cases} M_1(\hat{\theta}_{MM}) = m_1 \\ \vdots \\ M_p(\hat{\theta}_{MM}) = m_p \end{cases}$

- **Méthode du max de vraisemblance** : (Maximum Likelihood Estimator)

$$\hat{\theta}_{MV} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(X_1, \ldots, X_n, \theta)$$) on n'a pas besoin de (H1) etc.

Si on a (H1) à (H4) vérifiée alors $\hat{\theta}_{ML}$ tq :

$$\begin{cases} \frac{\partial}{\partial \theta} \ln L(X_1, \ldots, X_n, \theta)\big|_{\theta = \hat{\theta}_{ML}} = 0 \\ \frac{\partial^2}{\partial \theta^2} \ln L(X_1, \ldots, X_n, \theta)\big|_{\theta = \hat{\theta}_{ML}} < 0 \end{cases}$$

VII - Généralisation à un paramètre multidimensionnel

On considère le modèle stat

$$\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta \subset \mathbb{R}^{p}\right)^{n} \quad p \geqslant 2$$

Une stat T_n est une fonction de $X_1, \ldots, X_n$) pas la vraie def.
C'est un vecteur aléatoire de dimension p avec

$$T_n = (T_{n,1}, \ldots, T_{n,p})$$

et on a $\theta = (\theta_1, \ldots, \theta_p)$

Biais: $$B_\theta(T_n) = \mathbb{E}[T_n] - \theta = \begin{pmatrix} \mathbb{E}[T_{n,1}] \\ | \\ \mathbb{E}[T_{n,p}] \end{pmatrix} - \begin{pmatrix} \theta_1 \\ | \\ \theta_p \end{pmatrix} = \begin{pmatrix} \mathbb{E}[T_{n,1}] - \theta_1 \\ | \\ \mathbb{E}[T_{n,p}] - \theta_p \end{pmatrix} = \begin{pmatrix} B_{\theta_1}(T_{n,1}) \\ | \\ B_{\theta_p}(T_{n,p}) \end{pmatrix}$$

T_n est sans biais si $B_\theta(T_n) = 0_p$

T_n est asympt. sans biais si $B_\theta(T_n) \longrightarrow 0_p$ i.e $\forall j \in [\![1,p]\!], \mathbb{E}[T_{n,j}] \longrightarrow \theta_j$

Convergence: T_n est convergent pour θ si $T_n \xrightarrow{\mathbb{P}} \theta$.

C'est à dire si $\forall j \in [\![1,p]\!], T_{n,j} \xrightarrow{\mathbb{P}} \theta_j$

Risque quadratique: Rappel: en dim 1 $R(T_n,\theta) = \mathbb{E}[(T_n-\theta)^2]$

$$R(T_n,\theta) = \mathbb{E}[(T_n-\theta)'(T_n-\theta)] = \mathbb{E}\left[(T_{n,1}-\theta_1, \ldots, T_{n,p}-\theta_p)\begin{pmatrix} T_{n,1}-\theta_1 \\ | \\ T_{n,p}-\theta_p \end{pmatrix}\right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{p} (T_{n,j}-\theta_j)^2\right] = \sum_{j=1}^{p} \mathbb{E}[(T_{n,j}-\theta_j)^2]$$

$$= \sum_{j=1}^{p} R(T_{n,j}, \theta_j)$$

$$\Longrightarrow R(T_n,\theta) = \sum_{j=1}^{p} B_{\theta_j}(T_j)^2 + \sum_{j=1}^{p} V(T_{n,j})$$

$$= B_\theta(T_n)' B_\theta(T_n) + \sum_{j=1}^{p} V(T_{n,j})$$

Généralisation de Cramer-Rao:

On note $f(x,\theta)$ la vraisemblance du modèle en dimension 1 i.e si $n=1$
$= $ densité de X.

$L(x_1,\dots,x_n,\theta)$ la vraisemblance du modèle
$= $ densité du vecteur $(X_1,\dots,X_n)$

$\mathcal{X}$ est le support de f et $\mathcal{X}^n$ est le support de L.

Hypothèses de régularité

(H1') $\mathcal{X}$ est indépendant de θ

(H2') Θ est un ouvert.

On a $f(x,\theta)>0$ $\forall (x,\theta)\in\mathcal{X}\times\Theta$

(H3') $\forall j\in [\![1,p]\!]$, $\frac{\partial}{\partial\theta} f(x,\theta)$ est bien définie $\forall (x,\theta)\in\mathcal{X}\times\Theta$

$\forall (j,k)\in[\![1,p]\!]^2$, $\frac{\partial^2}{\partial\theta_j\partial\theta_k} f(x,\theta)$ est bien définie $\forall(x,\theta)\in\mathcal{X}\times\Theta$

(H4') $\forall (j,k)\in[\![1,p]\!]^2$, $\frac{\partial}{\partial\theta_j} f(x,\theta)$ et $\frac{\partial^2}{\partial\theta_j\partial\theta_k} f(x,\theta)$

Vérifient l'hypothèse (H4) de dominat° sur tout compact de Θ par des fonct° de x μ-intégrables

pour pouvoir $\frac{\partial}{\partial\theta}\int = \int\frac{\partial}{\partial\theta}$

Définition : (Score)

La fonction score définie par la fonct°

$$S : \mathcal{X}\times\Theta \longrightarrow \mathbb{R}$$
$$(x,\theta)\longmapsto \operatorname{grad}\ln f(x,\theta) = \begin{pmatrix}\frac{\partial}{\partial\theta_1}\ln f(x,\theta)\\ \vdots \\ \frac{\partial}{\partial\theta_n}\ln f(x,\theta)\end{pmatrix}$$

On peut définir $S_n(x_1,\dots,x_n,\theta) = \operatorname{grad}\ln L(x_1,\dots,x_n,\theta)$

Si on a le modèle d'échantillonage (i.e. $X_1,\dots,X_n$ iid)
$$S_n(x_1,\dots,x_n,\theta) = \sum_{i=1}^{n} S(x_i,\theta)$$

Si on a (H4') on a $\mathbb{E}[S(X,\theta)] = 0_p = \mathbb{E}[S_n(X_1,\dots,X_n,\theta)]$

Définition : (Matrice d'informat°)

La matrice d'informat° de Fisher est une matrice $p\times p$ définie par
$$I(\theta) = \mathbb{E}[S(X,\theta)\,S(X,\theta)']$$

C'est la matrice de terme général
$$\mathbb{E}\left[\frac{\partial}{\partial\theta_j}\ln f(x,\theta)\frac{\partial}{\partial\theta_k}\ln f(x,\theta)\right]$$

Sur la diagonale on a $\mathbb{E}\left[\left(\frac{\partial}{\partial\theta_j}\ln f(x,\theta)\right)^2\right] = \mathbb{E}[(S(X,\theta_j)^2] = I(\theta_j)$

$$I_n(\theta) = \mathbb{E}[S_n(X_1,\dots,X_n,\theta)\cdot S_n(X_1,\dots,X_n,\theta)'].$$

Si (H_4') vérifiée, la matrice $I(\theta)$ est la matrice de variance de $S(X,\theta)$.

$\Rightarrow$ $I(\theta)$ est une matrice symétrique et (semi) définie positive.
Le terme général de $I(\theta)$ est
$$-\mathbb{E}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ln f(X,\theta)\right]$$

Dans le modèle d'échantillonage $I_n(\theta) = n\,I(\theta)$

Exemple:

$X_1, \dots, X_n$ va iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

Ici $\theta = (\mu, \sigma^2) = (\mu, \alpha)$

La densité est:
$$f(x,\theta) = \frac{1}{\sqrt{2\pi}}\,\alpha^{-1/2}\exp\left\{-\frac{1}{2\alpha}(x-\mu)^2\right\}$$

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n}(2\pi)^{-1/2}\alpha^{-1/2}\exp\left\{-\frac{1}{2\alpha}(x_i-\mu)^2\right\}$$
$$= (2\pi)^{-\frac{n}{2}}\alpha^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\alpha}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\ln L(x_1, \dots, x_n, \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\alpha) - \frac{1}{2\alpha}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\alpha}\ln L(x_1, \dots, x_n, \theta) = -\frac{n}{2\alpha} + \frac{\alpha^{-2}}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$
$$\frac{\partial}{\partial\mu}\ln L(x_1, \dots, x_n, \theta) = \alpha^{-1}\sum_{i=1}^{n}(x_i-\mu)$$

$$\Rightarrow S_n(x_1, \dots, x_n, \theta) = \begin{pmatrix} \alpha^{-1}\sum(x_i-\mu)^2 \\ -\frac{n}{2\alpha} + \frac{1}{2}\alpha^{-2}\sum(x_i-\mu)^2 \end{pmatrix}$$

On utilise la formule $I_n(\theta) = \left(-\mathbb{E}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ln L(X_1, \dots, X_n, \theta)\right]\right)_{j,k}$

$$\frac{\partial^2}{\partial\alpha^2}\ln L(x_1, \dots, x_n, \theta) = \frac{n}{2\alpha^2} - \alpha^{-3}\sum(x_i-\mu)^2$$

$$\frac{\partial^2}{\partial\mu^2}\ln L(x_1, \dots, x_n, \theta) = -\frac{n}{\alpha}$$

$$\frac{\partial^2}{\partial\alpha\partial\mu}\ln L(x_1, \dots, x_n, \theta) = -\alpha^{-2}\sum_{i=1}^{n}(x_i-\mu)$$

$$-\mathbb{E}\left[-\alpha^{-2}\sum_{i=1}^{n}(x_i-\mu)\right] = \alpha^{-2}\sum(\mathbb{E}[X_i]-\mu)$$

$$I_n(\theta) = \begin{pmatrix} -\mathbb{E}\left[\frac{\partial^2}{\partial\mu^2}\ln L\right] & -\mathbb{E}\left[\frac{\partial^2}{\partial\mu\partial\alpha}\ln L\right] \\ -\mathbb{E}\left[\frac{\partial^2}{\partial\alpha\partial\mu}\ln L\right] & -\mathbb{E}\left[\frac{\partial^2}{\partial\alpha^2}\ln L\right] \end{pmatrix} = \begin{pmatrix} \frac{n}{\alpha} & 0 \\ 0 & \frac{n}{2\alpha^2} \end{pmatrix}$$

$$-\frac{n}{2\alpha^2} + \alpha^{-3}\underbrace{\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]}_{=n\alpha} = -\frac{n}{2\alpha^2} + \frac{2n}{2\alpha^2} = \frac{n}{2\alpha^2}$$

Car $\sigma^{-2}\sum(X_i-\mu)^2 \sim \chi^2_n$
$\mathbb{E}\left[\sigma^{-2}\sum(X_i-\mu)^2\right] = n$
$\Rightarrow \mathbb{E}\left[\sum(X_i-\mu)^2\right] = n\sigma^2$

$$I_n(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

On a 0 car on n'a pas d'interaction entre les paramètres.

$I_n(\mu)$

$I_n(\sigma^2)$

Inégalité de Cramer-Rao :

Soit T_n un estimateur de θ tq $\mathbb{E}[T_n] = g(\theta)$ i.e. $\forall j \in [\![1,p]\!], \mathbb{E}[T_{n,j}] = g_j(\theta)$

Soit $D_g(\theta)$ la matrice Jacobienne de g i.e. la matrice carrée de terme général $\frac{\partial}{\partial \theta_k} g_j(\theta)$ pour $(j,k) \in [\![1,p]\!]^2$

Soit $V(T_n)$ la matrice de variance de T_n.

On a besoin en plus

(HS') $I_n(\theta)$ est définie positive, alors la matrice

$$V(T_n) - \underbrace{D_g(\theta) I_n(\theta)^{-1} D_g(\theta)'}_{\text{Borne inf de l'inégalité de Cramer-Rao.}} \text{ est semi définie positive}$$

T_n est efficace si $V(T_n) = D_g(\theta) I_n^{-1}(\theta) D_g(\theta)'$

Dans le cas d'un estimateur sans biais, on a $V(T_n) - I_n(\theta)^{-1}$ est semi définie positive.

Exemple : $X_1, \ldots, X_n$ va iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

On sait que $\bar{X}_n$ est un ESB de μ et $\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ est un ESB de σ^2

On a $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ et on a $\sigma^{-2} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \sim \chi^2_{n-1} \Rightarrow \begin{cases} \mathbb{E}[\sum_{i=1}^{n} (X_i - \bar{X}_n)^2] = (n-1)\sigma^2 \\ V(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2) = 2(n-1)\sigma^4 \end{cases}$

Si $T_n = (\bar{X}_n, \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2)$ alors

$$V(T_n) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix}$$

cf TD 1 exo 4

$V(\bar{X}_n) = \frac{\sigma^2}{n}$

$V(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2) = \frac{1}{(n-1)^2} V(\sum_{i=1}^{n} (X_i - \bar{X})^2) = \frac{2\sigma^4}{n-1}$

$$I_n(\theta)^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^4}{2n} \end{pmatrix} \Rightarrow T_n \text{ est sans biais. } T_n \text{ efficace si } V(T_n) = I_n(\theta)^{-1}$$

Ici, T_n n'est pas efficace pour θ.

Généralisation de la famille exponentielle :

Définition : On dit qu'une loi de probabilité appartient à la famille exponentielle (à paramètre multidimensionnel) si sa vraisemblance s'écrit sous la forme

$$f(x,\theta) = \exp\Big(\sum_{j=1}^{p} \alpha_j(\theta) \beta_j(x) + \gamma(x) + \delta(\theta)\Big) \quad \text{si } x \in \mathcal{X}.$$

fonct° de $\mathbb{R} \to \mathbb{R}$

fonct° de $\mathbb{R}^2 \to \mathbb{R}$

Si $\forall j \in [\![1,p]\!], \alpha_j(\theta) = \theta_j$, on a la forme naturelle de la loi exponentielle

Si $\forall j \in [\![1,p]\!], \beta_j(x) = x$, on a la forme canonique de la loi exponentielle.

Exemple : $X \sim \mathcal{N}(\mu, \sigma^2)$, on a

$$f(x) = (2\pi)^{-1/2} (\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$= \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\ln(2\pi)\right)$$

ici, $\begin{cases} \alpha_1(\mu,\sigma^2) = \frac{-1}{2\sigma^2} & \beta_1(x) = x^2 \\ \alpha_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2} & \beta_2(x) = x \end{cases}$ $\quad \gamma(x) = 0 \quad \delta(\mu,\sigma^2) = \frac{-\mu^2}{2\sigma^2} - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\ln(2\pi)$

Généralisat° du maximum de vraisemblance :

Définit° : L'estimateur du maximum de vraisemblance du θ est défini par

$$\hat{\theta}_{MV} = \underset{\theta \in \Theta}{\arg\max}\ L(X_1, -, X_n, \theta)$$

Caractérisat° : Si on a (H1') à (H4') vérifiées, alors pour déterminer $\hat{\theta}_{MV}$, on résoud

$$S_n(X_1, -, X_n, \hat{\theta}_{MV}) = 0 \quad \text{i.e} \quad \begin{cases} \frac{\partial}{\partial \theta_1} \ln L(x_1, -, x_n, \theta)\big|_{\theta=\hat{\theta}_{MV}} = 0 \\ \quad | \\ \frac{\partial}{\partial \theta_p} \ln L(x_1, -, x_n, \theta)\big|_{\theta=\hat{\theta}_{MV}} = 0 \end{cases}$$

et on vérifie que la matrice Hessienne de $\ln L$ (matrice carrée de terme général $\left(\frac{\partial^2}{\partial\theta_j \partial\theta_k} \ln L\right)$ calculée en $\hat{\theta}_{MV}$ est définie négative.

Exemple : $X_1, -, X_n$ iid tq $X_1 \sim \mathcal{N}(\mu, \sigma^2 = \alpha)$

$$L(x_1, -, x_n, \mu, \alpha) = (2\pi)^{-\frac{n}{2}} \alpha^{-\frac{n}{2}} \exp\left(-\frac{1}{2\alpha} \sum_{i=1}^{n} (x_i - \mu)^2\right)$$

$$\ln L(x_1, -, x_n, \mu, \alpha) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\alpha) - \frac{1}{2\alpha}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \alpha} \ln L(x_1, -, x_n, \mu, \alpha) = \frac{-n}{2\alpha} + \frac{1}{2\alpha^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \ln L(x_1, -, x_n, \mu, \alpha) = \frac{1}{\alpha}\sum_{i=1}^{n}(x_i - \mu)$$

L'estimateur du maximum de vraisemblance de (μ, α) est tq

$$\begin{cases} -\frac{n}{2\hat{\alpha}} + \frac{1}{2\hat{\alpha}^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = 0 \\ \frac{1}{\hat{\alpha}}\sum_{i=1}^{n}(x_i - \hat{\mu}) = 0 \end{cases} \iff \begin{cases} n\hat{\alpha} = \sum_{i=1}^{n}(x_i - \hat{\mu})^2 \\ \sum_{i=1}^{n} x_i = n\hat{\mu} \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\mu}_{MV} = \frac{1}{n}\sum_{i=1}^{n} X_i \\ \hat{\sigma}_{MV}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \end{cases}$$

On calcule la matrice Hessienne

$$\frac{\partial^2}{\partial \mu^2} \ln L(x_1, -, x_n, \mu, \alpha) = -\frac{n}{\alpha}$$

$$\frac{\partial^2}{\partial \alpha^2} \ln L(x_1, \dots, x_n, \mu, \alpha) = \frac{n}{2\alpha} - \frac{1}{\alpha^3} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{\partial^2}{\partial \alpha \partial \mu} \ln L(x_1, \dots, x_n, \mu, \alpha) = -\frac{1}{\alpha^2} \sum_{i=1}^{n} (x_i - \mu)$$

$$\Rightarrow H(x_1, \dots, x_n, \mu, \alpha) = \begin{pmatrix} -\frac{n}{\alpha} & -\frac{1}{\alpha^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\ & \frac{n}{2\alpha} - \frac{1}{\alpha^3}\sum_{i=1}^{n}(x_i-\mu)^2 \end{pmatrix}$$

Pour que $(\hat{\mu}_{MV}, \hat{\alpha}_{MV})$ soit un EMV, il faut que $H(\hat{\mu}_{MV}, \hat{\alpha}_{MV})$ soit définie négative.

$$H(\hat{\mu}_{MV}, \hat{\alpha}_{MV}) = \begin{pmatrix} -\frac{n}{\hat{\alpha}_{MV}} & 0 \\ 0 & -\frac{n}{2\hat{\alpha}_{MV}^2} \end{pmatrix}$$

$$-\frac{1}{2\hat{\alpha}_{MV}^2} \underbrace{\sum_{i=1}^{n}(x_i - \hat{\mu}_{MV})}_{= \sum_{i=1}^{n} x_i - n\underbrace{\hat{\mu}_{MV}}_{= \sum x_i}} = 0$$

$$\frac{n}{2\hat{\alpha}_{MV}} - \frac{1}{\hat{\alpha}_{MV}^3} \underbrace{\sum_{i=1}^{n}(x_i - \hat{\mu}_{MV})^2}_{n\hat{\alpha}_{MV}}$$

$$= \frac{n}{2\hat{\alpha}_{MV}} - \frac{n}{\hat{\alpha}_{MV}^2} = -\frac{n}{2\hat{\alpha}_{MV}^2}$$

est bien définie négative.

9 Novembre CC.

Chapitre III : Comportement asymptotique des estimateurs

Idée : On a $X_1, \dots, X_n$ iid tq $X_i \sim \mathbb{P}_\theta$. On a un estimateur T_n de θ mais on ne peut / veut pas calculer sa loi. On regarde ce qu'il se passe quand $n \longrightarrow \infty$.

On a besoin de deux hypothèses supplémentaires :

(H6) $\theta \neq \theta' \Longrightarrow \mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ (on dit que le modèle est identifiable).

Exemple : $Y_i = a X_i + b + \varepsilon_i$ — $\varepsilon_i \sim \mathcal{N}(m, \sigma^2)$

(a, b, m) et $(a, b-m, 0)$ donne la m̂ distribut° des Y_i

(H7) $\dfrac{\partial^2}{\partial \theta^2} \ln f(x, \theta)$ est continue en θ, uniformément en x.

Théorème : Si on a les hypothèses (H1) à (H4) et (H6) vérifiées, alors il existe une suite d'estimateurs du max de vraisemblance $(\hat{\theta}_n^{MV})_n$ qui converge presque sûrement vers θ.

Rq : C'est la version de la LGN pour les EMV.

Théorème : Sous les hypothèses (H1) à (H7), on a

$$\sqrt{n}\,(\hat{\theta}_n^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta))$$

Preuve :

On considère

$$h_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\overbrace{\frac{\partial \ln f(X_i, \theta)}{\partial \theta}}^{= S(X_i, \theta)}}_{\text{V.a d'espérance nulle et de variance } I(\theta)}$$

On a $\sqrt{n}\, h_n(X_1, \dots, X_n, \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta))$ ou $\sqrt{n}\, \dfrac{h_n(X_1, \dots, X_n, \theta)}{\sqrt{I(\theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

On note

$$K_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln f(X_i, \theta) \longrightarrow I(\theta)$$

On a

$$h_n(X_1, \dots, X_n, t) = h_n(X_1, \dots, X_n, \theta) + (t - \theta) K_n(X_1, \dots, X_n, \theta_n)$$

avec $\theta^* \in \,]\min(t, \theta); \max(t, \theta)[$

On calcule en $\hat{\theta}_n^{MV}$

$$\underbrace{h_n(X_1, \dots, X_n, \hat{\theta}_n^{MV})}_{= 0 \text{ car } S_n(X_1, \dots, X_n, \hat{\theta}_n^{MV}) = 0} = h_n(X_1, \dots, X_n, \theta) + (\hat{\theta}_n^{MV} - \theta) K_n(X_1, \dots, X_n, \theta^*)$$

$$\Rightarrow \underbrace{h_n(X_1,\dots,X_n,\theta)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0,I(\theta))} + (\hat{\theta}_n^{MV}-\theta)\underbrace{K_n(X_1,\dots,X_n,\theta^*)}_{\to -I(\theta)}$$

On a $\hat{\theta}_n^{MV} \xrightarrow{p.s} \theta$
et $\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)$ continue en θ uniforme en x

$$\Rightarrow \sqrt{n}(\hat{\theta}_n^{MV}-\theta) \longrightarrow \mathcal{N}(0, I(\theta)^{-1})$$

Rappel:

Maximum de vraisemblance ($\theta \in \mathbb{R}^p$) $(\theta_1,\dots,\theta_p)$

$$\hat{\theta}_n^{MV} = \underset{\theta\in\Theta}{\arg\max}\, L(X_1,\dots,X_n,\theta)$$

Si on a toutes les hypothèses $\hat{\theta}_n^{MV}$ est tq

$$S_n(X_1,\dots,X_n,\hat{\theta}_n^{MV}) = 0$$

et la matrice Hessienne de $\ln L$ est def négative

Comportem^t asympt.

Si on a toutes les hypothèses (Surtout H1 $\mathcal{X}$ indep de θ et H_X: Identifiabilité du modèle i.e $\theta\neq\theta' \Rightarrow P_\theta \neq P_{\theta'}$

$$\hat{\theta}_n^{MV} \xrightarrow{p.s} \theta$$
$$\text{et } \sqrt{n}(\hat{\theta}_n^{MV}-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta))$$

Remarque:

* $\hat{\theta}_n^{MV} \xrightarrow{p.s} \theta$ justifie que dans la pratique on cherche une valeur qui annule l'équat° de vraisemblance et on prend cette valeur comme estimat° de $\hat{\theta}_n^{MV}$

(Rappel: $\underset{\text{v.a}}{T_n} = g_n(X_1,\dots,X_n)$ estimateur, $\underset{\text{nb}}{t_n} = g(x_1,\dots,x_n)$ estimat°)

En particulier, dans les cas "compliqués" où on ne peut pas obtenir d'express° analytiques de $\hat{\theta}_n^{MV}$, on utilise un algo. d'optimisation pour obtenir une estimat° du max de vraisemblance).

Par exemple, $X_1,\dots,X_n$ iid tq $X_i \sim GEV(\mu,\sigma,\xi)$
Generalized Extreme Value Distribution.

i.e de densité

$$f(x) = \frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{\xi+1}{\xi}} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right)$$

Sous R $\quad$ x observat°

```
loglike = function(theta){
   mu = theta[1]; sigma = theta[2]; xi = theta[3]
   resultat = sum(log(dgev(x, _)))}
optimize(loglike, _)
```

⊛ On peut écrire
$$\sqrt{n}\,\frac{\hat{\theta}_n^{MV}-\theta}{\sqrt{I^{-1}(\theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

ou
$$\sqrt{I_n(\theta)}\,(\hat{\theta}_n^{MV}-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

⊛ Si on est en dimension $p\geq 2$, en adoptant les hypothèses on a
$$\sqrt{n}\,(\hat{\theta}_n^{MV}-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \underline{I^{-1}(\theta)})$$
matrice info de Fisher.

II – Notations – Définitions

- Soit T_n un estimateur de θ.
  S'il existe Σ tq
$$\sqrt{n}\,(T_n-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)$$

On dit que T_n est asympt. normal.
La matrice Σ est la matrice de variance asymptotique de T_n.

asympt. normal

Si on a :
$$\sqrt{n}\,(T_n-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, I^{-1}(\theta))$$
alors
T_n est asympt. efficace pour θ.

Rq : L'EMV est asympt. normal et asympt. efficace.

Rapel :

⊛ Si $X_n - Y_n \xrightarrow{\mathbb{P}} 0$ et $X_n \xrightarrow{\mathcal{L}} X$ alors $Y_n \xrightarrow{\mathcal{L}} X$.

⊛ Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} a$ alors $X_n Y_n \xrightarrow{\mathcal{L}} aX$.

⊛ $a_n(X-\mu) \xrightarrow{\mathcal{L}} X$ et $a_n \to +\infty$ alors $X_n \xrightarrow{\mathbb{P}} \mu$.

Méthode Delta :

Cette méthode est utile si on a un estimateur asymptotiquement normal du paramètre θ et que je m'intéresse à $g(\theta)$

Soit g une fonct° C^1, on suppose qu'on a un estimateur T_n de θ tq :

on peut remplacer par $a_n \to +\infty$
$$\sqrt{n}\,(T_n-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2(\theta))$$
alors $g(T_n)$ converge en proba vers $g(\theta)$ et on a :
$$\sqrt{n}\,(g(T_n)-g(\theta)) \longrightarrow \mathcal{N}(0,\sigma(\theta)^2 (g'(\theta))^2)$$

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$

Maximum de vraisemblance :
$$L(x_1,\dots,x_n,\theta) = \prod_{i=1}^n f(x_i,\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta\sum_{i=1}^n x_i}$$

$$\ln L(x_1,\dots,x_n,\theta) = n\ln\theta - \theta\sum_{i=1}^n x_i$$

$$\frac{\partial}{\partial\theta}\ln L(x_1,\dots,x_n,\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

L'EMV $\hat{\theta}_n^{MV}$ est tq $\dfrac{n}{\hat{\theta}_n^{MV}} - \sum_{i=1}^{n} X_i = 0$

$$\Rightarrow \hat{\theta}_n^{MV} = \frac{n}{\sum_{i=1}^{n} X_i}$$

$$\frac{\partial^2}{\partial\theta^2}\ln L(x_1,\dots,x_n,\theta) = -\frac{n}{\theta^2} < 0$$

$\hat{\theta}_n^{MV}$ est l'EMV de θ.

$$\Downarrow$$

$$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln L(X_1,\dots,X_n,\theta)\right] = \frac{n}{\theta^2} \Rightarrow I(\theta) = \theta^{-2}$$

Comme toutes les hypothèses sont vérifiées, on a :

$$\sqrt{n}\,(\hat{\theta}_n^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta))$$

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

Méthode des moments : (on oublie tout)

Le paramètre θ est en dimension 1 donc on a besoin que de 1 équation $\mathbb{E}[X_1] = \dfrac{1}{\theta}$.

L'estimateur des moments $\hat{\theta}_n^{MV}$ est tq

$$M_1(\hat{\theta}_n^{MV}) = \bar{X}_n$$

i.e

$$\frac{1}{\hat{\theta}_n^{MV}} = \bar{X}_n \Rightarrow \hat{\theta}_n^{MV} = \frac{1}{\bar{X}_n}$$

Normalité asympt. de $\hat{\theta}_n^{MV}$?

On applique le TCL à $X_1, \dots, X_n$ i.e

$$\sqrt{n}\,\frac{\bar{X}_n - \mathbb{E}[X_1]}{\sqrt{V(X_1)}} \longrightarrow \mathcal{N}(0,1)$$

$$\Rightarrow \sqrt{n}\,\frac{\bar{X}_n - \frac{1}{\theta}}{\sqrt{1/\theta^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

dans l'écriture de la méthode Delta.

$$\Rightarrow \sqrt{n}\,(\underbrace{\bar{X}_n}_{T_n} - \overset{\theta}{\frac{1}{\theta}}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \overset{\sigma^2(\theta)}{\theta^{-2}})$$

On va appliquer la méthode Delta au TCL avec $g(x) = \frac{1}{x}$.

On a $g(\bar{X}_n) = \dfrac{1}{\bar{X}_n} = \hat{\theta}_n^{MM}$; $g(\frac{1}{\theta}) = \theta$; $g'(x) = -\dfrac{1}{x^2}$ $\quad (g'(\frac{1}{\theta}))^2 = \theta^4$

$$\Rightarrow \sqrt{n}\,\left(g(\bar{X}_n) - g(\tfrac{1}{\theta})\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \theta^{-2}(g'(\tfrac{1}{\theta}))^2\right)$$

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

On obtient bien le même résultat que le max de vraisemblance (car dans ce cas, on a $\hat{\theta}_n^{MV} = \hat{\theta}_n^{MM}$)

Méthode Delta en dimension Sup :

On considère T_n un vect aléatoire de $\mathbb{R}^k$, Σ est une matrice de variance tq

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

Soit g une fonct° C^1 alors,
$$\sqrt{n}\,(g(T_n)-g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, D_g \Sigma D_g')$$

où D_g est la matrice Jacobienne de g calculée en θ.

Rq:
Si on veut la normalité asympt. d'un estimateur des moments $\hat{\theta}_n^{MM}$ de $\theta \in \mathbb{R}^2$
On a $\hat{\theta}_n^{MM} = g(\overline{X_n}, \overline{X_n^2})$

On prend le TCL multivarié
$$\sqrt{n}\begin{pmatrix} \overline{X_n} - \mathbb{E}[X_1] \\ \overline{X_n^2} - \mathbb{E}[X_1^2]\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma) \qquad \Sigma = \begin{pmatrix} V(X_1) & Cov(X_1, X_1^2) \\ & V(X_1^2)\end{pmatrix}$$

On applique la méthode Delta au TCL et à g

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \Gamma(k, \theta)$ $\quad k, \theta \in \mathbb{R}_+^*$

i.e de densité $f(x) = \dfrac{x^{k-1}\theta^k}{\Gamma(k)} e^{-\theta x}$

On veut estimer k et θ par la méthode des moments.
On a besoin de deux moments.

$$\mathbb{E}[X] = \int x \frac{x^{k-1}\theta^k}{\Gamma(k)} e^{-\theta x}\,dx = \int \frac{x^k\theta^k}{\Gamma(k)} e^{-\theta x}\,dx = \frac{k}{\theta}\int \underbrace{\frac{x^k\theta^{k+1}}{\Gamma(k+1)} e^{-\theta x}}_{\text{densité d'une } \Gamma(k+1,\theta)}\,dx = \frac{k}{\theta}$$

$$\mathbb{E}[X^2] = \int x^2 \frac{x^{k-1}\theta^k}{\Gamma(k)} e^{-\theta x}\,dx = \frac{k(k+1)}{\theta^2}\int \frac{x^{k+1}\theta^{k+2}}{\Gamma(k+2)} e^{-\theta x}\,dx = \frac{k(k+1)}{\theta^2}$$

L'estimateur des moments $(\hat{k}_n^M, \hat{\theta}_n^M)$ est tq

$$\begin{cases} \dfrac{\hat{k}_n^M}{\hat{\theta}_n^M} = \overline{X_n} \\ \dfrac{\hat{k}_n^M(\hat{k}_n^M+1)}{(\hat{\theta}_n^M)^2} = \overline{X_n^2} \end{cases} \Rightarrow \begin{cases} \hat{\theta}_n^M = \dfrac{\hat{k}_n^M}{\overline{X_n}} \\ \dfrac{(\hat{k}_n^M+1)}{\hat{k}_n^M/(\overline{X_n})^2} = \overline{X_n^2} \end{cases} \qquad \begin{cases} \hat{\theta}_n^M = \dfrac{\hat{k}_n^M}{\overline{X_n}} \\ (\overline{X_n})^2(\hat{k}_n^M+1) = \overline{X_n^2}\,\hat{k}_n^M \end{cases} \Rightarrow \begin{cases} \hat{\theta}_n^M = \dfrac{\hat{k}_n^M}{\overline{X_n}} \\ \hat{k}_n^M(\overline{X_n^2} - \overline{X_n}^2) = \overline{X_n}^2 \end{cases} \qquad \begin{cases} \hat{\theta}_n^M = \dfrac{\overline{X_n}}{\overline{X_n^2} - \overline{X_n}^2} \\ \hat{k}_n^M = \dfrac{\overline{X_n}^2}{\overline{X_n^2} - \overline{X_n}^2} \end{cases}$$

Si on veut trouver la normalité asympt. de $(\hat{k}_n^M, \hat{\theta}_n^M)$ on applique la méthode Delta au TCL bidimensionnel appliqué à $(X_i, X_i^2)_{i\geq 1}$ et $g(x,y) = \left(\dfrac{x}{y-x^2}, \dfrac{x^2}{y-x^2}\right)$

i.e on a le TCL

$$\sqrt{n}\begin{pmatrix} \overline{X_n} - \frac{k}{\theta} \\ \overline{X_n^2} - \frac{k(k+1)}{\theta^2}\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma) \qquad \Sigma = \begin{pmatrix} V(X) & Cov(X, X^2) \\ & V(X^2)\end{pmatrix}$$

$V(X) = \dfrac{k}{\theta^2}$

$Cov(X,X^2) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = \dfrac{k(k+1)(k+2)}{\theta^3} - \dfrac{k}{\theta}\dfrac{(k+1)k}{\theta^2}$

$V(X^2) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = \dfrac{k(k+1)(k+2)(k+3)}{\theta^4} - \dfrac{(k(k+1))^2}{\theta^4}$

On prend la fonct° $g(x,y) = \left(\dfrac{x}{y-x^2}, \dfrac{x^2}{y-x^2}\right)$ avec $g\left(\dfrac{k}{\theta}, \dfrac{(k+1)k}{\theta^2}\right) = (\theta, k)$

... finir les calculs

⚠ Pour le CC du 09/11/22. On s'arrête ici, le reste ne sera pas au programme.

Chapitre IV : Estimation par intervalle de confiance

TD Actuariat

+ 10/11 AM // Access
09/11 PM → 10/11 PM // SAS
16/11 PM → 17/11 PM // SAS
30/11 PM → 28/11 PM // Access
07/12 PM → 08/12 PM // Access
14/12 PM → 12/12 PM // Économétrie

CM

09/11 PM ← CC
16/11 PM
30/11 PM

Idée : Si on a une distribution continue $\mathbb{P}_\theta$, on estime θ par T_n alors $\mathbb{P}(T_n=\theta)=0$ et en plus on sait pas si T_n est proche ou pas de θ.

$\Rightarrow$ On va donner un intervalle qui va contenir θ avec une proba très élevée (style 95% ou +)

Modèle : On se place dans le modèle
$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_\theta, \theta \in \Theta \subseteq \mathbb{R})^n$

Définition : Soit $\alpha \in [0,1]$. On appelle intervalle de confiance de niveau $1-\alpha$ de paramètre θ la donnée de 2 statistiques A_n et B_n tq : $\mathbb{P}(A_n \leq \theta \leq B_n) = 1-\alpha$.

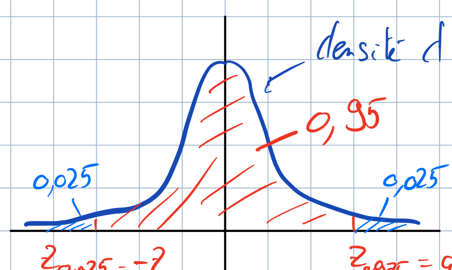
Exemple : Soit $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{N}(\theta,1)$

On veut construire un intervalle de confiance pour θ de niveau 95% i.e $\alpha = 5\%$

donc $\sqrt{n}(\bar{X}_n - \theta) \sim \mathcal{N}(0,1)$

fonct° pivot = fonct° de T_n et de θ mais dont la loi ne dépend pas de θ

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



$\Rightarrow \mathbb{P}(-z_{0,025} \leq Z \leq z_{0,975}) = 0,95$

$z_{0,025} \approx -1,96$; $z_{0,975} \approx 1,96$

$z_{0,025} = -z_{0,975}$ = quantile d'ordre 0,025 d'une $\mathcal{N}(0,1)$ i.e le réel a tq $\mathbb{P}(Z \leq a) = 0,025$

$z_{0,975}$ = quantile d'ordre 0,975 d'une $\mathcal{N}(0,1)$ i.e le réel tq $\mathbb{P}(Z \leq a) = 0,975$

Comme $\sqrt{n}(\overline{X}_n - \theta) \sim \mathcal{N}(0,1)$ on a

$$\mathbb{P}(-z_{0,975} \leq \sqrt{n}(\overline{X}_n - \theta) \leq z_{0,975}) = 0,95$$

$$\Rightarrow \mathbb{P}\left(-\frac{z_{0,975}}{\sqrt{n}} \leq \overline{X}_n - \theta \leq \frac{z_{0,975}}{\sqrt{n}}\right) = 0,95$$

$$\Rightarrow \mathbb{P}\Big(\underbrace{\overline{X}_n - \frac{z_{0,975}}{\sqrt{n}}}_{A_n} \leq \theta \leq \underbrace{\overline{X}_n + \frac{z_{0,975}}{\sqrt{n}}}_{B_n}\Big) = 0,95$$

de la def

Remarques – Commentaires :

⊛ α est un risque i.e le risque que le paramètre θ ne soit pas dans l'intervalle de confiance. C'est nous qui le choisissons.

Si α est trop petit, l'IC est très grand et pas informatif

Si α est trop grand, l'IC sert à rien car proba trop forte de ne pas contenir θ.

$\Rightarrow$ En général, on prend $\alpha = 5\%$ (ou 1% ou 10%).

Le niveau $1-\alpha$ est appelé $\begin{cases}\text{coefficient de confiance} \\ \text{niveau}\end{cases}$

⊛ On a $\mathbb{P}(A_n \leq B_n) = 1$

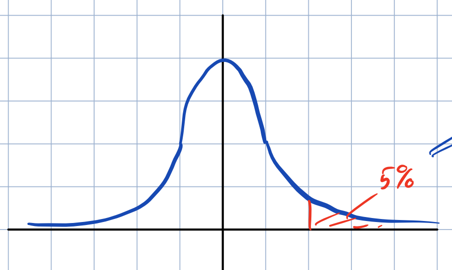
⊛ Soit $x_1, \dots, x_n$ des réalisat° de $X_1, \dots, X_n$

On a $a_n = A_n(x_1, \dots, x_n)$ et $b_n = B_n(x_1, \dots, x_n)$
On a $[a_n, b_n]$ est un intervalle réel inclus dans $\mathbb{R}$

Par contre, on a $\underbrace{\mathbb{P}(a_n \leq \theta \leq b_n)}_{\text{rien d'aléatoire.}} = 0$ ou 1.

⊛ C'est plus raisonnable d'estimer θ par IC que par estimat° ponctuelle.

⊛ L'intervalle de confiance n'est pas unique.
Si on revient à l'exemple



$\Rightarrow \mathbb{P}(Z \leq z_{0,95}) = 0,95.$

Remarque : La notion d'intervalle de confiance est liée aux tests.

On pense que le montant moyen des sinistres sur un portefeuille (i.e l'espérance de la v.a qui génère les sinistres) est 7800€

On regarde les observations et on calcule un intervalle de confiance de niveau 95% (basé sur le modèle choisi) pour $\mu = \mathbb{E}[X]$ et on trouve $[7857, 8216]$

2 possibilités
- → on n'a pas eu de chance et on est tombé sur un des 5% d'échantillons ne contenant pas μ.
- → on s'est planté et $\mu \neq 7800$ €) on retient ça.

Exemple :

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

μ et σ^2 sont inconnus et on veut un intervalle de confiance pour μ.

On ne peut pas utiliser $\sqrt{n}\,\dfrac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$

On remplace σ par $\sqrt{S_n^2}$ i.e $\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$

On a :
$$\begin{cases} \bar{X}_n \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right) \text{ i.e } \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \\ \sigma^{-2}\underbrace{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}_{(n-1)S_n^2} \sim \chi^2_{n-1} \text{ i.e } \sigma^{-2}(n-1)S_n^2 \sim \chi^2_{n-1} \end{cases}$$

Rappel : Si $Z \sim \mathcal{N}(0,1)$ $W \sim \chi^2_k$ et Z indep de W

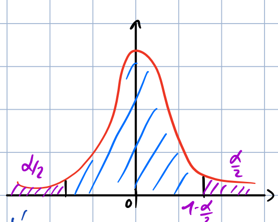
alors $\dfrac{Z}{\sqrt{W/k}} \sim t_k$

Ici :

$$\frac{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2(n-1)}}} \sim t_{n-1} \quad \text{(Student à } (n-1) \text{ degrés de liberté)}$$

$$= \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1}$$

Si $T \sim t_{n-1}$



t_{n-1} = quantile d'ordre $1-\frac{\alpha}{2}$ d'une Student à $n-1$ degrés de liberté.

Rappel : Si X_n est une suite de v.a tq $X_n \sim t_n$ alors $X_n \xrightarrow{\mathcal{L}} X$ avec $X \sim \mathcal{N}(0,1)$

$t_{n-1} = x$ tq $P(t_{n-1} \le x) = 1-\frac{\alpha}{2}$

$= qt\left(1-\frac{\alpha}{2}, df = n-1\right)$ Sous R.

quantile Student degree of freedom

t_{n-1} = quantile d'ordre $\frac{\alpha}{2}$ de la $t_{n-1} = -(t_{n-1})^{1-\frac{\alpha}{2}}$

↳ distribut° de Student est symétrique.

Si $T \sim t_{n-1}$ alors $P\left(-t_{n-1}^{1-\alpha/2} \le T \le t_{n-1}^{1-\alpha/2}\right) = 1-\alpha$

Comme $\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1}$

On a $P\left(-t_{n-1}^{1-\alpha/2} \le \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \le t_{n-1}^{1-\alpha/2}\right) = 1-\alpha$

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

$$P\Big(\underbrace{\bar{X}_n - t_{n-1}^{1-\alpha/2}\sqrt{\frac{S_n^2}{n}}}_{A_n} \le \mu \le \underbrace{\bar{X}_n + t_{n-1}^{1-\alpha/2}\sqrt{\frac{S_n^2}{n}}}_{B_n}\Big) = 1-\alpha$$

A.N : On suppose qu'on a des pertes sur des contrats $x_1, \dots, x_n$.
On suppose qu'elles sont indep et qu'elles sont les observat° de $N(\mu, \sigma^2)$. On a observé $\bar{x}_n = 8012€$ et $S_n^2 = 1616$ (Variance estimée)
↳ Valeur estimée de $\bar{X}_n$ ↳ Valeur observée de S_n^2

On veut un intervalle de confiance de niveau 95% pour μ.

Pour $n = 100$ et $\alpha = 5\%$. On a $t_{99}^{0,975} = qt(0.975, df = 99) = 1,984$

=> L'intervalle de confiance observée de niveau 95% pour μ est :

$$\left[8012 - 1,984 \times \sqrt{\frac{1616}{100}} \,;\, 8012 + 1,984 \times \sqrt{\frac{1616}{100}}\right]$$

$$= [8012 \pm 7,97] \simeq [8004 \,;\, 8020].$$

=> L'intervalle est petit, a priori on a une « bonne estimation ».

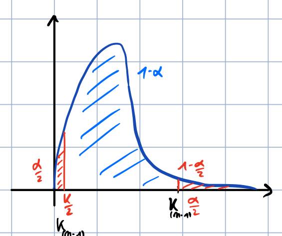
Exemple :

$X_1, \dots, X_n$ iid tq $X_i \sim N(\mu, \sigma^2)$

μ et σ^2 sont inconnus et on veut construire un intervalle de confiance de niveau $1-\alpha$ pour σ^2.

On a $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2_{n-1}$

Si $W \sim \chi^2_{n-1}$



$k_{(n-1)}$ = Quantile d'ordre $1-\frac{\alpha}{2}$ d'une χ^2_{n-1}

$= x \in \mathbb{R}$ tq $P(\chi^2_{n-1} \le x) = 1-\frac{\alpha}{2}$

$= qchisq\left(1-\frac{\alpha}{2}, df = n-1\right)$

quantile Chisquare degree of freedom

quantile d'ordre $\frac{\alpha}{2}$ d'une χ^2_{n-1}

$\Rightarrow \mathbb{P}(k_{(n-1)}^{\alpha/2} \leq W \leq k_{(n-1)}^{1-\alpha/2}) = 1-\alpha$.

Comme $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \chi^2_{n-1}$, on a $\mathbb{P}(k_{(n-1)}^{\alpha/2} \leq \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \leq k_{(n-1)}^{1-\frac{\alpha}{2}}) = 1-\alpha$.

$$\Rightarrow \mathbb{P}\Big(\underbrace{\frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{k_{(n-1)}^{1-\alpha/2}}}_{A_n} \leq \sigma^2 \leq \underbrace{\frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{k_{(n-1)}^{\alpha/2}}}_{B_n}\Big) = 1-\alpha$$

$$\mathbb{P}\Big(\frac{(n-1)S_n^2}{k_{(n-1)}^{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{k_{(n-1)}^{\alpha/2}}\Big) = 1-\alpha.$$

__A.N:__ $n = 100$, $\alpha = 5\%$, $\bar{x} = 8012$ et $S_n^2 = 1616$

Ici, $k_{n-1}^{1-\alpha/2} = \text{qchisq}(0.975, df = 99) \simeq 129$

$k_{n-1}^{\alpha/2} = \text{qchisq}(0.025, df = 99) \simeq 74$

et $(n-1)S_n^2 = 159\,984$.

L'intervalle de confiance estimé de niveau $1-\alpha$ pour σ^2 vaut :

$$\Big[\frac{(n-1)S_n^2}{k_{n-1}^{1-\frac{\alpha}{2}}} ; \frac{(n-1)S_n^2}{k_{n-1}^{\alpha/2}}\Big] = \Big[\frac{159\,984}{129} ; \frac{159\,984}{74}\Big]$$

$$\simeq [1240 ; 2162].$$

II – Intervalles de confiance asymptotique.

__Idée :__ C'est pas facile de faire des I.C de niveau exact $1-\alpha$ (on doit bien connaître l'estimateur etc.)

$\Rightarrow$ On va chercher A_n et B_n tq $\mathbb{P}(A_n \leq \theta \leq B_n) \longrightarrow 1-\alpha$.

__Définition :__ Un intervalle de confiance asymptotique ou de niveau de confiance asymptotique est deux statistiques A_n et B_n tq $\mathbb{P}(A_n \leq \theta \leq B_n) \longrightarrow 1-\alpha$.

__Utilisation du TCL__

Soit $X_1, \dots, X_n$ v.a. iid tq $X_i \sim P_\theta$, $\mathbb{E}[X_i] = \mu$ et $V(X_i) = \sigma^2$ alors on sait que :

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{quand } n \to \infty.$$

En utilisant cela, on peut construire des IC asymptotiques :

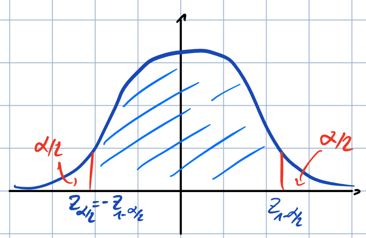
- de μ si σ est connu
- de σ si μ est connu
- de θ si μ et σ^2 dépendent de θ (Style Poisson, exponentielle ...)

Exemple : $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$ Exponentielle

On veut construire un IC de niveau asymptotique $1-\alpha$ pour θ.

On a $\sqrt{n}\,\dfrac{\bar{X}_n - 1/\theta}{\sqrt{1/\theta^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

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



On a $\mathbb{P}(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1-\alpha$

Si on a $Z_n \xrightarrow{\mathcal{L}} Z$ alors
$\mathbb{P}(-z_{1-\alpha/2} \leq Z_n \leq z_{1-\alpha/2}) \longrightarrow 1-\alpha$.

Ici on a : $\mathbb{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\,\dfrac{\bar{X}_n - 1/\theta}{\sqrt{1/\theta^2}} \leq z_{1-\alpha/2}\right) \longrightarrow 1-\alpha$.

$\Rightarrow \mathbb{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\,(\theta \bar{X}_n - 1) \leq z_{1-\frac{\alpha}{2}}\right) \longrightarrow 1-\alpha$.

$\Rightarrow \mathbb{P}\left(-\dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \theta \bar{X}_n - 1 \leq \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) \longrightarrow 1-\alpha$.

$\Rightarrow \mathbb{P}\left(\dfrac{1}{\bar{X}_n} - \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}\,\bar{X}_n} \leq \theta \leq \dfrac{1}{\bar{X}_n} + \dfrac{z_{1-\frac{\alpha}{2}}}{\bar{X}_n \sqrt{n}}\right) \longrightarrow 1-\alpha$.

Exemple : $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{P}(\lambda)$ Poisson

On a $\bar{X}_n$ est un super estimateur de λ et $\bar{X}_n \xrightarrow{\mathbb{P}} \lambda$

On a

$$\sqrt{n}\,\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Rappel (Slutsky) : Si $\left.\begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathbb{P}} 1 \end{array}\right\} \Rightarrow X_n Y_n \xrightarrow{\mathcal{L}} X$.

Ici on a $\dfrac{\lambda}{\bar{X}_n} \xrightarrow{\mathbb{P}} 1$ et $\sqrt{\dfrac{\lambda}{\bar{X}_n}} \xrightarrow{\mathbb{P}} 1$

Idée : On remplace dans le TCL le paramètre σ par un estimateur convergent : $\sqrt{n}\,\dfrac{\bar{X}_n - \mu}{\sigma} \longrightarrow \mathcal{N}(0,1)$

Ici on a : $\underbrace{\sqrt{\dfrac{\lambda}{\bar{X}_n}}}_{\downarrow \mathbb{P} \ 1} \times \underbrace{\sqrt{n}\,\dfrac{\bar{X}_n - \lambda}{\sqrt{\lambda}}}_{\downarrow \mathcal{L} \ \mathcal{N}(0,1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

$$\Longrightarrow \sqrt{n}\,\frac{\bar{X}_n-\lambda}{\sqrt{\bar{X}_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

On a $\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\bar{X}_n-\lambda}{\sqrt{\bar{X}_n}} \leq z_{1-\frac{\alpha}{2}}\right) \longrightarrow 1-\alpha.$

$$\Longrightarrow \mathbb{P}\Big(\underbrace{\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n}{n}}}_{A_n} \leq \lambda \leq \underbrace{\bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n}{n}}}_{B_n}\Big) \longrightarrow 1-\alpha.$$

Utilisation de la normalité asymptotique de l'EMV.

__Rappel:__ Si on a toutes les hypothèses alors si $\hat{\theta}_n$ est la suite d'EMV de θ.

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

ou

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

ou

$$\sqrt{I_n(\theta)}\,(\hat{\theta}_n-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

En général, on prend $\hat{I}(\theta)$ un estimateur convergent pour $I(\theta)$.
On a $\sqrt{\frac{\hat{I}(\theta)}{I(\theta)}} \xrightarrow{\mathbb{P}} 1.$ et donc $\sqrt{n}\,\frac{\hat{\theta}_n-\theta}{\sqrt{\hat{I}(\theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$

...

$$\Longrightarrow \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\hat{\theta}_n-\theta}{\sqrt{\hat{I}(\theta)^{-1}}} \leq z_{1-\frac{\alpha}{2}}\right) \longrightarrow 1-\alpha.$$

$$\Longrightarrow \mathbb{P}\left(\hat{\theta}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{I}(\theta)^{-1}}{n}} \leq \theta \leq \hat{\theta}_n + z_{1+\frac{\alpha}{2}}\sqrt{\frac{\hat{I}(\theta)^{-1}}{n}}\right) \longrightarrow 1-\alpha.$$

__Exemple:__ $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{E}(\theta)$

$$L(x_1, \dots, x_n, \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

$$\Longrightarrow \ln(L(x_1, \dots, x_n, \theta)) = n\ln(\theta) - \theta\sum_{i=1}^n x_i$$

$$\Longrightarrow \frac{\partial}{\partial\theta}\ln L = \frac{n}{\theta} - \sum_{i=1}^n x_i$$

$$\Longrightarrow \frac{\partial^2}{\partial\theta^2}\ln L = -\frac{n}{\theta^2}$$

$$\Longrightarrow \hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^n x_i} \qquad I_n(\theta) = \frac{n}{\theta^2} \Longrightarrow I(\theta) = \frac{1}{\theta^2}$$

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

On remplace θ (en bas) par $\hat{\theta}_{MV}$ (rappel: $\hat{\theta}_{MV} \xrightarrow{\mathbb{P}} \theta$ car EMV)
car $\frac{\theta}{\hat{\theta}_n} \xrightarrow{\mathbb{P}} 1$.

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

$$\Longrightarrow \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\hat{\theta}_{MV}-\theta}{\hat{\theta}_{MV}} \leq z_{1-\frac{\alpha}{2}}\right) \longrightarrow 1-\alpha.$$

$$\Longrightarrow \mathbb{P}\left(\hat{\theta}_{MV} - z_{1-\frac{\alpha}{2}}\frac{\hat{\theta}_{MV}}{\sqrt{n}} \leq \theta \leq \hat{\theta}_{MV} + z_{1-\frac{\alpha}{2}}\frac{\hat{\theta}_{MV}}{\sqrt{n}}\right) \longrightarrow 1-\alpha.$$

Rappel :

Intervalle de confiance :

Def : Un IC pour θ de niveau $1-\alpha$ est deux stats A_n et B_n tq
$$\mathbb{P}(A_n \leq \theta \leq B_n) = 1-\alpha.$$

(de niveau asympt. $1-\alpha$ si $\mathbb{P}(A_n \leq \theta \leq B_n) \longrightarrow 1-\alpha$.

Technique : On prend une fonction pivot i.e une fonction de $X_1, \ldots, X_n$ et de θ mais dont la loi ne dépend pas de θ.

Exemple : $X_i \sim \mathcal{N}(\mu, \underbrace{\sigma^2}_{\text{non connu}})$ iid

On sait que $\underbrace{\sqrt{n}\,\dfrac{\bar{X}_n - \mu}{\sqrt{S_n^2}}}_{\text{fonct° de } \mu \text{ et } X_1, \ldots, X_n} \sim \underbrace{t_{n-1}}_{\mathbb{P} \text{ indep de } \mu}$

On encadre la fonction pivot avec une probabilité $1-\alpha$ et on isole le paramètre.

Exemple : Si $T \sim t_{n-1}$, alors $\mathbb{P}(-t^{n-1}_{1-\alpha/2} \leq T \leq t^{n-1}_{1-\alpha/2}) = 1-\alpha$.

$$\Longrightarrow \mathbb{P}\left(-t^{n-1}_{1-\alpha/2} \leq \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \leq t^{n-1}_{1-\alpha/2}\right) = 1-\alpha.$$

$$\Longrightarrow \mathbb{P}\left(\bar{X}_n - t^{n-1}_{1-\alpha/2}\sqrt{\frac{S_n^2}{n}} \leq \mu \leq \bar{X}_n + t^{n-1}_{1-\alpha/2}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha.$$

Pour les IC asympt., de même sauf qu'on veut la distribut° asympt indep du paramètre.

Exemple : $X_1, \ldots, X_n$ iid tq $X_i \sim \mathcal{P}(\lambda)$

On a vu que l'EMV $\hat{\lambda}_n$ de λ est $\bar{X}_n$ et $I(\lambda) = \frac{1}{\lambda}$

Donc comme on a toutes les hypothèses, on a

$$\sqrt{n}\,(\bar{X}_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda)$$

ou

$$\underbrace{\sqrt{n}\,\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}}}_{\text{pivot de la loi asympt indep de } \lambda} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Par Slutsky, comme $\sqrt{\dfrac{\lambda}{\bar{X}_n}} \xrightarrow{\mathbb{P}} 1$, on a

$$\underbrace{\sqrt{n}\,\frac{\bar{X}_n-\lambda}{\sqrt{\bar{X}}}}_{\text{fonct° pivot de loi asympt indep de } \lambda} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Si $Z \sim \mathcal{N}(0,1)$, alors $\mathbb{P}(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1-\alpha$

on remplace Z par la fonct° pivot.

quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ de $\mathcal{N}(0,1)$

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \sqrt{n}\,\frac{\bar{X}_n-\lambda}{\sqrt{\bar{X}_n}} \leq z_{1-\alpha/2}\right) \longrightarrow 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\bar{X}_n}{n}} \leq \lambda \leq \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\bar{X}_n}{n}}\right) \longrightarrow 1-\alpha.$$

Application numérique

On a un portefeuille avec 10 000 contrats et $X_1, \dots, X_{10\,000}$ représentant le nbr de sinistre par contrat.

On suppose que les $X_i \sim \mathcal{P}(\lambda)$

L'an dernier, sur le même portefeuille, on avait estimé le paramètre à 3,6. Cette année, on observe $\bar{x}_n = 3{,}72$

$\longrightarrow$ On veut un IC à 95 % pour λ.

On applique l'exemple : $z_{1-\alpha/2} = z_{0,975} = 1{,}96$.

L'intervalle estimée est

$$\left[\bar{x}_n \pm 1{,}96\sqrt{\frac{\bar{x}_n}{n}}\right] = [3{,}682\,;\,3{,}758]$$

Si on regarde d'un point de vue estimation ponctuelle, d'une année à l'autre la distribution a changé.

D'un point de vue IC, si $3{,}6 \in IC$ alors ça signifie que 3,72 est un résultat "compatible" avec la variabilité d'une $\mathcal{P}(3{,}6)$.

Ici, $3{,}6 \notin IC$, donc la distribution a changé.

Chapitre V — Tests

Objectif : On veut construire une aide à la décision (ça aide à décider mais ne décide pas pour autant à notre place) pour choisir entre deux hypothèses H_0 et H_1.

Exemple : On a des nombres de sinistres par contrats et on a calculé les années passées $\lambda = 3{,}6$ (paramètre de la loi de Poisson), cette année on a observé $\hat{\lambda}_n = 3{,}72$.

On veut choisir entre $H_0 : \lambda = 3{,}6$ et $H_1 : \lambda \neq 3{,}6$

Exemple : On a fait une régression linéaire $(y = a_0 + a_1 x + a_2 x_2)$ et on a estimé $\begin{cases} a_0 = 4{,}12 \\ a_1 = 3{,}48 \\ a_2 = 0{,}07 \end{cases}$

On veut décider entre $H_0 : a_2 = 0$ et $H_1 : a_2 \neq 0$.

Exemple : On prend $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$

On veut choisir entre $H_0 : \mu = 0$ et $H_1 : \mu = 1$
On a observé $n = 100$ et $\bar{x} = 0{,}48$

Idée : On estime μ par $\bar{X}_n$, on rejette H_0 si $\bar{X}_n$ devient trop grand.

I — Généralités sur les tests

Le but d'un test statistique est de fournir un critère qui permet de retenir une hypothèse $H_0 : \theta \in \Theta_0$ contre une hypothèse alternative $H_1 : \theta \in \Theta_1$

La mise en œuvre d'un test statistique détermine une zone de rejet W (on rejette H_0) et une zone d'acceptation W^c (on accepte H_0).

Comme c'est basé sur un modèle proba, on a différents risques concernant les hypothèses, l'hypothèse H_0 est celle qu'on pense être vraie et qui est privilégiée. C'est l'hypothèse qu'on va rejeter uniquement si on a de vraies arguments pour le rejeter.

Parallèle : Au tribunal, un prévenu est supposé innocent jusqu'à ce qu'on amène les preuves de sa culpabilité.

H_0 : il est innocent contre H_1 : il est coupable.

⚠ Le rôle des hypothèses n'est pas symétrique.

⚠ Risques :

* Le premier risque est le **risque de première espèce** = proba. On risque de rejeter H_0 alors que H_0 est vraie

On le note $\alpha = \underbrace{\mathbb{P}(W \mid H_0)}_{\text{proba d'être dans la zone de rejet sachant que } H_0 \text{ est vraie}} \longleftarrow$ on le choisit (d'habitude)

La puissance du test est la proba de rejet H_0 avec $\beta = \mathbb{P}(W \mid H_1)$

* Le **risque de deuxième espèce** est la proba d'accepter H_0 alors que H_1 est vraie.

$$1-\beta = \mathbb{P}(W^c \mid H_1)$$

réalité \ résultat test	H_0	H_1
H_0	$1-\alpha$ ①	α ②
H_1	$1-\beta$ ③	β ④

1. proba d'accepter H_0 avec raison
2. proba de rejeter H_0 à tort
3. proba d'accepter H_0 à tort
4. proba de rejeter H_0 avec raison.

Définition : On se place dans le modèle statistique $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\theta, \theta \in \Theta)^n$

* Un **test statistique** est donné par une fonction $\phi : \mathbb{R}^n \longrightarrow \{0,1\}$

On retient H_0 si $\phi(X_1, \dots, X_n) = 0$ et on rejette H_0 si $\phi(X_1, \dots, X_n) = 1$.

* La **zone de rejet W** est l'ensemble $\{(x_1, \dots, x_n) \in \mathbb{R}^n,\ \phi(x_1, \dots, x_n) = 1\}$

Si on a une région de rejet W alors $\phi = \mathbb{1}_W$

Si $H_0 : \Theta = \{\theta_0\}$ et $H_1 : \Theta = \{\theta_1\}$, on parle d'hypothèses simples.

* Le **niveau d'un test**, ou **sensibilité**, est la proba de rejeter H_0 à tort.

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta((X_1, \dots, X_n) \in W)$$

Si on a $H_0 : \theta = \theta_0$
$\alpha = \mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in W)$

* La puissance du test est la fonction $\beta : \Theta_1 \longrightarrow [0,1]$ tel que
$\beta(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in W)$

Si on a $H_1 : \theta = \theta_1$
$\beta = \mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in W)$
(spécificité du test).

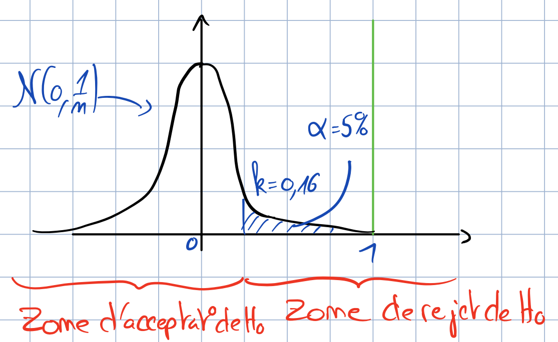
Le test est sans biais si :
$\forall \theta \in \Theta_1,\ \beta(\theta) \geq \alpha$

Exemple : On reprend l'exemple où $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$. Applicat° numérique, on a $n = 100$ et $\bar{x}_n = 0{,}28$.

On veut choisir entre $H_0 : \mu = 0$ et $H_1 : \mu = 1$.

Idée : On estime μ par $\bar{X}_n$ et on rejette H_0 si $\bar{X}_n$ s'éloigne trop de 0 (vers 1) ou $\Leftrightarrow$ on rejette H_0 si $\bar{X}_n > k$ ← à déterminer

Sous H_0 (i.e. $\mu = 0$), $\bar{X}_n \sim \mathcal{N}(0, 1/n)$



On a $\alpha = \mathbb{P}(\text{rejeter } H_0 \mid H_0 \text{ est vraie}) = 5\% = 0{,}05$
$= \mathbb{P}(\bar{X}_n > k \mid \bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})) = \mathbb{P}(\mathcal{N}(0, \frac{1}{n}) > k) = \alpha$

$$\alpha = \mathbb{P}_0(\underbrace{\sqrt{n}\,\bar{X}_n}_{\sim \mathcal{N}(0,1)} > \underbrace{\sqrt{n}\,k}_{z_{1-\alpha}})$$

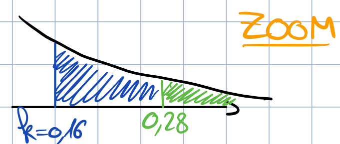
$z_{1-\alpha}$ quantile d'ordre $1-\alpha$ d'une $\mathcal{N}(0,1)$.

Si on prend $\alpha = 0{,}05$, $z_{0,95} = 1{,}6$, $n = 100$ $\longrightarrow$ $k \simeq 0{,}16$.

On a observé $0{,}28$, $0{,}28 > 0{,}16 \Rightarrow$ On rejette H_0.

Quelle est la valeur minimale de α pour qu'on accepte H_0 alors qu'on a observé $0{,}28$?

Le plus petit α tq on accepte H_0
en observant $0{,}28$ = p.value
$\mathbb{P}(\bar{X}_n \geq 0{,}28 \mid H_0)$
$= \mathbb{P}(\mathcal{N}(0, \frac{1}{n}) > 0{,}28)$



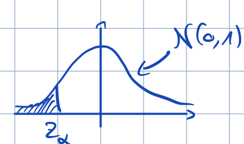
Réponse : Le plus petit α tq on accepte H_0 si on observe $0{,}28$
est $\mathbb{P}(\mathcal{N}(0, \frac{1}{n}) > 0{,}28) = \mathbb{P}(\mathcal{N}(0,1) > \sqrt{n}\,0{,}28)$
$\underset{n=100}{=} \mathbb{P}(\mathcal{N}(0,1) > 2{,}8) = 0{,}0026$
$= p.\text{value}$.

$\Rightarrow$ On accepte H_0 si la p.value est supérieure à α et on rejette si elle est inférieure à α.

Remarque : Les hypothèses ne sont pas symétriques :

Parce que H_0 est l'hypothèse privilégiée et H_1 n'intervient pas dans la construction.

Si on inverse les hypothèses : $H_0 : \mu = 1$ contre $H_1 : \mu = 0$
On rejette H_0 quand $\bar{X}_n < k$ i.e. on cherche k tq $\mathbb{P}(\bar{X}_n < k \mid \mu = 1) = \alpha$
$$\Rightarrow \mathbb{P}(\underbrace{\sqrt{n}\,(\bar{X}_n - 1)}_{\sim \mathcal{N}(0,1)} < \underbrace{\sqrt{n}\,(k-1)}_{z_\alpha} \mid \mu = 1) = \alpha$$



Si $\alpha = 5\%$, $n = 100$, on a $10(k-1) = -1{,}6$
$\Rightarrow \underline{k = 0{,}84}$

$\Longrightarrow$ On rejette H_0 si $\bar{x}_m < 0,84$
Comme on a observé $\bar{x} = 0,28$, on rejette H_0

Conclusion : Si $\bar{x}_m = 0,28$

Il faut faire attention à comment on choisit les hypothèses

$\begin{cases} H_0 : \mu = 0 \text{ et } H_1 : \mu = 1 \Rightarrow \text{on rejette } H_0 \\ H_0 : \mu = 1 \text{ et } H_1 : \mu = 0 \Rightarrow \text{on rejette } H_0 \end{cases}$

II – Tests uniformément plus puissants.

Définition : Étant donné deux tests ϕ_1 et ϕ_2 de niveau $< \alpha$ pour tester

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_1 .$$

Le test ϕ_1 est uniformément plus puissant que ϕ_2 ssi

$$\forall \theta \in \Theta_1, \ \beta_1(\theta) \geq \beta_2(\theta)$$

Dans le cas où $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1$, on dit que ϕ_1 est plus puissant que ϕ_2 si $\beta_1 \geq \beta_2$

On prend les hypothèses $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$

Soit $L(x_1, \dots, x_n, \theta)$ la fonction de vraisemblance.

Définition : On considère $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$

Soit $V_{\theta_0, \theta_1} = \dfrac{L(x_1, \dots, x_n, \theta_1)}{L(x_1, \dots, x_n, \theta_0)}$ le rapport de vraisemblance.

On considère la famille du test ϕ_α de région de rejet $V_{\theta_0, \theta_1} > k_\alpha$

On appelle ces tests des tests de Neyman-Pearson

Remarque : $L(x_1, \dots, x_n, \theta)$ est le "degré de vraisemblance" de θ sachant qu'on a observé $x_1, \dots, x_n$
i.e $\simeq$ la proba d'observer $x_1, \dots, x_n$ sous l'hypothèse θ.

Dans un test de NP, on rejette H_0 si θ_1 devient plus vraisemblable que θ_0

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$, on veut $H_0 : \mu = 0$ contre $H_1 : \mu = 1$

On veut construire un test de Neyman-Pearson

$$L(x_1, \dots, x_n, \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

$$\Rightarrow L(x_1,\dots,x_n,\mu) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$V_{0,1} = \frac{L(x_1,\dots,x_n,1)}{L(x_1,\dots,x_n,0)} = \frac{(2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-1)^2\right)}{(2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_i^2\right)}$$

$$= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i^2-2x_i+1) + \frac{1}{2}\sum_{i=1}^{n}x_i^2\right)$$

$$= \exp\left(\sum_{i=1}^{n} x_i - \frac{n}{2}\right)$$

On rejette H_0 si $\exp\left(\sum_{i=1}^{n} x_i - \frac{n}{2}\right) > k_\alpha$

On détermine k_α tel que

$$\mathbb{P}\left(\exp\left(\sum_{i=1}^{n} X_i - \frac{n}{2}\right) > k_\alpha \,\middle|\, \underbrace{H_0}_{X_i \sim \mathcal{N}(0,1)}\right) = \alpha$$

Donc on cherche k_α tq :

$$\mathbb{P}_0\left(\exp\left(\sum_{i=1}^{n} X_i - \frac{n}{2}\right) > k_\alpha\right) = \alpha \quad \text{avec } X_i \sim \mathcal{N}(0,1)$$

$$\Leftrightarrow \mathbb{P}_0\left(\sum_{i=1}^{n} X_i > \ln(k_\alpha) + \frac{n}{2}\right) = \alpha$$

$$\Leftrightarrow \mathbb{P}_0\left(\underbrace{\sqrt{n}\,\bar{X}_n}_{\sim\mathcal{N}(0,1)} > \underbrace{\frac{\ln(k_\alpha)}{\sqrt{n}} + \frac{\sqrt{n}}{2}}_{z_{1-\alpha}}\right) = \alpha$$

$\Rightarrow$ On rejette H_0 si $\sqrt{n}\,\bar{x}_n > z_{1-\alpha}$ où $\bar{x}_n = \frac{z_{1-\alpha}}{\sqrt{n}}$

On retombe sur le régime de rejet qu'on avait trouvé avant.

Théorème Neyman - Pearson

* Soit $\alpha > 0$ et ϕ_{k} un test de Neyman-Pearson de niveau α alors il est plus puissant que tout autre test de niveau $\leq \alpha$ et en plus il est sans biais.
  ↳ C'est le meilleur.

* Si $\alpha \in]0,1[$, si les lois $\mathbb{P}_\theta$ sont absolument continues, il existe $k_\alpha \in \mathbb{R}$ tq ϕ_{k_α} est de niveau α.

  Si les lois sont discrètes alors il existe un plus petit k_α tq ϕ_{k_α} est de niveau $\leq \alpha$.

* Soit ϕ un test plus puissant de niveau α alors $\forall \theta \in \{\theta_0, \theta_1\}$,

$$\mathbb{P}_\theta\left(\phi(X) \neq \phi_{k_\alpha}(X) \text{ et } V(X) \neq k\right) = 0.$$

Rappel : On veut choisir entre deux hypothèses

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

Définition : Un test est une fonct° $\phi : \mathbb{R}^n \longrightarrow \{0,1\}$

(Si $\phi(x_1, \dots, x_n) = 1$, on rejette H_0 et inversement)

W = région de rejet $= \{(x_1, \dots, x_n) \in \mathbb{R}^n \text{ tq } \phi(x_1, \dots, x_n) = 1\}$

Si on veut tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$

$\Longrightarrow$ Test de Neyman-Pearson (test plus puissant dans ce cas)

On calcule $V_{\theta_0,\theta_1} = \dfrac{L(x_1, \dots, x_n, \theta_1)}{L(x_1, \dots, x_n, \theta_0)}$

On rejette H_0 si V_{θ_0,θ_1} si $V_{\theta_0,\theta_1} \geq k_\alpha$ et k_α est déterminé par le fait que

$$\underset{\text{Sous l'hypothèse } H_0}{\mathbb{P}_{\theta_0}} \big(\underbrace{V_{\theta_0,\theta_1} \geq k_\alpha}_{\text{rejet de } H_0} \big) = \alpha$$

risque de $1^{ère}$ espèce = proba de rejeter H_0 alors que H_0 est vraie

$\leadsto$ on en déduit k_α

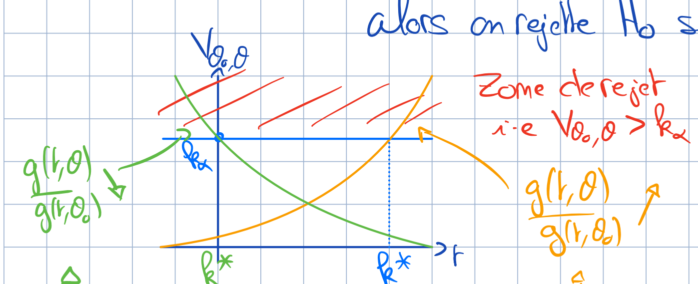
Si on veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$ (ou $H_1 : \theta < \theta_0$)

Idée : On fait Neyman-Pearson avec $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1 > \theta_0$ $\leftarrow$ tous les θ_1

Si on a une stat exhaustive T_n de vraisemblance g.

$$V_{\theta_0,\theta} = \frac{L(x_1, \dots, x_n, \theta)}{L(x_1, \dots, x_n, \theta_0)} = \frac{g(t, \theta)}{g(t, \theta_0)} \quad \text{monotone en } t.$$

alors on rejette H_0 si $\{T_n \geq k\}$ si croissante et $\{T_n \leq k\}$ si décroissante



$\{V_{\theta_0,\theta} \geq k_\alpha\} \Leftrightarrow \{t \leq k^*\}$

$\{V_{\theta_0,\theta} \leq k_\alpha\} \Leftrightarrow \{t \geq k^*\}$

On détermine k^* tq $\mathbb{P}_{\theta_0}(T > k^*) = \alpha$ ou $\mathbb{P}_{\theta_0}(T \leq k^*) = \alpha$

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, 1)$

On veut tester $H_0 : \mu = 0$ contre $H_1 : \mu > 0$

$$L(x_1, \dots, x_n, \mu) = (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2)\right]$$

$$V_{0,\mu} = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(-2x_i\mu - \mu^2)\right] = \exp\left[\mu \sum x_i - \frac{n\mu^2}{2}\right].$$

On sait que $\overline{X}_n$ est une stat exhaustive.

Si on pose $t = \frac{1}{n}\sum_{i=1}^{n} x_i$

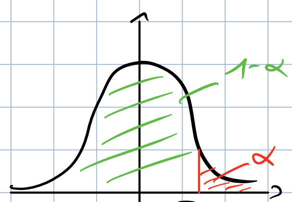
$V_{0,\mu} = \exp\left[n\mu t - \frac{n\mu^2}{2}\right]$ fonction croissante de t.

On rejette si $t > k$ avec k tq $\mathbb{P}_0(\overline{X}_n > k) = \alpha$

Sous H_0, $\overline{X}_n \sim \mathcal{N}(0, \frac{1}{n})$ donc $\sqrt{n}\,\overline{X}_n \sim \mathcal{N}(0,1)$

k tq $\mathbb{P}_0(\overline{X}_n > k) = \alpha \iff \mathbb{P}_0(\sqrt{n}\,\overline{X}_n > \sqrt{n}\,k) = \alpha.$

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



$Z_{1-\alpha}$ = quantile d'ordre $1-\alpha$

$$\mathbb{P}(Z > Z_{1-\alpha}) = 1-\alpha \implies \mathbb{P}_0(\sqrt{n}\,\overline{X}_n > \underbrace{Z_{1-\alpha}}_{=\sqrt{n}\,k}) = 1-\alpha$$

Remarque : On retombe sur la même région de rejet que pour tester $H_0 : \mu = 0$ contre $H_1 : \mu = 1$ (ou n'importe quoi > 0)

La région de rejet c'est $\{\sqrt{n}\,\overline{x}_n \geq Z_{1-\alpha}\}$

Application numérique : On a 100 observations qui viennent d'une loi $\mathcal{N}(\mu, 1)$.

On a observé $\overline{x}_n = 0{,}196$ et on veut tester $H_0 : \mu = 0$ contre $H_1 : \mu > 0$ de niveau $\alpha = 5\%$

On calcule la statistique du test $(\sqrt{n}\,\overline{X}_n)$, $\sqrt{n}\,\overline{x}_n = 1{,}96$.

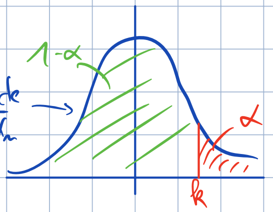
Si on a $\alpha = 5\%$, alors $Z_{1-\alpha} = Z_{0,95} \simeq 1{,}6$.
Ici, $\sqrt{n}\,\overline{x}_n = 1{,}96 > 1{,}6 = Z_{1-\alpha} = Z_{0,95}$. On rejette H_0 de niveau $\alpha = 5\%$

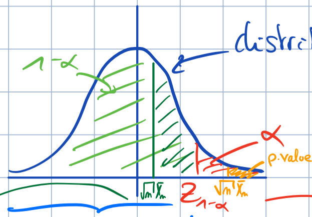
A quel niveau de α est-ce qu'on aurait accepté le test ?

$\Rightarrow$ p.value

On cherche le plus petit α tq $\sqrt{n}\,\overline{x}_n = Z_{1-\alpha}$ équivalent à calculer $\mathbb{P}(\underbrace{\sqrt{n}\,\overline{X}_n}_{\text{stat théorique}} > \underbrace{\sqrt{n}\,\overline{x}_n}_{\text{valeur qu'on observe}})$

Loi de $\sqrt{n}\,\overline{X}_n$





distribut° de $\sqrt{n}\,\overline{X}_n = \mathcal{N}(0,1)$ ici

On veut calculer la proba :
p.value $= \mathbb{P}(\sqrt{n}\,\overline{X}_n > \sqrt{n}\,\overline{x}_n)$

Si $\sqrt{n}\,\overline{X}_n < z_{1-\alpha}$
On accepte H_0 et
p.value $< \alpha$.

Si $\sqrt{n}\,\overline{X}_n < z_{1-\alpha}$
On ne rejette pas H_0

Si $\sqrt{n}\,\overline{X}_n \geq z_{1-\alpha}$
On rejette H_0

Si $\sqrt{n}\,\overline{X}_n > z_{1-\alpha} \Rightarrow$ p.value $< \alpha$

Ici, on a $\sqrt{n}\,\overline{x}_n = 1{,}96$.
p.value $= \mathbb{P}(\sqrt{n}\,\overline{X}_n > 1{,}96) = 0{,}025$

$\longrightarrow$ p.value = plus petit α tq on acceptait H_0 i.e on aurait accepté H_0 au niveau $\alpha = 2{,}5\%$

$\Rightarrow$ La p.value permet de contrôler la décision qu'on prend.
Si p.value très grande $\Longrightarrow$ on accepte H_0
.... très petit $\Longrightarrow$ on rejette H_0
.... pas trop loin de $\alpha = 5\%$ $\Longrightarrow$ on réfléchit.

III - Tests fondés sur le rapport du maximum de vraisemblance.

On appelle test du rapport du max de vraisemblance tout test de région de rejet

$$W = \left\{ x \in \mathbb{R}^n, \ \frac{\sup_{\theta \in \Theta_1} L(x_1, \dots, x_n, \theta)}{\sup_{\theta \in \Theta_0} L(x_1, \dots, x_n, \theta)} > k_\alpha \right\}$$

pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ avec k_α tq $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(W) = k_\alpha$

Dans le cas où $\Theta_1 = \Theta_0^c$; $W = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \text{ tq } \frac{\sup_{\theta \in \Theta} L(x_1, \dots, x_n, \theta)}{\sup_{\theta \in \Theta_0} L(x_1, \dots, x_n, \theta)} > k_\alpha \right\}$

Résumé :

$H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1 \Longrightarrow$ Neyman-Pearson = test plus puissant

$H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0 \Longrightarrow V_{\theta_0, \theta} + T$ exhaustive = test unif plus puissant.

$H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1 \Longrightarrow$ Test rapport max vraisemblance = pas de propriété

Exemple : $X_1, \dots, X_n$ iid tq $X_i \sim \mathcal{N}(\mu, \sigma^2)$

On veut tester $H_0 : \mu = \mu_0$

⊛ $H_1 : \mu = \mu_1 \Longrightarrow$ déjà fait

⊛ $H_1 : \mu > \mu_0 \Rightarrow$ déjà fait si σ connue

Si σ^2 inconnue, on a toujours $V_{\mu_0,\mu} \nearrow$ en $t = \bar{x}$
$\Rightarrow$ on rejette si $\bar{x}_m > k$ avec $\mathbb{P}_{\mu_0}(\bar{X}_m > k_\alpha) = \alpha$.

Sous H_0, $\bar{X}_m \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{m})$.
On a :
$$\sqrt{m}\, \frac{\bar{X}_m - \mu_0}{\sigma} \sim \mathcal{N}(0,1).$$
σ ← inconnue.

On remplace σ^2 par $S_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X}_m)^2$

Rappel : Si :
- ⊛ $Z \sim \mathcal{N}(0,1)$
- ⊛ $W \sim \chi^2_a$
- ⊛ $Z \perp W$

$\Rightarrow \frac{Z}{\sqrt{W/a}} \sim t_a$

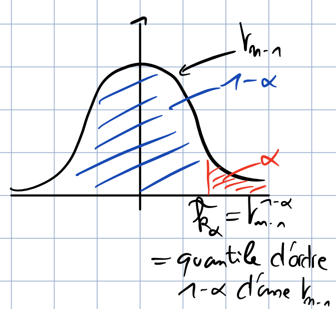
⚠ Parce que ce sont des $X_i \sim$ Normale

On sait que $\sigma^{-2} \sum_{i=1}^{m} (X_i - \bar{X}_m)^2 \sim \chi^2_{m-1}$
et $\bar{X}_m$ est indépendant de $\sum_{i=1}^{m} (X_i - \bar{X}_m)^2$

$$\frac{\overbrace{\sqrt{m}\, \frac{\bar{X}_m - \mu_0}{\sigma}}^{\mathcal{N}(0,1)}}{\sqrt{\frac{\sigma^{-2}}{m-1} \underbrace{\sum_{i=1}^{m} (X_i - \bar{X}_m)^2}_{\chi^2_{m-1}}}} \sim t_{m-1} \Rightarrow \sqrt{m}\, \frac{\bar{X}_m - \mu_0}{\sqrt{S_m^2}} \sim t_{m-1}$$

À la base, on cherchait k_α tq $\mathbb{P}_{\mu_0}(\bar{X}_m > k_\alpha) = \alpha$ équivaut à chercher
$$\mathbb{P}_{\mu_0}\left(\sqrt{m}\, \frac{\bar{X}_m - \mu_0}{\sqrt{S_m^2}} > \tilde{k}_\alpha\right) = \alpha. \Rightarrow \tilde{k}_\alpha = t_{m-1}^{1-\alpha}$$

Si $T \sim t_{m-1}$,



On rejette donc H_0 si :
$$\sqrt{m}\, \frac{\bar{x}_m - \mu_0}{\sqrt{s_m^2}} > t_{m-1}^{1-\alpha}$$

⊛ On veut tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$ (test bilatéral)

On ne peut pas faire un test unif plus puissant mais on peut quand même en faire un basé sur le rapport du max de vraisemblance.

Ici, $\Theta = \{(\mu, \sigma^2) \text{ avec } \mu \in \mathbb{R} \text{ et } \sigma^2 > 0\}$ ← on a (μ, σ^2) paramètres

$\Theta_0 = \{(\mu_0, \sigma^2) \text{ avec } \sigma^2 > 0\}$ ← que σ^2 paramètre

$\Theta_1 = \{(\mu, \sigma^2) \text{ avec } \mu \in \mathbb{R} / \{\mu_0\}, \sigma^2 > 0\}$

On a $\Theta_1 = \Theta_0^c$. Le test va être basé sur la zone de rejet

$$W = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n, \ \frac{\overbrace{\sup_{(\mu,\sigma^2)\in\Theta} L(x_1,\dots,x_n,\mu,\sigma^2)}^{= L(x_1,\dots,x_n,\hat\mu_n,\hat\sigma^2)}}{\underbrace{\sup_{\sigma^2\in\Theta_0} L(x_1,\dots,x_n,\mu_0,\sigma^2)}_{= L(x_1,\dots,x_n,\mu_0,\hat\sigma_0^2)}} > k \right\}$$

Si l'EMV sur Θ est $(\hat\mu, \hat\sigma_n)$ et l'EMV sur Θ_0 est $\hat\sigma_0^2$

On a déjà vu que $\hat\mu_n = \bar X_n$ et $\hat\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar X_n)^2$ (On estime par MV (μ,σ^2) d'une $\mathcal{N}(\mu,\sigma^2)$)

$$\hat\sigma_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$$ (On estime par MV σ^2 quand μ_0 est connue)

On sait aussi que $L(x_1,\dots,x_n,\mu,\sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right]$.

$$\lambda(x_1,\dots,x_n) = \frac{\sup_\Theta L(x_1,\dots,x_n,\mu,\sigma^2)}{\sup_{\Theta_0} L(x_1,\dots,x_n,\mu_0,\sigma^2)} = \frac{L(x_1,\dots,x_n,\hat\mu_n,\hat\sigma_n^2)}{L(x_1,\dots,x_n,\mu_0,\hat\sigma_0^2)}$$

$$= \frac{(2\pi)^{-n/2}(\hat\sigma_n^2)^{-n/2}\overbrace{\exp\left[-\frac{1}{2\hat\sigma_n^2}\sum_1^n (x_i-\bar x_n)^2\right]}^{=\exp[-\frac{n}{2}]}}{(2\pi)^{-n/2}(\hat\sigma_0^2)^{-n/2}\underbrace{\exp\left[-\frac{1}{2\hat\sigma_0^2}\sum_1^n (x_i-\mu_0)^2\right]}_{=\exp[-\frac{n}{2}]}}$$

$$= \left(\frac{\hat\sigma_0^2}{\hat\sigma_n^2}\right)^{\frac{n}{2}}$$

$$= \left(1 + \frac{(\bar x_n - \mu_0)^2}{\hat\sigma_n^2}\right)^{\frac{n}{2}}$$

$$\hat\sigma_0^2 = \frac{1}{n}\sum_1^n (x_i - \mu)^2$$
$$= \frac{1}{n}\sum_1^n (x_i - \bar x_n + \bar x_n - \mu)^2$$
$$= \underbrace{\frac{1}{n}\sum_1^n (x_i - \bar x_n)}_{\hat\sigma_n^2} + (\bar x_n - \mu_0)^2$$

Si on pose $t_n = \frac{\bar x_n - \mu_0}{\sqrt{\hat\sigma_n^2}}$

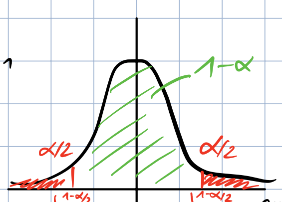
On a $\lambda(x_1,\dots,x_n)$ est une fonct° croissante de $|t_n|$

$\{\lambda(x_1,\dots,x_n) > k_\alpha\}$ équivalent à $\{|t_n| > \tilde k_\alpha\}$

Sous H_0 $(\mu=\mu_0)$, $T_n = \underbrace{\sqrt{n-1}\,\frac{\overbrace{\bar X_n - \mu_0}^{\mathcal{N}(0,1)}}{\sqrt{\hat\sigma_n^2}}}_{\approx\sqrt{\chi^2_{n-1}}} \sim t_{n-1}$ car $T_n = \frac{\sqrt{n}\,\frac{\bar X_n - \mu_0}{\sigma}}{\sqrt{\frac{\sigma^{-2}}{n(n-1)}\sum_1^n (X_i-\bar X)^2}}$ $\begin{matrix}\leftarrow \mathcal{N}(0,1) \\ \leftarrow \chi^2_{n-1}\end{matrix}$

On cherche $\tilde k_\alpha$ tq sous H_0, $\mathbb{P}_{\mu_0}(\underbrace{|T_n|}_{\sim t_{n-1}} > \tilde k_\alpha) = \alpha$

Si $T \sim t_{n-1}$



$-t_{n-1}^{1-\alpha/2}$

$t_{n-1}^{1-\alpha/2}$ = quantile d'ordre $1-\alpha/2$ d'une t_{n-1}

On rejette donc H_0 si $\left|\sqrt{n-1}\,\frac{\bar{x}-\mu_0}{\sqrt{\hat{\sigma}_n^2}}\right| > t_{n-1}^{1-\alpha/2}$

On calcule la p.value comme $\mathbb{P}_{\mu_0}\left(\left|\sqrt{n-1}\,\frac{\bar{X}_n-\mu_0}{\sqrt{\hat{\sigma}_n^2}}\right| > \left|\sqrt{n-1}\,\frac{\bar{x}_n-\mu_0}{\sqrt{\hat{\sigma}_n^2}}\right|\right)$

$$= \mathbb{P}_{\mu_0}\left(t_{n-1} \geq \left|\sqrt{n-1}\,\frac{\bar{x}_n-\mu_0}{\sqrt{\hat{\sigma}_n^2}}\right|\right)$$

Exemple : Comparaison de moyenne.

On a deux echantillons $(X_1, \dots, X_{n_x})$ et $(Y_1, \dots, Y_{n_y})$ indep entre eux tq

$$X_i \sim \mathcal{N}(\mu_x, \sigma^2) \text{ et } Y_i \sim \mathcal{N}(\mu_y, \sigma^2)$$

On veut tester $H_0 : \mu_x = \mu_y$ contre $H_1 : \mu_x \neq \mu_y$

On a $\Theta = \{(\mu_x, \mu_y, \sigma^2)$ avec $\mu_x \in \mathbb{R}, \mu_y \in \mathbb{R}$ et $\sigma^2 > 0\}$

L'estimateur du max de vraisemblance (sur Θ) est $\hat{\mu}_x = \bar{X}_n = \frac{1}{n_x}\sum_{i=1}^{n_x} X_i$, $\hat{\mu}_y = \bar{Y}_n = \frac{1}{n_y}\sum_{i=1}^{n_y} Y_i$
et $\hat{\delta} = \frac{1}{n}\left(\sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)$ avec $n = n_x + n_y$

On a $\Theta_0 = \{(\mu, \sigma^2), \mu \in \mathbb{R}$ et $\sigma^2 > 0\}$

L'EMV sur Θ_0 est $\hat{\mu}_0 = \frac{1}{n}\left(\sum_{i=1}^{n_x} X_i + \sum_{i=1}^{n_y} Y_i\right)$

$$\text{et } \hat{\delta}_0 = \frac{1}{n}\left(\sum_{i=1}^{n_x}(X_i-\hat{\mu}_0)^2 + \sum_{i=1}^{n_y}(Y_i-\hat{\mu}_0)^2\right)$$

La vraisemblance vaut $L(x_1, \dots, x_n, \mu_x, \mu_y, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n_x}(x_i-\mu_x)^2\right]\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n_y}(y_i-\mu_y)^2\right]$

$$\lambda(x_1, \dots, y_{n_y}) = \frac{\sup_{\Theta} L(x_1, \dots, y_{n_y}, \mu_x, \mu_y, \sigma^2)}{\sup_{\Theta_0} L(x_1, \dots, y_{n_y}, \mu, \sigma^2)} = \left(\frac{\hat{\delta}_0}{\hat{\delta}_n}\right)^{\frac{n}{2}}$$

On pose $T_n = \sqrt{\frac{n_x n_y}{n}}\,\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{n}{n-2}\hat{\delta}_n}}$

On a $T_n \sim t_{n-2}$, on mq $\lambda(x_1, \dots, y_{n_y})$ est une fonction croissante en $|t_n|$

$\Longrightarrow$ On rejette si $|t_n| > t_{n-2}^{1-\alpha/2}$ $\longleftarrow$ quantile d'ordre $1-\alpha/2$ d'une t_{n-2}.

Ref à l'exam