Chap 1 : ESTIMATION

I Introduction :

L'estimation consiste à donner des valeurs approchées à 1 ou plusieurs paramètres (souvent noté $\theta \in \mathbb{R}$ ou $\mathbb{R}^n$) à partir d'observations notées généralement $x_1, \dots, x_n$ issues d'une même "population"

<u>Le modèle statistique</u> retenu par la plupart des statisticiens consiste à supposer que $x_1, \dots, x_n$ (observations) est la réalisation d'un vecteur $(X_1, \dots, X_n)$ <u>où X_i sont des variables aléatoires indépendantes et de même loi</u> (va iid).

On dit que $(X_1, \dots, X_n)$ est un <u>échantillon</u>.

Un <u>estimateur</u> d'un paramètre $\theta \in \mathbb{R}$ ou $\mathbb{R}^n$ est une fonction des n VA X_i notée $\hat{\theta}$ ou $\hat{\theta}(X_1, \dots, X_n)$

<u>Remarque</u> : $\hat{\theta}(X_1, \dots, X_n)$ est une <u>fonction aléatoire</u> appelée <u>statistique</u>

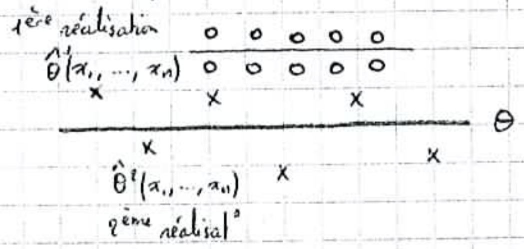
<u>Qualités d'un estimateur : $\hat{\theta}$</u>

- **Biais de l'estimateur :**

$$\text{Biais}(\hat{\theta}) = \underbrace{E[\hat{\theta}]}_{\text{moyenne de l'estimé}} - \underbrace{\theta}_{\text{la vraie valeur}}$$

moyenne → non aléatoire

C'est une fonction non aléatoire de θ notée $b(\theta)$



l'estimateur X est non biaisé : pas d'erreur systématique.

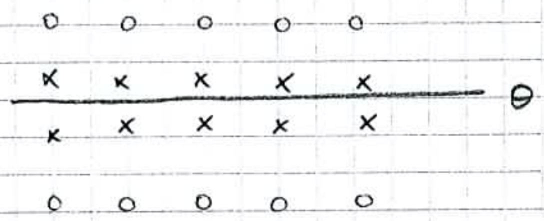
l'estimateur o est biaisé : il y a une erreur systématique. (ex : balance car 0 mal réglé)

Dans les applications pratiques, on veut $E[\hat{\theta}] - \theta = 0$; on dit que l'estimateur est <u>non biaisé</u>. On veut biais faible

- **Variance de l'estimateur**

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

Var $\hat{\theta}$ fournit une précision sur l'estimation. Plus var $\hat{\theta}$ est faible meilleur est l'estimateur. On veut var faible



Var X < Var O

$\Rightarrow$ X préférable

- Erreur quadratique moyenne d'un estimateur :

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

l'estimateur $\hat{\theta}$ est d'autant meilleur que $EQM(\hat{\theta})$ est faible

Remarque : $\boxed{EQM(\hat{\theta}) = Var\,\hat{\theta} + \left(Biais(\hat{\theta})\right)^2}$

Si on a var $\hat{\theta}$ et Biais $\hat{\theta}$, pas la peine de calculer $EQM(\hat{\theta})$

En effet $EQM(\hat{\theta}) = E\left[\left(\hat{\theta}-E[\hat{\theta}] + E[\hat{\theta}]-\theta\right)^2\right]$

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

donc $EQM(\hat{\theta}) = Var\,\hat{\theta} + \underbrace{E\left[(Biais\,\hat{\theta})^2\right]}_{(Biais\,\hat{\theta})^2} + 2\left(E[\hat{\theta}]-\theta\right)\underbrace{E\left[\hat{\theta}-E[\hat{\theta}]\right]}_{E[\hat{\theta}]-E[E[\hat{\theta}]]}$

$E[\hat{\theta}] - E[\hat{\theta}] = 0$

Définition : On dit que l'estimateur $\hat{\theta}$ est convergent si

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

Condition suffisante de convergence :

Si biais $(\hat{\theta}) \xrightarrow[n\to+\infty]{} 0$ et var $(\hat{\theta}) \xrightarrow[n\to+\infty]{} 0$ alors

$\hat{\theta}$ est Convergent

$\left.\begin{array}{l}\text{biais} \to 0\\ \text{var} \to 0\end{array}\right\} \Rightarrow EQM(\hat{\theta}) \to 0$

$\Rightarrow$ cvg en moyenne quadratique $\Rightarrow$ cvg en proba

En effet biais $(\hat{\theta}) \xrightarrow[n\to+\infty]{} 0$

Var $\hat{\theta} \xrightarrow[n\to+\infty]{} 0$

$\Rightarrow EQM(\hat{\theta}) \xrightarrow[n\to+\infty]{} 0 \Leftrightarrow \hat{\theta} \xrightarrow[n\to+\infty]{mq} \theta$

On sait que $\hat{\theta} \xrightarrow[n\to+\infty]{mq} \theta \Rightarrow \hat{\theta} \xrightarrow[n\to+\infty]{P} \theta$

II Exemples:

Exemple 1: Soit $(X_1, \dots, X_n)$ un échantillon tel que $E[X_i] = m$ et $\text{var } X_i = \sigma^2$. On suppose que σ^2 est connue et on désire estimer $\theta = m$.

Comparer $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$ et $\hat{\theta}_2 = \frac{2}{n(n+1)} \sum_{i=1}^{n} i X_i$

Biais: $* \; E[\hat{\theta}_1] - \theta = E[\hat{\theta}_1] - m$

$$= E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] - m$$

$$= \frac{1}{n} \sum_{i} \underbrace{E[X_i]}_{m} - m = \boxed{0}$$

$\hat{\theta}_1$ est un estimateur non biaisé de m $\quad 1+2+\dots+n$

$$* \; E[\hat{\theta}_2] - m = \frac{2}{n(n+1)} \sum_{i=1}^{n} i \, E[X_i] - m$$

$$= \frac{2}{n(n+1)} m \frac{n(n+1)}{2} - m = \boxed{0}$$

$\hat{\theta}_2$ non biaisé

Variances:

Rq: Si $X_1, \dots, X_n$ sont n VA indépendantes, alors

$$\boxed{\text{Var} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \text{Var } X_i}$$

$$* \; \text{Var } \hat{\theta}_1 = \frac{1}{n^2} \text{Var}\left(\sum X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \overbrace{\text{Var } X_i}^{\sigma^2} = \cancel{\frac{n\sigma^2}{n^2}} = \boxed{\frac{\sigma^2}{n}}$$

Rq: $\text{Var } \hat{\theta}_1 \xrightarrow[n \to +\infty]{} 0$
Biais $\hat{\theta}_1 = 0$ $\quad$ Donc $\hat{\theta}_1$ est convergent

$$* \; \text{Var } \hat{\theta}_2 = \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} \underbrace{\text{Var}(i X_i)}_{i^2 \underbrace{\text{Var}(X_i)}_{\sigma^2}} = \frac{4\sigma^2}{n^2(n+1)^2} \underbrace{\sum_{i=1}^{n} i^2}_{\frac{n(n+1)(n+2)}{6}}$$

$$\text{Var } \hat{\theta}_2 = \frac{2}{3} \sigma^2 \frac{2n+1}{n(n+1)}$$

$\text{Var } \hat{\theta}_2 \xrightarrow[n \to +\infty]{} 0$
Biais $\hat{\theta}_2 = 0$ $\quad$ Donc $\hat{\theta}_2$ est convergent

On préfèrera $\hat{\theta}_1$ à $\hat{\theta}_2$ si $Var\,\hat{\theta}_1 < Var\,\hat{\theta}_2$

si $\frac{\sigma^2}{n} < \frac{2}{3}\sigma^2 \frac{2n+1}{n(n+1)} \sim \frac{4}{3}\frac{\sigma^2}{n}$

$\boxed{\text{Ok on préfèrera } \hat{\theta}_1 \text{ à } \hat{\theta}_2}$

$\hat{\theta}_3 = \hat{m}^n = \prod_{i=1}^{n} X_i$

$E[\hat{m}^n] = \prod_{i=1}^{n} \underbrace{E[X_i]}_{m} = m^n$

$Var(\hat{m}^n) = E[(\prod_{i=1}^{n} X_i)^2] - (m^n)^2$

$= E[X_1^2 X_2^2 \dots X_n^2] - \dots$

$= \prod_{i=1}^{n} \underbrace{E[X_i^2]}_{m^2+\sigma^2} - m^{2n} = (m^2+\sigma^2)^n - m^{2n}$

i.i.d.

Exemple 2 :

mêmes hypothèses mais on désire estimer σ^2

$m = E[X_i]$

$\sigma^2 = E[(X_i - m)^2]$

- 1er Cas : m connu

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

$$E[\hat{\sigma}^2] - \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E[(X_i-m)^2]}_{n\sigma^2} - \sigma^2 = \sigma^2 - \sigma^2 = \boxed{0}$$

$\hat{\sigma}^2$ non biaisé

$$Var\,\hat{\sigma}^2 = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i - m)^2 = \frac{n\,Var(X_1 - m)^2}{n^2}$$

les X_i ont la même loi => $\hat{m}$ variance

$$Var\,\hat{\sigma}^2 = \frac{Var(X_1 - m)^2}{n} \xrightarrow[n\to+\infty]{} 0$$ donc $\hat{\sigma}^2$ converge

2ème Cas : m inconnu :

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \underbrace{\bar{X}}_{\hat{m}})^2$$

Biais ? $\hat{\sigma}^2$ Converge ?

$$\text{Biais}(\tilde{\sigma}^2) : E[\tilde{\sigma}^2] - \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] - \sigma^2$$

$$E[(X_i - \bar{X})^2] = E[(X_i - \bar{X} + m - m)^2] = E[(X_i - m + m - \bar{X})^2]$$

$$= E[(X_i - m)^2] + E[(m - \bar{X})^2] + 2E[(X_i - m)(m - \bar{X})]$$

$$\text{Biais}\,\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} E[(X_i - m)^2] + \frac{1}{n}\sum_{i=1}^{n} E[(m - \bar{X})^2] + \frac{2}{n}\sum_{i=1}^{n} E[(X_i - m)(m - \bar{X})] - \sigma^2$$

$$= \sigma^2 + 0 + \quad - \sigma^2$$

biais $\tilde{\sigma}^2 = E[\tilde{\sigma}^2] - \sigma^2$

$$E[\tilde{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n} E\left[\left\{(X_i - m) + (m - \bar{X})\right\}^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E\left[(X_i - m)^2\right]}_{\sigma^2} + \underbrace{E\left[(\underset{\uparrow E[\bar{X}]}{\bar{X} - m})^2\right]}_{Var\,\bar{X} = \frac{\sigma^2}{n}} + \underbrace{2\,E\left[(X_i - m)(m - \bar{X})\right]}_{-\frac{2\sigma^2}{n}}$$

$$E\Big[(X_i - m)\big(\underbrace{m - \frac{1}{n}\sum_{j=1}^{n} X_j}_{\bar{X}}\big)\Big] = E\left[(X_i - m)\,\frac{1}{n}\sum_{j=1}^{n}(m - X_j)\right]$$

$$= \frac{1}{n}\sum_{j=1}^{n} - \underbrace{E\left[(X_i - m)(X_j - m)\right]}$$

$i \neq j$ $cov(X_i, X_j) = 0$ car X_i et X_j ind

$i = j$ $Var\,X_i = \sigma^2$

$$= -\frac{\sigma^2}{n}$$

$$E[\tilde{\sigma}^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

$\boxed{\tilde{\sigma}^2 \text{ est biaisé}}$

donc $E\left[\frac{n}{n-1}\tilde{\sigma}^2\right] = \sigma^2$

On pose $\boxed{\sigma^{2*} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ non biaisé

faudrait montrer que cvg

13/12/99

Bon estimateur :
- Biais proche de 0 (moyenne de l'estimat° proche de la valeur moyenne
- var faible
- Convergence

III Inégalité de RAO-CRAMER - ESTIMATEUR EFFICACE

1) Cas d'un paramètre réel $\theta \in \mathbb{R}$:

On peut tjs éliminer le biais en général (ex : balance)

La var est bornée ...

Définition : On appelle vraisemblance de $(X_1, \dots, X_n)$ la fonction $L(x_1, \dots, x_n ; \theta)$ définie par :

* Si X_i VA continues $L(x_1 \dots x_n ; \theta) = f(x_1, \dots, x_n ; \theta)$ densité de probabilité de $(X_1, \dots, X_n)$
* Si X_i VA discrètes $L(x_1, \dots, x_n ; \theta) = P[X_1 = x_1, \dots, X_n = x_n ; \theta]$

Inégalité de RAO-CRAMER :

$$\boxed{\mathrm{Var}\,\hat{\theta} \geqslant \frac{[1 + b'(\theta)]^2}{-E\left[\frac{\partial^2 \ln L(x_1, \dots, x_n ; \theta)}{\partial \theta^2}\right]}} = \mathrm{BRC}(\theta)$$

en général non biaisé $\Rightarrow$ $b'(\theta) = 0$ $\Rightarrow$ 1 au numérateur

$b'(\theta)$: dérivée du biais

* $\mathrm{BRC}(\theta)$ est appelée borne de Cramer-Rao de θ

Cette inégalité suppose que $L(x_1, \dots, x_n ; \theta)$ est 2 fois dérivable par rapport à θ, et que les bornes du support de L sont indépendantes de θ.

Par exemple, cette inégalité n'est pas valable lorsque $X_i \sim U(0, \theta)$ et qu'on cherche à estimer θ.

Définition : Un estimateur non biaisé de θ noté $\hat{\theta}$ vérifiant $\mathrm{Var}\,\hat{\theta} =$ Borne de RAO-CRAMER est unique et appelé l'estimateur efficace de θ.
(c'est le meilleur)

Exemples

ex 1 : $X_i \sim N(m, \sigma^2)$, X_i indépendantes, σ^2 connu, on cherche à estimer m.

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

ind

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

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

$$\frac{\partial \ln L}{\partial m} = \cancel{-}\frac{1}{\cancel{2}\sigma^2} \sum_{i=1}^{n} \cancel{2}(x_i - m)\cancel{(-1)} = \frac{1}{\sigma^2} \Sigma (x_i - m)$$

$$\frac{\partial^2 \ln L}{\partial m^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (-1) = -\frac{n}{\sigma^2}$$

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

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

Si $\hat{m}$ est un estimateur non biaisé de m alors $\boxed{\operatorname{Var} \hat{m} \geq \frac{\sigma^2}{n}}$

Rq 1: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ on a vu $E[\bar{X}] = m$

$\operatorname{Var} \bar{X} = \frac{\sigma^2}{n}$

$\bar{X}$ est l'estimateur efficace de m dans le cas de VA indépendantes de loi $N(m, \sigma^2)$

Rq 2: la variance optimale concernant l'estimation de m est $\frac{\sigma^2}{n} \propto \frac{1}{n}$

→ précis° de l'estimation optimale

Rq 3: lorsque les X_i sont indépendantes

$$L(x_1 \dots x_n; \theta) = \prod_{i=1}^{n} L(x_i; \theta)$$

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

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

$$E\left[\frac{\partial^2 \ln L(X_1 \dots X_n; \theta)}{\partial \theta^2}\right] = \sum_{i=1}^{n} E\left[\frac{\partial^2 \ln L(X_i; \theta)}{\partial \theta^2}\right]$$

Si les VA X_i ont la même loi, toutes les espérances sont égales donc

$$= n \cdot E\left[\frac{\partial^2 \ln L(X_1; \theta)}{\partial \theta^2}\right]$$

$$\boxed{\operatorname{Var} \hat{\theta} \geq \frac{(1 + b'(\theta))^2}{n\, E\left[\frac{\partial^2 \ln L(X_1; \theta)}{\partial \theta^2}\right]}}$$

Ex 2 : $X_i \sim \mathcal{P}(\lambda)$ chercher la borne de Rao-Cramer sur λ pour un estimateur non biaisé.

2) Cas où $\theta = (\theta_1 \dots \theta_p) \in \mathbb{R}^p$

La matrice de covariance de $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$ (définie par

$\mathrm{Cov}(\hat{\theta}) = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right) {}^t\!\left(\hat{\theta} - E[\hat{\theta}]\right) \right]$) vérifie $\rightarrow$ matrice $x({}^t x)$ $\quad$ ${}^t x . x \rightarrow a \in \mathbb{R}$

$$\mathrm{Cov}(\hat{\theta}) \geqslant I_\theta^{-1}$$ pour un estimateur non biaisé de θ

avec $I_\theta = \left(I_{ij}\right)_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$ et $I_{ij} = E\left[\dfrac{-\partial^2 \ln L(X_1 \dots X_n ; \theta)}{\partial \theta_i \partial \theta_j} \right]$

$A \geqslant B$ signifie $A - B$ matrice semi définie positive

$\left({}^t x (A-B) x \geqslant 0 \quad \forall x \right)$ $\quad$ Semi définie positive

En particulier $\boxed{\mathrm{Var}\, \hat{\theta}_i \geqslant \left(I_\theta^{-1}\right)_{ii}}$

$\boxed{\hat{\theta}_{mo} \text{ est simple à implanter mais possède peu de bonnes propriétés...}}$

IV ~~V~~ Méthode du maximum de Vraisemblance

V.1. Définition :

L'estimateur du maximum de vraisemblance de θ est défini par

$$\boxed{\hat{\theta}_{MV} = \arg\max_{\theta} \underbrace{L(x_1, \ldots, x_n ; \theta)}_{\text{fct° de } \theta}}$$

$L(x_1 \ldots x_n ; \theta)$ étant la vraisemblance de θ $\left(f(x_1 \ldots x_n ; \theta) \text{ ou } P[X_1 = x_1 \ldots X_n = x_n ; \theta] \right)$

Dans la mesure où les hypothèses de la borne de Rao Cramer sont vérifiées (ie $L(x_1 \ldots x_n ; \theta)$ 2 fois dérivable et bornes du support de $L(x_1 \ldots x_n ; \theta)$ indépendantes de θ). la détermination de $\hat{\theta}_{MV}$ se fait à l'aide des équations suivantes :

$\theta \in \mathbb{R}$ $\qquad \dfrac{\partial L(x_1 \ldots x_n ; \theta)}{\partial \theta} = 0$ ou $\dfrac{\partial \ln L(x_1 \ldots x_n ; \theta)}{\partial \theta} = 0$

On vérifie que la solution donne un maximum en étudiant

$\dfrac{\partial L(x_1 \ldots x_n ; \theta)}{\partial \theta} \geq 0$ ou en cas de problème

$\dfrac{\partial^2 L(x_1 \ldots x_n ; \hat{\theta}_{MV})}{\partial \theta^2} < 0$ (Condition assurant que $\hat{\theta}_{MV}$ est un maximum local)

θ		$\hat{\theta}_{MV}$	
$\dfrac{\partial^2 L}{\partial \theta^2}$		< 0	
$\dfrac{\partial L}{\partial \theta}$	$+$	0	$-$
L	$\nearrow$		$\searrow$

ex 3: $X_i \sim P(\lambda)$ $\quad \lambda$?

$$L(x_1 \dots x_n ; \lambda) = \prod_{i=1}^{n} P[X_i = x_i ; \lambda] = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= \frac{\lambda^{\Sigma x_i}}{\Pi x_i!} e^{-n\lambda}$$

$$\ln L(\,) = \Sigma x_i \ln \lambda - n\lambda - \ln(\Pi x_i!)$$

$$\frac{\partial \ln L(\,)}{\partial \lambda} \geq 0 \iff \frac{\Sigma x_i}{\lambda} - n \geq 0 \iff \lambda \leq \bar{x}$$

λ	0	$\bar{x}$	$+\infty$
$\frac{\partial \ln L}{\partial \lambda}$	$+$	0	$-$
L	$\nearrow$		$\searrow$

$\lambda = \bar{x}$ est le maximum global unique de $L(x_i ; \lambda)$

on pose $\boxed{\hat{\lambda}_{MV} = \bar{X}}$

04/01/2000

$$f(x_1 \dots x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

$\boxed{\hat{\theta}_{MV}}$ vérifie $f(x_1 \dots x_n ; \hat{\theta}_{MV}) \geq f(x_1 \dots x_n ; \theta) \quad \forall \theta$

$\theta \in \mathbb{R}^p$ On résoud $\frac{\partial L(x_1 \dots x_n ; \theta)}{\partial \theta_i} = 0 \quad \forall i = 1, \dots, p$

ou $\frac{\partial \ln L(x_1 \dots x_n ; \theta)}{\partial \theta_i} = 0 \quad \forall i = 1, \dots, p$

V - 2 - Propriétés :

* $\widehat{\theta}_{MV}$ est asymptotiquement sans biais $\lim_{n \to +\infty} E[\widehat{\theta}_{MV}] - \theta = 0$
* $\widehat{\theta}_{MV}$ est convergent
* $\widehat{\theta}_{MV}$ est asymptotiquement efficace $Var\, \widehat{\theta}_{MV} \underset{n \to +\infty}{\sim} BRC(\theta)$ Borne de Rao Cramer

(ie $\lim_{n \to +\infty} \frac{Var\, \widehat{\theta}_{MV}}{BRC(\theta)} = 1$)

* Invariance fonctionnelle

Soit $\mu = h(\theta)$, où h est une fct° bijective d'un ouvert $O \subset \mathbb{R}^p$ dans un ouvert $V \subset \mathbb{R}^p$ alors

$$\widehat{\mu}_{MV} = h(\widehat{\theta}_{MV})$$

* $\sqrt{I_n(\theta)}\, (\widehat{\theta}_{MV} - \theta) \xrightarrow[n \to +\infty]{L} N(0, I_p)$

où $I_n(\theta) = E\left[\frac{-\partial^2 \ln(X_1 \dots X_n ; \theta)}{\partial \theta^2} \right]$

est l'information de FISHER (c'est le dénominateur de la borne de Rao Cramer).

$\boxed{\widehat{\theta}_{MV} \text{ possède plein de bonnes propriétés mais peut être difficile à implanter}}$

V - 3 - Exemples :

ex 1 : $X_i \sim N(m, \sigma^2)$ σ^2 connu m inconnu

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

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

$$\frac{\partial \ln L}{\partial m} \geq 0 \iff + \frac{1}{2\sigma^2} \sum_i 2(x_i - m)(-1) \geq 0$$

$$\iff \sum x_i \geq nm \iff \boxed{m \leq \frac{1}{n} \sum x_i = \bar{x}}$$

m	$-\infty$		$\bar{x}$		$+\infty$
$\frac{\partial \ln L}{\partial m}$		$+$	0	$-$	
$\ln L$		$\nearrow$		$\searrow$	

$\bar{x}$ est un maximum global unique de $L(x_1 \dots x_n ; m)$

$$\boxed{\widehat{m}_{nv} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}}$$

Ex 2 : m inconnu , σ^2 inconnu

$$\frac{\partial \ln L}{\partial m} = 0 \quad \text{et} \quad \frac{\partial \ln L}{\partial \sigma^2} = 0$$

$$m = \bar{x} \quad \text{et} \quad -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - m)^2 = 0$$

$$m = \bar{x} \quad \text{et} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2$$

j'admets que $\left(m = \bar{x} , \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2\right)$ est l'argument du maximum de $L(x_1 \dots x_n ; \theta)$

On pose
$$\boxed{\begin{aligned} \widehat{m}_{nv} &= \overline{X} \\ \widehat{\sigma}^2_{nv} &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \end{aligned}}$$

16/12/99

~~IV~~ V Méthode des moments :

IV.1. Définition

Soient n VA $x_1, \dots, x_n$ iid dont la loi dépend d'un paramètre $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$. Le ~~paramètre~~ principe de l'estimation de θ consiste à donner une valeur approchée de θ à l'aide de $x_1 \dots x_n$.

En général θ s'exprime en fonction de $m_1, \dots, m_p$ avec $m_i = E[X_j^i]$ (indépendant de j car toutes les VA X_j ont la même loi), de la façon suivante

$$\theta = h(m_1, \dots, m_p)$$

L'estimateur des moments de θ est défini par

$$\hat{\theta}_{ro} = h(\hat{m}_1, \dots, \hat{m}_p) \quad \text{avec } \hat{m}_i = \frac{1}{n} \sum_{j=1}^{n} X_j^i$$

IV.2. Exemple :

$X_j \sim N(m, \sigma^2)$ $\quad X_j$ iid $\quad \theta = (m, \sigma^2)$

$$m_1 = E[X_j] = m$$

$$m_2 = E[X_j^2] = \operatorname{Var} X_j + E[X_j]^2 = \sigma^2 + m^2$$

Rq : $m = m_1$
$\sigma^2 = m_2 - m_1^2$

On pose

$$\boxed{\begin{aligned} \hat{m}_{ro} &= \frac{1}{n} \sum_{j=1}^{n} X_j \\ \hat{\sigma}^2_{ro} &= \frac{1}{n} \sum_{j=1}^{n} X_j^2 - \left(\frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \end{aligned}}$$

IV.3. Propriétés :

* $\hat{\theta}_{ro}$ est convergent (vers θ)
* $\sqrt{n}\,(\hat{\theta}_{ro} - \theta) \xrightarrow[n \to +\infty]{L} N(0, \Sigma)$ $\quad$ Σ matrice $p \times p$ indépendante de n

$\hat{\theta}_{ro}$ peut être biaisé, non biaisé, efficace ou non. $\Rightarrow$ pas terrible niveau propriétés

$\hat{\theta}_{ro}$ simple à implanter mais il possède peu de bonnes propriétés.

04/01/2000

VI Estimation BAYESIENNE

L'estimat° Bayésienne consiste à supposer que le paramètre inconnu θ est la réalisation d'une va Θ qui possède une appelée LOI A PRIORI (densité de probabilité à priori dans le cas continu).

le but de l'estimation Bayésienne consiste à estimer θ (la réalisation de la va Θ) à l'aide d'un estimateur noté $\hat{\theta}(X)$ de façon à minimiser une fonction (appelée fonction de coût) : $E[c(\theta, \hat{\theta}(x))]$ qui représente l'erreur entre θ et $\hat{\theta}(x)$

<u>Remarque</u>: L'estimateur $\hat{\theta}$ sera construit à partir de la loi des va X_i <u>mais aussi à l'aide de la loi à priori</u>

VI.1- ESTIMATEUR de la moyenne à Posteriori ($\hat{\theta}_{MOAP}(X)$)

$\hat{\theta}_{MOAP}(X)$ est l'estimateur qui minimise $E\left[(\theta - \hat{\theta}(x))^2\right]$ (coût quadratique). Il est défini par:

$$\hat{\theta}_{MOAP}(x) = E[\theta / X]$$

moyenne de la loi de θ/X appelée loi à POSTERIORI

<u>Preuve</u>: On cherche à minimiser $E\left[(\theta - \hat{\theta}(x))^2\right]$ ie On cherche $\hat{\theta}(X)$ qui minimise $\underset{X}{E}\left[\underset{\theta}{E}\left[(\theta - \hat{\theta}(x))^2 \,|\, x\right]\right]$

$$E\left[(\theta - \hat{\theta}(x))^2 / x\right] = E\left[\theta^2/x\right] - \underbrace{2\,E\left[\theta\,\hat{\theta}(x) / x\right]}_{-2\hat{\theta}(x)\,E[\theta/x]} + \underbrace{E\left[\hat{\theta}^2(x) \,|\, x\right]}_{\hat{\theta}^2(x)}$$

$$= E\left[\theta^2/x\right] - E[\theta/x]^2 + E[\theta/x]^2 - 2\,\hat{\theta}(x)\,E[\theta/x] + \hat{\theta}^2(x)$$

$$= Var(\theta/x) + \left(\hat{\theta}(x) - E[\theta/x]\right)^2$$

~~$\hat{\theta}$ simple à implanter mais possède peu de bonnes propriétés~~

à X fixé, on a $E\left[\left(\theta - \hat{\theta}(X)\right)^2 / X\right] \geqslant Var(\theta/X)$

il y a égalité pour $\hat{\theta}(X) = E[\theta/X]$

donc $E[\theta/X]$ minimise $E\left[\left(\theta - \hat{\theta}(X)\right)^2 / X\right]$

Limitons nous au cas continu pour simplifier ($X_1 \dots X_n$ va de densités $f(x_i ; \theta)$)

$$E\left[\left(\theta - \hat{\theta}(X)\right)^2\right] = \int_{\mathbb{R}^n} E\left[\left(\theta - \hat{\theta}(X)\right)^2 / X\right] f(x_1 \dots x_n ; \theta)\, dx_1 \dots dx_n$$

puisque $\hat{\theta}(X) = E[\theta/X]$ minimise $\forall x$, $\hat{\theta}(x) = E[\theta/X]$ minimise aussi $E\left[\left(\theta - \hat{\theta}(x)\right)^2\right]$ $\boxed{\text{CQFD}}$

ex exemple (Ex 3 TD3) :

- Résistance inconnue θ
- On a N mesures $x_i = \theta + \underbrace{b_i}_{\text{erreur}}$ avec $b_i \sim \mathcal{N}(0, \sigma_b^2)$

Il est clair que $\boxed{x_i \sim \mathcal{N}(\theta, \sigma_b^2)}$

La loi $N(\theta, \sigma_b^2)$ est la loi de x_i / θ

- On suppose qu'on dispose d'une loi a priori sur θ qui est $N(m, \sigma^2)$

ie $p(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(\theta - m)^2}{2\sigma^2}$

$\hat{\theta}_{MAP} = E[\theta / X]$ $X = (X_1 \dots X_n)$

$\boxed{\text{Calcul de } E[\theta/X]}$ 1) loi de θ/X

Règle de BAYES $p(\theta / \underbrace{x_1 \dots x_n}_{X}) = \frac{P(x_1 \dots x_n / \theta)\, P(\theta)}{P(x_1 \dots x_n)}$

$p(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(\theta - m)^2}{2\sigma^2}$ connu

$$p(x_1 \ldots x_n / \theta) \underset{X_i \text{ ind}}{=} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_b^2}} \exp - \frac{(x_i - \theta)^2}{2 \sigma_b^2}$$

$$p(x_1 \ldots x_n) = \int p(x_1 \ldots x_n, \theta)\, d\theta = \int \underbrace{p(x_1 \ldots x_n / \theta)}_{\text{connu}} \underbrace{p(\theta)}_{\text{connu}}\, d\theta$$

$p(x_1 \ldots x_n)$ est évidemment indépendant de θ. Souvent, il n'est pas nécessaire de le calculer.

$$p(x_1 \ldots x_n / \theta)\, p(\theta) = C \exp - \frac{1}{2}\left[\frac{(\theta - m)^2}{\sigma^2} + \frac{1}{\sigma_b^2} \sum_{i=1}^{n} (x_i - \theta)^2 \right]$$

donc $p(\theta / x_1 \ldots x_n) \underset{\text{proportionnel}}{\propto} \exp - \frac{1}{2}\left[\theta^2 \underbrace{\left(\frac{1}{\sigma^2} + \frac{n}{\sigma_b^2} \right)}_{\frac{1}{\sigma_p^2}} - 2\theta \underbrace{\left(\frac{m}{\sigma^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma_b^2} \right)}_{\frac{m_p}{\sigma_p^2}} \right.$

$$\exp - \frac{1}{2} \frac{(\theta - m_p)^2}{\sigma_p^2} = \exp - \frac{1}{2\sigma_p^2} \left(\theta^2 - 2\theta m_p + m_p^2 \right)$$

$$p(\theta / x_1 \ldots x_n) \propto \exp - \frac{1}{2} \frac{(\theta - m_p)^2}{\sigma_p^2}$$

avec $\frac{1}{\sigma_p^2} = \frac{1}{\sigma^2} + \frac{n}{\sigma_b^2}$

$$m_p = \left(\frac{m}{\sigma^2} + \frac{n \bar{x}}{\sigma_b^2} \right) \sigma_p^2 = \frac{\frac{m}{\sigma^2} + \frac{n \bar{x}}{\sigma_b^2}}{\frac{1}{\sigma^2} + \frac{n}{\sigma_b^2}}$$

ie $\boxed{m_p = \bar{x} \left(\frac{\sigma^2}{\sigma^2 + \frac{\sigma_b^2}{n}} \right) + m \left(\frac{\sigma_b^2 / n}{\sigma_b^2 / n + \sigma^2} \right)}$

$$\theta / X \sim \mathcal{N}(m_p, \sigma_p^2)$$

valeur nominale

$$\boxed{\hat{\theta}_{MAP} = E[\theta / X] = m_p = \bar{X} \left(\frac{\sigma^2}{\sigma^2 + \frac{\sigma_b^2}{n}} \right) + m \left(\frac{\sigma_b^2 / n}{\sigma_b^2 / n + \sigma^2} \right)}$$

précision valeur R — précis appareil de mesure

<u>quand n gd</u> $\quad \frac{\sigma_b^2}{n} \ll \sigma^2 \qquad \hat{\theta}_{MAP} \# \bar{X}$

On ne tient plus compte de la valeur à priori m (on ne fait pas confiance au fournisseur)

<u>quand n petit</u> $\quad \frac{\sigma_b^2}{n} \gg \sigma^2 \qquad \hat{\theta}_{MAP} \# m$

(n=0)

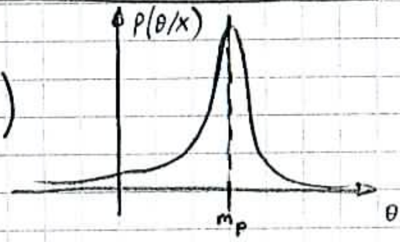
VI. 2. Estimateur du Maximum A Posteriori ($\hat{\theta}_{MAP}$)

L'estimateur $\hat{\theta}(x)$ qui minimise $E[C(\theta, \hat{\theta}(x))]$ avec

$$C(\theta, \hat{\theta}(x)) = \begin{cases} 1 & \text{si } |\theta - \hat{\theta}(x)| > \Delta \\ 0 & \text{si } |\theta - \hat{\theta}(x)| < \Delta \end{cases}$$

Δ "arbitrairement" petit est défini par

$$\hat{\theta}_{MAP} = \arg\max p(\theta / x) \quad \text{ie} \quad p(\hat{\theta}_{MAP} / x) \geqslant p(\theta / x) \quad \forall \theta$$



Pour l'exemple précédent $\theta / x \sim N(m_p, \sigma_p^2)$

donc $\boxed{\hat{\theta}_{MAP} = \hat{\theta}_{MOAP}}$

VII Estimation par intervalle de confiance :

1) **Principe :** Dans certaines applications, il est plus réaliste de donner une information de la forme $a < \theta < b$ plutôt que de dire $\hat{\theta} = c$.
Un intervalle de confiance pour l'estimation du paramètre θ est un intervalle de cette forme $[d_1, d_2]$ tel que

$$P[d_1 < \theta < d_2] = \alpha$$

En général $\alpha = 0,95$; $0,99$ ou $0,90$ et s'appelle le paramètre de confiance.

2) **Détermination pratique de l'intervalle**

On cherche à l'aide d'un estimateur $\hat{\theta}$ de θ (obtenu grâce à la méthode des moments ou du maximum de vraisemblance) une statistique $T(X_1 \dots X_n)$ dont la loi est connue et qui dépend de θ.
On exprime la probabilité $P[a < T(X_1 \dots X_n) < b] = \alpha$ sous la forme

$$P[d_1 < \theta < d_2] = \alpha$$

3) Exemples

Ex 1 : $X_i \sim \mathcal{N}(m, \sigma^2)$ σ^2 connue, m inconnue

On sait que $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ est un bon estimateur de m
(pour loi $\mathcal{N}$, $\hat{\theta}$ sans biais, convergent, efficace)

et que $\hat{\theta} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$ (dans la mesure où les VA sont indépendantes)
(en effet $E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum \underbrace{E[X_i]}_{m} = m$ et

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum X_i\right) \underset{X_i \text{ ind}}{=} \frac{\sum \overbrace{Var X_i}^{\sigma^2}}{n^2} = \frac{\sigma^2}{n} \quad)$$

la loi est connue et dépend de θ

On pose $T(X_1 \dots X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$ ou

$$\boxed{T(X_1 \dots X_n) = \frac{\frac{1}{n} \sum_{i=1}^{n} X_i - m}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)} \to \text{Tabulé}$$

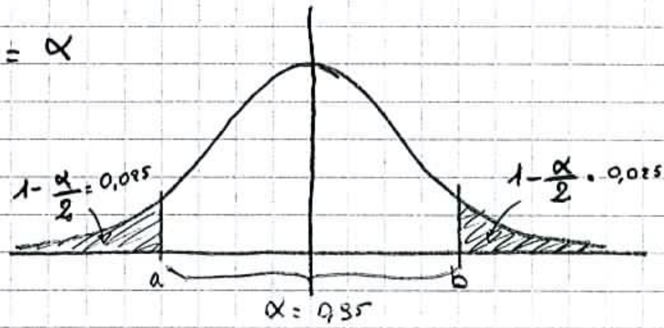
<u>Rq:</u> $T(X_1 \dots X_n)$ <u>n'est pas unique</u>

<u>Intervalle de confiance pour m:</u>

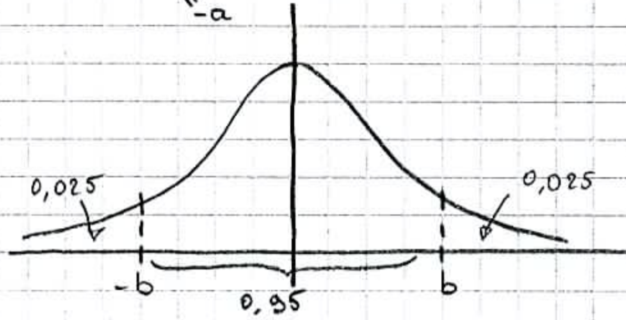
il se détermine grâce à

$$P[\, a < T(X_1 \dots X_n) < b \,] = \alpha$$

<u>Prenons $\alpha = 0{,}95 = 95\%$</u>

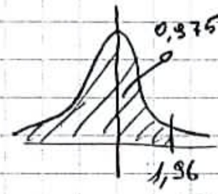


On détermine a et b à l'aide des tables
$b = -a$



$\boxed{b = 1{,}96}$

$\boxed{a = -1{,}96}$



Donc $P\left[-1{,}96 < \frac{\frac{1}{n}\sum X_i - m}{\sigma/\sqrt{n}} < 1{,}96\right] = 95\%$

$$-1{,}96 < \frac{\overline{X} - m}{\sigma/\sqrt{n}} < 1{,}96 \iff -1{,}96 \frac{\sigma}{\sqrt{n}} < \overline{X} - m < 1{,}96 \frac{\sigma}{\sqrt{n}}$$

$$\iff \overline{X} - 1{,}96 \frac{\sigma}{\sqrt{n}} < m < \overline{X} + 1{,}96 \frac{\sigma}{\sqrt{n}}$$

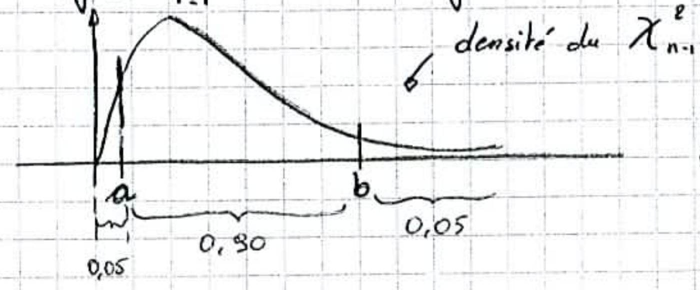
$\left[\overline{X} - 1{,}96 \frac{\sigma}{\sqrt{n}} \,,\, \overline{X} + 1{,}96 \frac{\sigma}{\sqrt{n}}\right]$ est l'intervalle de confiance pour m avec $\alpha = 0{,}95$

Ex 2 : $X_i \sim N(m, \sigma^2)$ $\quad$ m inconnue

On cherche un intervalle de confiance pour σ^2

On sait que $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ est un estimateur non biaisé et convergent de σ^2.

On montre que $T(X_1 \dots X_n) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n-1}{\sigma^2} \hat{\sigma}^2$ suit une loi du χ^2_{n-1}



$\boxed{\text{Prenons } \alpha = 0,90}$

les tables donnent $\begin{cases} a = \_\_\_\_\_\_ \\ b = \_\_\_\_\_\_ \end{cases}$

$$a < T(X_1 \dots X_n) < b \iff a < \frac{1}{\sigma^2} \sum (X_i - \overline{X})^2 < b$$

$$\iff \boxed{\frac{\sum (X_i - \overline{X})^2}{b} < \sigma^2 < \frac{\sum (X_i - \overline{X})^2}{a}}$$

Ex 3 : $X_i \sim N(m, \sigma^2)$ $\quad$ Intervalle de confiance pour m, σ^2 étant inconnue

On sait que $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ est un bon estimateur de m

et que $\overline{X} \sim N(m, \frac{\sigma^2}{n})$

On ne peut pas prendre $T(X_1 \dots X_n) = \overline{X}$ car $\overline{X}$ dépend non seulement de m mais aussi de σ^2.

$$U = \frac{\overline{X} - m}{\sigma / \sqrt{n}} \sim N(0,1)$$

On sait que $V = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \sim \chi^2_{n-1}$

Donc $\dfrac{U}{\sqrt{\frac{V}{n-1}}} \sim \underbrace{t_{n-1}}_{\text{loi de Student ou du t, à n-1 d° de libertés}}$

On pose $T(X_1 \dots X_n) = \dfrac{\dfrac{\overline{X}-m}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{1}{\sigma^2}\Sigma(X_i-\overline{X})^2}\dfrac{1}{\sqrt{n-1}}}$

$$\boxed{= \frac{\overline{X}-m}{\sqrt{\Sigma(X_i-\overline{X})^2}}\sqrt{n(n-1)} \sim t_{n-1}}$$

VIII Méthode des moindres carrés : Exemple de la régression linéaire

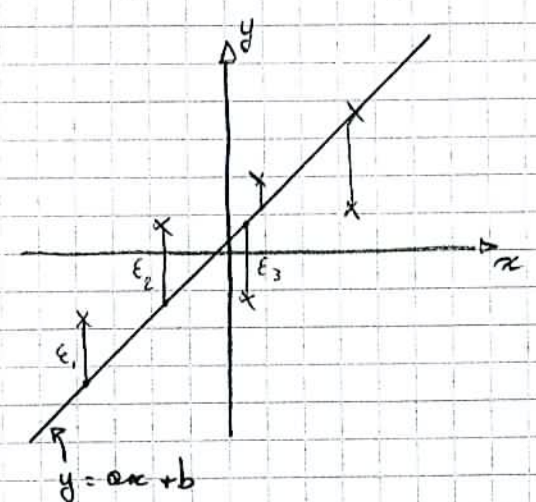
1) Principe

On a un nuage de points (x_i, y_i) et on cherche à estimer / déterminer une relation fonctionnelle entre y_i et x_i.

Par exemple, on cherche la "meilleure" droite qui passe par ces points (x_i, y_i)

2) Méthode :

On modélise les couples (x_i, y_i) de la façon suivante : (cas de la meilleure droite)



$$Y_i = a x_i + b + \varepsilon_i$$

y_i observée ; abscisses observées non aléatoires ; erreur

Y_i va associée à y_i

On détermine alors a et b de façon à minimiser

$$g(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - a x_i - b)^2$$

(cf cours d'optimisation, $g(a,b)$ admet un minimum global unique

qui vérifie : $\frac{\partial g(a,b)}{\partial a} = 0 \iff \sum_{i=1}^{n} 2(Y_i - a x_i - b)(-x_i) = 0$ (1)

$\frac{\partial g(a,b)}{\partial b} = 0 \iff \sum_{i=1}^{n} 2(Y_i - a x_i - b)(-1) = 0$ (2)

(2) $n\overline{Y} - an\bar{x} - nb = 0 \implies \boxed{b = \overline{Y} - a\bar{x}}$

avec $\overline{Y} = \frac{1}{n}\sum Y_i$ et $\bar{x} = \frac{1}{n}\sum x_i$

(1) $\sum_i Y_i x_i - a\sum_i x_i^2 - nb\bar{x} = 0$

$\frac{1}{n}\sum_i Y_i x_i - \frac{a}{n}\sum_i x_i^2 - (\overline{Y} - a\bar{x})\bar{x} = 0$

d'où $\boxed{a = \dfrac{\frac{1}{n}\sum_{i=1}^{n} Y_i x_i - \overline{Y}\bar{x}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}}$

Si $\sigma_X \neq 0$: $a = \frac{E(XY)-E(X)E(Y)}{\sigma_X^2} = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}$ et $b = E(Y) - aE(X) = E(Y) - \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}E(X)$.

Nous trouvons alors l'approximation affine :

$$\hat{Y} = E(Y) + \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}[X - E(X)].$$

La droite $y = ax+b$ porte le nom de *droite de régression de Y en X*. Son équation peut s'écrire sous la forme :

$$\frac{y-m_Y}{\sigma_Y} = \rho_{X,Y}\frac{x-m_X}{\sigma_X} \qquad \text{[III.21a]}$$

Cette droite passe par le point [E(X), E(Y)] appelé *centre de gravité de la probabilité unité* sur $\mathbb{R}^2$.

L'écart quadratique moyen minimum obtenu entre Y et son approximation affine $\hat{Y}$ est alors de :

$$E[Y-E(Y)-\rho_{X,Y}\frac{\sigma_Y}{\sigma_X}[X-E(X)]]^2 = E[Y-E(Y)]^2+\rho_{X,Y}^2\frac{\sigma_Y^2}{\sigma_X^2}E[X-E(X)]^2-2\rho_{X,Y}\frac{\sigma_Y}{\sigma_X}E[Y-E(Y)][X-E(X)]]$$

$$= \sigma_Y^2 + \sigma_Y^2\rho_{X,Y}^2 - 2\sigma_Y^2\rho_{X,Y}^2 = \sigma_Y^2[1-\rho_{X,Y}^2].$$

Cet écart est d'autant plus grand que la variance de Y est grande (grande dispersion des valeurs que peut prendre Y) et que la corrélation linéaire de X et de Y est petite.

Si $\sigma_X = 0$: nous avons $X - m_X = 0$ presque sûrement. La minimisation de l'expression [III.20] est indéterminée et tout revient en se fixant $a = 0$, à chercher un nombre b tel que $E[Y-b]^2$ soit minimum. D'après ce que nous avons dit sur la variance de Y, on trouve $b = m_Y$.

Nous constatons que si $\rho_{X,Y} = 0$, la droite de régression de Y en X devient $y-m_Y=0$. Comme dans le cas précédent, la meilleure approximation de Y au sens des moindres carrés, sous forme d'une fonction affine de X est $Y = m_Y$.

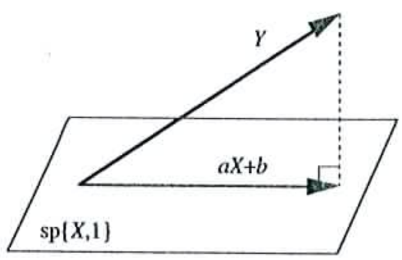
Remarque : Ces résultats peuvent s'obtenir plus simplement de manière géométrique en considérant l'espace de Hilbert $L^2(\Omega, \mathcal{Q}, P)$ des variables aléatoires du second ordre définies sur Ω :

D'après les propriétés des espaces de Hilbert, trouver l'approximation affine $aX+b$ de Y qui minimise $E[Y-aX-b]^2$ est équivalent à projeter orthogonalement le vecteur Y sur l'espace vectoriel noté sp$\{X,1\}$ engendré par les vecteurs X et 1 (variable aléatoire $1_\Omega(\omega)$) :

$$Y-aX-b \perp X \quad \Leftrightarrow \quad E[(Y-aX-b)X] = 0$$

$$Y-aX-b \perp 1 \quad \Leftrightarrow \quad E[(Y-aX-b)1] = 0$$

et l'on retrouve les équations précédentes.



On pourrait chercher également une approximation de X comme fonction affine de Y. Un raisonnement équivalent donnerait une *droite de régression de X en Y* qui aurait pour équation si $\sigma_Y \neq 0$: $x = E(X) + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}[y - E(Y)]$. Soit :

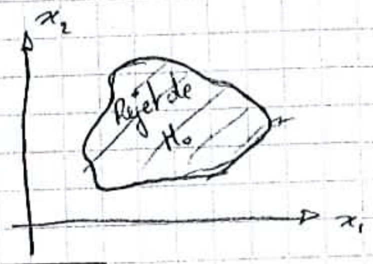
Chap 2: **Tests statistiques**

I Généralités:

Un test statistique est un mécanisme qui permet de décider à partir d'un échantillon $X = (X_1 \dots X_n)$ entre deux (ou plus) hypothèses notées H_0 et H_1. (ou $H_0, H_1, H_2 \dots$)

H_0 s'appelle <u>l'hypothèse nulle</u>. Tandis que H_1 s'appelle <u>l'hypothèse alternative</u>.

Rq: On se limite dans ce cours à $\frac{2}{3}$ hypothèses



Elaborer une stratégie de Test, c'est déterminer une statistique notée $T(X_1, \dots, X_n)$ telle que:

si $T(x_1 \dots x_n) \in \Delta$ alors on rejette l'hypothèse H_0 (c'est à dire on accepte H_1)

Si $T(x_1 \dots x_n) \notin \Delta$ alors on accepte l'hypothèse H_0

<u>Vocabulaire</u>: $\{(x_1 \dots x_n) / \underbrace{T(x_1 \dots x_n) \in \Delta}_{\text{on rejette } H_0}\}$ s'appelle <u>la région critique du test</u>

On distinguera dans ce cours <u>les tests dits paramétiques</u> des <u>tests non paramétiques</u>. <u>Un test est paramétique</u> si la forme de la loi des VA X_i est connue et que l'on cherche à tester la valeur d'un ou de plusieurs de ses paramètres.

<u>Exemple</u>: $X_i \sim N(m, \sigma^2)$ σ^2 connue

<u>Test n°1</u> H_0 $m = 1$ H_1 $m = 2$

On parle d'hypothèses simples lorsqu'elles sont réduites à un point. (1 seule valeur du paramètre)

le test n°1 est un test d'hypothèses simples.

Test n°2 : (H₀) $m > 0$ (H₁) $m \leq 0$

Ce test est un test paramétique à hypothèses appelées hypothèses Composées.

Test n°3 : (H₀) $X \sim N(m, \sigma^2)$ (H₁) non H₀

Ce test est non paramétique car il porte sur la loi des va X_i et non pas sur ses paramètres.

Rq : Pour le test n°1 on pourrait envisager la stratégie suivante

Rejet de H₀ si $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i > 1,5$

La région critique est séparée par un hyperplan (droite pour $n = 2$, plan pour $n = 3$.

Pour déterminer la performance d'un test, on s'intéressera aux risques de se tromper :

① $\boxed{\text{risque de } 1^{\text{ère}} \text{ espèce, risque } \alpha}$ $\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$

② $\boxed{\text{risque de } 2^{\text{nde}} \text{ espèce, risque } \beta}$ $\beta = P[\text{Rejeter } H_1 / H_1 \text{ vraie}]$

Un test sera d'autant meilleur que α et β seront "petits"

α et β jouent ~~des~~ rôles parfaitement symétriques et $\boxed{\text{pourtant}}$:

$\boxed{\text{Ex 1 :}}$ (H₀) le patient est sain

(H₁) le patient est atteint d'une maladie X

(Rq : le patient est sain signifie absence de maladie ⇒ hypothèse nulle)

$\alpha = P$[on décide que le patient est malade / le patient est sain]
probabilité de fausse alarme (PFA)

$\beta = P$[on décide que le patient est sain / le patient est malade]
probabilité de non détection (PND)

Dans la mesure où on a le choix, on privilégiera β

Ex 2 : H_0 non H_1 (absence d'avion)
H_1 un avion bombarde l'N7 (présence d'avion)

α = PFA
β = PND

Définition : On appelle **puissance du test** $\pi = 1-\beta$
(probabilité de détection)

II Exemple :

$X_i \sim N(m, \sigma^2)$ σ^2 connue
$(X_1, \dots, X_n)$ échantillon H_0 $m = m_0$ H_1 $m = m_1 > m_0$

Etudions la stratégie suivante :

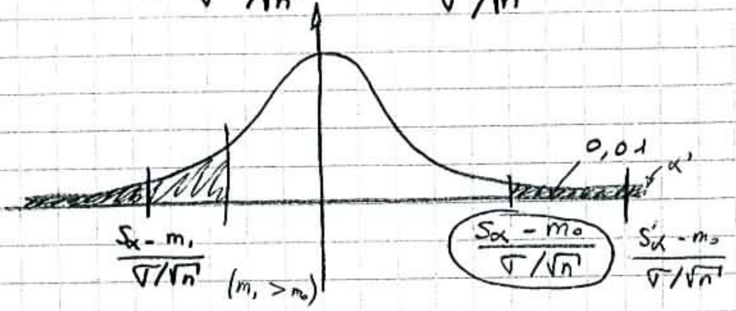
Rejet de H_0 si $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i > S_\alpha$ (seuil S_α dépend du risque α)

Détermination de S_α : On se fixe α (en général 0,1 ; 0,5 ; 0,01 ; ou 0,05,

$\alpha = 0{,}01 = P$[rejeter H_0 / H_0 vraie]
$= P[\overline{X} > S_\alpha / m = m_0]$

On sait que $\overline{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$
$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_0, \frac{\sigma^2}{n}\right)\right]$

$$\alpha = 0{,}01 = P\left[U = \frac{\bar{X} - m_0}{\sigma/\sqrt{n}} > \frac{S_\alpha - m_0}{\sigma/\sqrt{n}} \;/\; U \sim N(0,1)\right] = \int_{\frac{S_\alpha - m_0}{\sigma/\sqrt{n}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$



les tables donnent $\frac{S_\alpha - m_0}{\sigma/\sqrt{n}} = \dots$

d'où $S_\alpha = \dots$

$\boxed{\text{plus } \alpha \searrow \text{, plus } S_\alpha \nearrow}$

<u>Calcul de β</u> (ou de $\pi = 1-\beta$)

$$\beta = P[\text{rejeter } H_1 \;/\; H_1 \text{ vraie}] = P[\bar{X} < S_\alpha \;/\; m = m_1]$$

$$= P\left[V = \frac{\bar{X} - m_1}{\sigma/\sqrt{n}} < \frac{S_\alpha - m_1}{\sigma/\sqrt{n}} \;/\; V \sim N(0,1)\right]$$

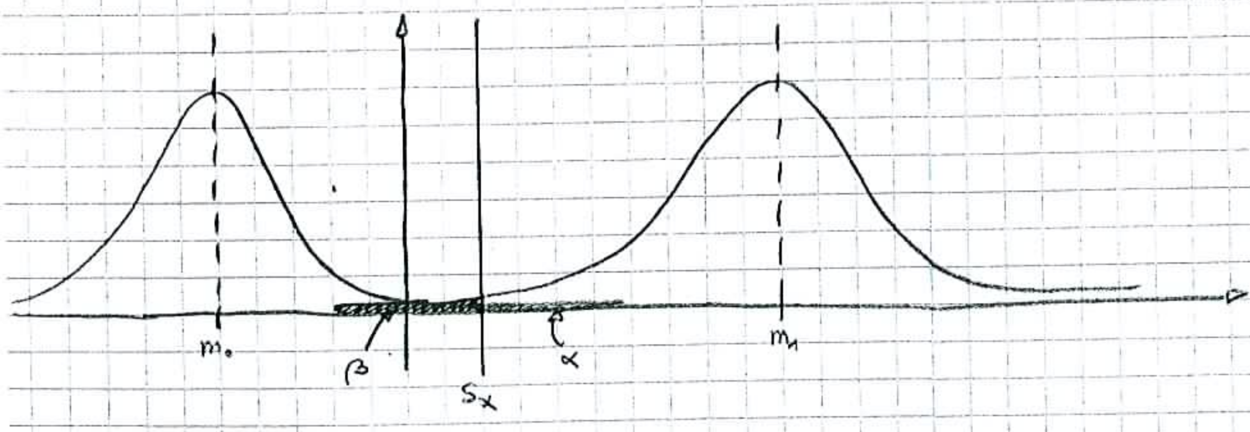
On détermine $\beta = \int_{-\infty}^{\frac{S_\alpha - m_1}{\sigma/\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$ à l'aide des tables de la loi normale.

plus $\alpha \searrow$, plus $S_\alpha \nearrow$ et donc plus $\beta \nearrow$

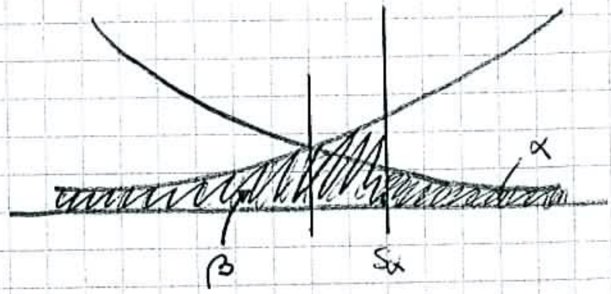
<u>α et β varient en sens inverse</u> (<u>β f$^{ct°}$ décroissante de α</u>)

<u>Autre représentation</u> : $\alpha = P\left[\bar{X} > S_\alpha \;/\; \bar{X} \sim N(m_0, \frac{\sigma^2}{n})\right]$

$\beta = P\left[\bar{X} < S_\alpha \;/\; \bar{X} \sim N(m_1, \frac{\sigma^2}{n})\right]$



Zoom



Si on cherche à minimiser $\alpha+\beta$, il faut placer S_α à l'intersection des 2 courbes.

III Théorème (ou LEMME) de NEYMAN-PEARSON (1933) (NP)

L'approche de NP consiste à fixer α et à chercher un test qui minimise β. (pour cette valeur de α fixé). On dit souvent que le test de NP est optimal.

III.1. Cas de variables aléatoires continues et d'hypothèses simples.

(H0) $\theta = \theta_0$ (H1) $\theta = \theta_1$

X_i possède une densité de probabilité sous chacune des 2 hypothèses notées

$$\underbrace{f(x_i\,;\,\theta_0)}_{\text{densité de } X_i \text{ sous } H_0} \quad \text{et} \quad \underbrace{f(x_i\,;\,\theta_1)}_{\text{densité de } X_i \text{ sous } H_1}$$

> A α fixé, le test qui minimise β est de la forme
>
> Rejet de H_0 si $\dfrac{L(x_1,\dots,x_n/H_1)}{L(x_1,\dots,x_n/H_0)} > S_\alpha$

Rq: $L(x_1,\dots,x_n/H_i) = \prod_{j=1}^{n} f(x_j,\theta_i) = f(x_1,\dots,x_n\,;\,\theta_i)$

Exemple: $X_i \sim N(m,\sigma^2)$ σ^2 connue

(H0) $m = m_0$ (H1) $m = m_1$

Le test de NP est défini par

Rejet de H_0 si
$$\frac{\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x_j-m_1)^2}{2\sigma^2}}{\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x_j-m_0)^2}{2\sigma^2}} > S_\alpha$$

si $\quad \exp -\frac{1}{2\sigma^2}\left[\sum_{j=1}^{n}(x_j-m_1)^2 - \sum_{j=1}^{n}(x_j-m_0)^2\right] > S_\alpha$

(on prend le ln)

si $\quad \sum_{j=1}^{n}\left(\cancel{x_j^2} - 2m_1 x_j + m_1^2 - \cancel{x_j^2} + 2m_0 x_j - m_0^2\right) < k\alpha$

$$\underbrace{-2(m_1-m_0)n}_{<0}\,\bar{x} + n\left(m_1^2 - m_0^2\right) < k_\alpha$$

$\boxed{\text{Si } m_1 > m_0}$ $\quad \boxed{\text{Rejet de } H_0 \text{ si } \quad \bar{x} > \mu_\alpha}$

$\boxed{\text{Si } m_1 < m_0}$ $\quad \boxed{\text{Rejet de } H_0 \text{ si } \quad \bar{x} < \mu_\alpha}$

On retrouve la stratégie du § II

<u>Rq</u>: Effectuer un test de NP c'est :

1) déterminer la zone de rejet de H_0 (zone critique)
région

2) On se fixe α ($= 0{,}01$ par ex) et on détermine μ_α

3) On calcule la statistique de test (ici $\bar{x} = \frac{1}{n}\sum x_i$)

que l'on compare à μ_α et on accepte ou rejette l'hypothèse H_0 avec le risque $\alpha = -$

4) On calcule β

17/01/2000

Une représentation habituelle permettant d'analyser les performances d'un test consiste à tracer les courbes COR (Caractéristiques Opérationnelles du Récepteur) définies par $PD = 1-\beta = f(\alpha)$
↳ probabilité de Détection

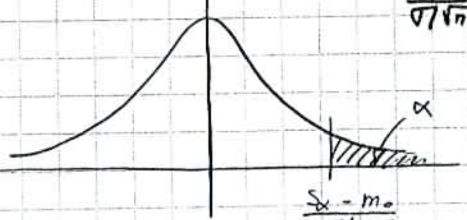
$\alpha = PFA = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$ $\qquad \beta = PND = P[\text{Rejeter } H_1 / H_1 \text{ vraie}]$

__exemple__ : On se fixe $\alpha = P[\overline{X} > S_\alpha / m = m_0]$

$$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_0, \frac{\sigma^2}{n}\right)\right]$$

$$= P\left[\underbrace{\frac{\overline{X} - m_0}{\sigma/\sqrt{n}}}_{U} > \frac{S_\alpha - m_0}{\sigma/\sqrt{n}} / U \sim N(0,1)\right]$$

$$= \int_{\frac{\overline{X}-m_0}{\sigma/\sqrt{n}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$



$$\frac{S_\alpha - m_0}{\sigma/\sqrt{n}}$$

$$\int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = erfc(x)$$ __Tabulé__
↳ fct° MATLAB

$+ x \nearrow, + erfc \searrow$

Donc $\frac{S_\alpha - m_0}{\sigma/\sqrt{n}} = erfc^{-1}(\alpha)$

$$\boxed{S_\alpha = m_0 + \frac{\sigma}{\sqrt{n}} erfc^{-1}(\alpha)}$$

$$PD = 1 - \beta = P[\text{accepter } H_1 / H_1 \text{ vraie}]$$

$$= P[\text{rejeter } H_0 / H_1 \text{ vraie}]$$

$$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_1, \frac{\sigma^2}{n}\right)\right]$$

$$= \boxed{erfc\left(\frac{S_\alpha - m_1}{\sigma/\sqrt{n}}\right) = PD}$$

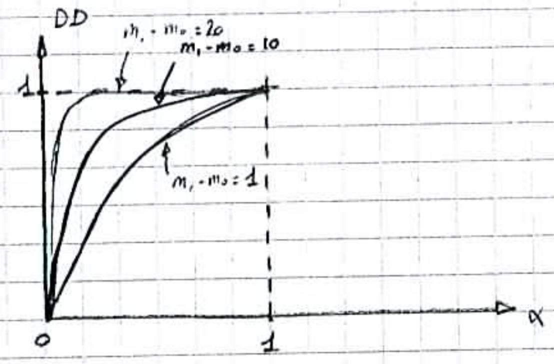
__Courbes COR__ : $\boxed{PD = erfc\left[\frac{-\sqrt{n}(m_1 - m_0)}{\sigma} + erfc^{-1}(\alpha)\right]}$

__Interprétations__ : ① plus $n \nearrow$ plus $erfc \nearrow$ plus $PD \nearrow$

qd $n \to +\infty$ $PD \to 1$

② plus $m_1 - m_0$ ↗ (H₀ et H₁ "éloignés") plus PD ↗

③ plus σ ↘ plus PD ↗



III.2 Cas discret : ($X = (X_1 \dots X_n)$ est un échantillon de VA discrètes)

Lemme de Neyman - Pearson :

> Parmi tous les tests dont le risque de 1ère espèce est $\leq \alpha$, il en existe un de puissance ($PD = 1-\beta$) maximale défini par
>
> Rejet de H_0 si $\dfrac{L(x_1 \dots x_n / H_1)}{L(x_1 \dots x_n / H_0)} > S_\alpha$

Exemple : $X_i \sim \mathcal{P}(\lambda)$ H_0 $\lambda = \lambda_0$ H_1 $\lambda = \lambda_1$ $(\lambda_1 > \lambda_0)$

Rejet de H_0 si $\dfrac{P[X_1 = x_1, \dots, X_n = x_n / H_1]}{P[X_1 = x_1, \dots, X_n = x_n / H_0]} > S_\alpha$

si $\dfrac{\prod_{i=1}^{n} e^{-\lambda_1} \dfrac{\lambda_1^{x_i}}{\cancel{x_i!}}}{\prod_{i=1}^{n} e^{-\lambda_0} \dfrac{\lambda_0^{x_i}}{\cancel{x_i!}}} > S_\alpha$

$$-n(\lambda_1 - \lambda_0) + \Sigma x_i (\ln \lambda_1 - \ln \lambda_0) > \ln S_\alpha$$

$$\boxed{\text{Rejet de } H_0 \text{ si } \sum_{i=1}^{n} x_i > \mu_\alpha}$$

A.N.: $n = 2 \quad \lambda_0 = 1 \quad \lambda_1 = 2 \quad \alpha = 0,05$

$\alpha = 0,05 = P[\text{Rejeter } H_0 / H_0 \text{ vraie}] = P\left[\sum_{i=1}^{2} X_i > \mu_\alpha / \lambda = \lambda_0\right]$

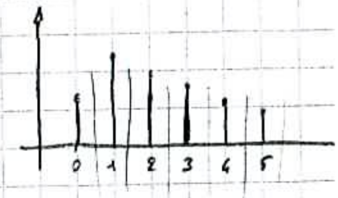
Si $\lambda = \lambda_0$ $X_i \sim P(\lambda_0)$ alors on montre que $X_1 + X_2 \sim P(2\lambda_0)$

$$\alpha = 0,05 = P[U > \mu_\alpha / U \sim P(2)]$$

Si $\mu_\alpha \leq 0$ $\quad P[U > \mu_\alpha] = 1 \quad \underline{\alpha = 1}$

si $\mu_\alpha \in]0,1]$ $\quad P[U > \mu_\alpha] = 1 - P[U = 0] = 1 - \frac{(2\lambda_0)^0}{0!} e^{-2\lambda_0} \# 0,865$ Table

si $\mu_\alpha \in]1,2]$ $\quad P[U > \mu_\alpha] = 1 - P[U = 0 \text{ ou } U = 1]$ ...



...

Si $\mu_\alpha \in]3,4]$ $\quad P[U > \mu_\alpha] = P[U > 4] = 0,0527$

$\boxed{\text{Si } \mu_\alpha \in]4,5] \quad P[U > \mu_\alpha] = 0,0166}$

Conclusion: Le test de risque $\alpha \leq 0,05$ le plus puissant est
Rejet de H_0 si $x_1 + x_2 > 5$
il admet un risque de 1ère espèce $\alpha \# 0,0166$

Rq: Si n grand, on approche la loi de U par une loi normale en vertu du théorème de la limite centrale

ici on aurait $\quad U = \sum_{i=1}^{n} X_i \ \tilde{\#} \ \mathcal{N}(n\lambda, n\lambda)$

IV Test du rapport des vraisemblances généralisé:

(ou Test GLR (Generalized Likelihood ratio))

Ce test est adapté aux hypothèses composées $\begin{cases} H_0 & \theta \in A \\ H_1 & \theta \in B \end{cases}$

où les ensembles A et B ne sont pas tous deux réduits à un point.

Le test GLR est défini par

$$\text{Rejet de } H_0 \text{ si } \frac{L(x_1, \dots, x_n \,;\, \widehat{\theta_1}_{mv})}{L(x_1, \dots, x_n \,;\, \widehat{\theta_0}_{mv})} > S_\alpha$$

où $\widehat{\theta}_{0\,mv}$ et $\widehat{\theta}_{1\,mv}$ sont les estimateurs du maximum de vraisemblance de θ sous les hypothèses H_0 et H_1 respectivement.

Rq: $$\frac{L(x_1, \dots, x_n \,;\, \widehat{\theta_1}_{mv})}{L(x_1, \dots, x_n \,;\, \widehat{\theta_0}_{mv})} = \frac{\sup_{\theta \in B} L(x_1 \dots x_n \,;\, \theta)}{\sup_{\theta \in A} L(x_1 \dots x_n \,;\, \theta)}$$

V Test du χ^2 :

Le test du χ^2 est un test d'ajustement (ou d'adéquation) non paramétique, de la forme (H_0) $L = L_0$ une certaine loi

(H_1) $L \neq L_0$

On se propose de tester si l'échantillon $(X_1 \dots X_n)$ est de loi L_0 (discrète ou continue) ou non.

Principe: On découpe le support de la loi L_0 connue en k parties appelées classes et notées $C_1 \dots C_k$.

On note $P_{0k} = P[X_j \in C_k \,/\, X_j \sim L_0]$

Exemple: $X_j \sim N(m, \sigma^2)$, $m = 0$, $\sigma^2 = 1$ $\quad C_1 =]-\infty, 0[$ $\quad C_2 = [0, 1]$ $\quad C_3 =]1, +\infty[$

$P_{01} = P[X_j \in C_1] = \frac{1}{2}$ $\qquad P_{02} = \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\,)\, dx$

$P_{03} = \int_1^{+\infty} \_\_\_$

On définit les va Y_j^k de la façon suivante :

$$Y_j^k = \begin{cases} 1 & \text{si } x_j \in C_k \\ 0 & \text{si } x_j \notin C_k \end{cases}$$

Alors $Z_k = \sum_{j=1}^{n} Y_j^k$ représente la va "nbre d'observations $\in C_k$"

D'après le théorème de la limite centrale, (les va Y_j^k sont ind. dans la mesure où les x_j le sont); pour n grand

$$\boxed{Z_k \underset{\#}{\approx} \mathcal{N}(nP_{0k}, nP_{0k}(1-P_{0k}))}$$ car $E[Y_j^k] = 1 \cdot P[x_j \in C_k] + 0 \cdot P[x_j \notin C_k]$

$\operatorname{Var} Y_j^k = P_{0k}(1-P_{0k})$ $\left(= E[(Y_j^k)^2] - E[Y_j^k]^2 = P_{0k} - (P_{0k})^2 \right)$

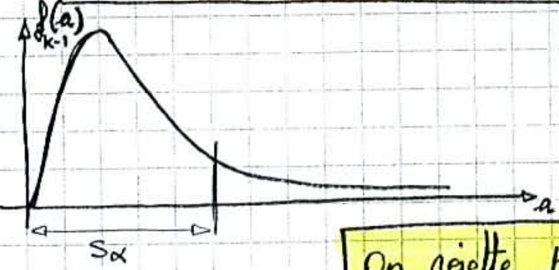
De manière équivalente $U_k = \dfrac{Z_k - nP_{0k}}{\sqrt{nP_{0k}}} \sim N(0, 1-P_{0k})$

Rq : les VA U_k ne sont pas indépendantes car $\sum_{k=1}^{n} Z_k = n$

(n = nbre de x_i)

On admettra que sous l'hypothèse H_0 ($X_i \sim L_0$)

$$\boxed{A = \sum_{k=1}^{k} \frac{(Z_k - nP_{0k})^2}{nP_{0k}} \xrightarrow[n \to \infty]{L} \chi^2_{k-1}}$$



$$\boxed{\text{On rejette } H_0 \text{ si } A > S_\alpha}$$

Le calcul de S_α se fait comme suit, on se fixe α (par ex 0,01)

$$\alpha = 0{,}01 = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$$

$$= P[A > S_\alpha / X_i \sim L_0] = \int_{S_\alpha}^{+\infty} f_{k-1}(a)\, da$$ → Tables du χ^2

les tables du χ^2 donnent S_α

$$\sum_{k=1}^{K} \frac{(Z_k - np_{0k})^2}{np_{0k}} \xrightarrow[n\to+\infty]{L} \chi^2_{K-1}$$

$Z_k = \sum_{j=1}^{n} Y_j^k$ donc $E[Z_k] = n\,E[Y_j^k] = n\,p_{0k}$

$\mathrm{cov}(Z_i, Z_k) = E[Z_i Z_k] - E[Z_i]E[Z_k]$

$$E[Z_i Z_k] = \sum_{j_1}\sum_{j_2} \underbrace{E[Y_{j_1}^k Y_{j_2}^i]}_{0 \text{ pour } j_1 \neq j_2} = \sum_j \underbrace{E[Y_j^k Y_j^i]}_{=0 \text{ pour } k\neq i}$$

donc $\boxed{\mathrm{cov}(Z_i, Z_k) = -n^2 p_{0i} p_{0k}}$

$$\mathrm{Var}\, Z_i : \mathrm{Var}\left(\sum_j Y_j^k\right) = \sum_j \mathrm{Var}(Y_j^k) = n\left[\underbrace{1\,p_{0k} + 0\,(1-p_{0k})}_{E[(Y_j^k)^2]} \underbrace{- p_{0k}^2}_{-(E[Y_j^k])^2}\right]$$

$\boxed{E[Z_k] = np_{0k}}$ et $\boxed{\mathrm{Var}\, Z_k = n\, p_{0k}(1-p_{0k})}$

les variables aléatoires Z_k sont liées car $\sum_{k=1}^{K} Z_k = n$

nbre d'observations $\in$ à la classe C_k

Donc, on ne considère que

$$Z = \left(\frac{Z_1 - np_{01}}{\sqrt{n}}, \frac{Z_2 - np_{02}}{\sqrt{n}}, \ldots, \frac{Z_{K-1} - np_{0K-1}}{\sqrt{n}}\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & & & \\ -np_1p_2 & p_2(1-p_2) & & \\ & & \ddots & \\ -np_1p_{K-1} & & & p_{K-1}(1-p_{K-1}) \end{pmatrix}$$

<u>1er théorème sur les formes quadratiques</u>

$X \sim N_p(\mu, \Sigma) \Rightarrow D^2 = (X-\mu)^T \Sigma^{-1} (X-\mu) \sim \chi^2_p$

on applique ce théorème à Z et ça marche ...

Rq 1, $A = \sum_{k=1}^{k} \frac{n}{P_{0k}} \left(\frac{Z_k}{n} - P_{0k} \right)^2$

estimation de P_{0k}

On comprend que si H_0 est vraie $\frac{Z_k}{n} \neq P_{0k}$ donc A est "petit"

On aura alors $A < S_\alpha$ et on acceptera H_0

Rq 2: la loi de A est valable en théorie pour $n = +\infty$.

En pratique n est fini. On montre que la loi reste valable pour n fini si $n P_{0k} > 4$ ou 5 $\forall k$

On a intérêt à choisir des classes de façon à avoir P_{0k} "grand" $\forall k$

$\Rightarrow$ classes équiprobables

Rq 3: Si L_0 est imparfaitement connue (N_p paramètres inconnus
ex: $N(m, \sigma^2)$, $N_p = 2$)

On estime m et σ^2 et $A \sim \chi^2_{K-1-N_p}$

VI Test de kolmogorov : (va continues)

Test d'ajustement : H_0 $L = L_0$

H_1 $L \neq L_0$

On teste si la loi des VA X_i est L_0 ou non à partir de l'observation

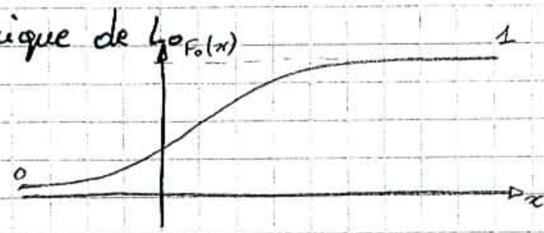
$x_1, \dots, x_n$.

le test de kolmogorov est défini par :

$$\boxed{\text{Rejet de } H_0 \text{ si } \sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F_0(x) \right| > S_\alpha}$$

où $F_0(x)$ est la fct° de répartition théorique de L_0

$F_0(x) = \int_{-\infty}^{x} P_0(u)\, du$
densité de proba associée à L_0

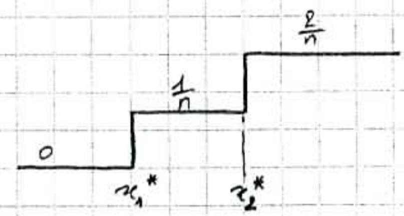


et $\hat{F}(x)$ est la fct° de répartition empirique

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)| = \sup_{i \in \{1,\dots,n\}} \left[\max \left(E_i^+, E_i^- \right) \right]$$

$$E_i^+ = \left| \hat{F}(x_i^{*+}) - F_0(x_i^*) \right|$$
$$E_i^- = \left| \hat{F}(x_i^{*-}) - F_0(x_i^*) \right|$$

<u>Si tous les x_i sont différents</u>



$$\hat{F}(x_i^{*+}) = \frac{i}{n} \quad \text{et} \quad \hat{F}(x_i^{*-}) = \frac{i-1}{n}$$

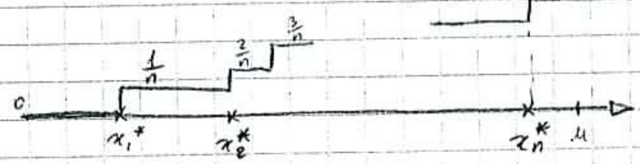
$$\boxed{D_n = \sup_i \left[\max \left| \frac{i}{n} - F_0(x_i^*) \right|, \left| \frac{i-1}{n} - F_0(x_i^*) \right| \right]}$$

<u>Rq 3:</u> • Le test de Kolmogorov est plus puissant que le test du χ^2 dans la plupart des applications (puissance $= 1 - \beta$)
• Attention, son utilisation est limitée aux lois <u>continues</u>

<u>Rq 4:</u> $\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$
et on rejette H_0 si $D_n > S_\alpha$, donc plus $\alpha \nearrow$
plus $S_\alpha \searrow$

et donc plus il est difficile d'accepter H_0

Si on accepte H_0, le résultat est d'autant plus fiable que α est grand



$x_1^* \leq x_2^* \leq \dots \leq x_n^*$

(On dit que $(x_1^*, \dots, x_n^*)$ est la statistique d'ordre de $(x_1, \dots, x_n)$)

Rq : Si tous les x_i^* sont différents deux à deux, les sauts de la fonction en escalier sont de $\frac{1}{n}$

On montre que $D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)|$ possède la même loi asymptotique $(n \to \infty)$ sous l'hypothèse H_0 quelle que soit la loi à tester L_0

$$P[\sqrt{n}\, D_n < y] \xrightarrow[n \to +\infty]{} \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y) = k(y)$$

Détermination de S_α :

$$\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$$
$$\alpha = P[D_n > S_\alpha / H_0 \text{ vraie}]$$
$$\alpha = 1 - P[D_n \leq S_\alpha]$$
$$\alpha = 1 - P[\sqrt{n}\, D_n \leq \sqrt{n}\, S_\alpha]$$

$$\boxed{\alpha \underset{n\,gd}{\#} 1 - k(\sqrt{n}\, S_\alpha)}$$

On a alors construit des tables qui donnent S_α en fonction de α

Rq 1 : pour $n > 80$ $\quad \alpha = 0{,}05 \quad S_\alpha \# \frac{1{,}3581}{\sqrt{n}}$

$\alpha = 0{,}01 \quad S_\alpha \# \frac{1{,}6276}{\sqrt{n}}$.

Rq 2 : Calcul de D_n