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"

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

On dit que $(X_1, \dots, X_n)$ est un échantillon.

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

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

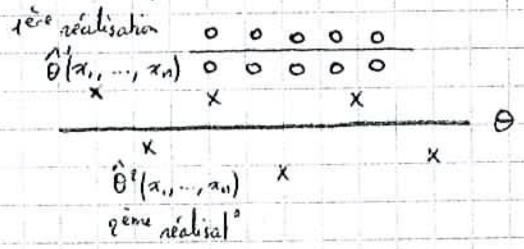
Qualités d'un estimateur : $\hat{\theta}$

- **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.

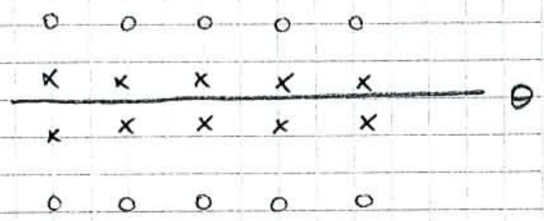
l'estimateur o est biaisé : il y a une erreur systématique. (ex : balance car 0 mal réglé)

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

- **Variance de l'estimateur**

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

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



$\operatorname{Var} X < \operatorname{Var} O$

$\Rightarrow X$ 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 : $\boxed{EQM(\hat\theta) = \operatorname{Var}\hat\theta + (\operatorname{Biais}(\hat\theta))^2}$

Si on a $\operatorname{var}\hat\theta$ 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
$$EQM(\hat\theta) = \operatorname{Var}\hat\theta + \underbrace{E\left[(\operatorname{Biais}\hat\theta)^2\right]}_{(\operatorname{Biais}\hat\theta)^2} + 2\left(E[\hat\theta] - \theta\right)\underbrace{E\left[\hat\theta - E[\hat\theta]\right]}_{E[\hat\theta] - E[E[\hat\theta]]}$$

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

Définition : On dit que l'estimateur $\hat\theta$ est convergent si
$$\hat\theta \xrightarrow[n\to\infty]{P} \theta$$

Condition suffisante de convergence :

Si $\operatorname{biais}(\hat\theta) \xrightarrow[n\to+\infty]{} 0$ et $\operatorname{var}(\hat\theta) \xrightarrow[n\to+\infty]{} 0$ 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 $\left.\begin{array}{l}\operatorname{biais}(\hat\theta) \xrightarrow[n\to+\infty]{} 0 \\ \operatorname{Var}\hat\theta \xrightarrow[n\to+\infty]{} 0\end{array}\right. \Rightarrow EQM(\hat\theta) \xrightarrow[n\to+\infty]{} 0 \iff \hat\theta \xrightarrow[n\to+\infty]{mq} \theta$

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 $\mathrm{var}\, X_i = \sigma^2$. 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

$1+2+\dots+n$

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

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

$\hat{\theta}_2$ non biaisé

Variances:

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

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

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

Rq: $\mathrm{Var}\, \hat{\theta}_1 \xrightarrow[n \to +\infty]{} 0$

Biais $\hat{\theta}_1 = 0$

Donc $\hat{\theta}_1$ est convergent

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

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

$\mathrm{Var}\, \hat{\theta}_2 \xrightarrow[n \to +\infty]{} 0$

Biais $\hat{\theta}_2 = 0$

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

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

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

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

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

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

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

li de ca

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

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

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

$\hat{\sigma}^2$ 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 $\Rightarrow$ $\hat{m}$ 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 ?

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

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

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

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

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

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

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

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

($m = E[\bar{X}]$)

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

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

$i \neq j$ $cov(X_i, X_j) = 0$ 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 :

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

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 1 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

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

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

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

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

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

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

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

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

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

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

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

→ précis° de l'estimation optimale

<u>Rq 3:</u> 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 : $X_i \sim P(\lambda)$ 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

$(\text{—}) (|) \to a \in \mathbb{R}$ ${}^t x . x$

$(|)$ (—) $\to$ matrice $x({}^t x)$

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

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

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

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

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

$\hat{\theta}_{MO}$ 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

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

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

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

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

On vérifie que la solution donne un maximum en étudiant

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

$\dfrac{\partial^2 L(x_1 \dots x_n ; \hat{\theta}_{MV})}{\partial \theta^2} < 0$ (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	↗ ↘

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 / 2020

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

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

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

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

V - 2 - Propriétés :

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

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

* 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 $I_n(\theta) = E\left[\dfrac{-\partial^2 \ln(X_1 \cdots X_n ; \theta)}{\partial \theta^2} \right]$

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

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

V - 3 - Exemples :

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

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

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

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

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

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

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

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

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

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 $\hat{\theta}(x)$

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

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

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

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

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

Preuve : On cherche à minimiser $E\left[\left(\theta - \hat{\theta}(X)\right)^2\right]$ ie On cherche $\hat{\theta}(X)$ qui minimise $\underset{X}{E}\left[\underset{\theta}{E}\left[\left(\theta - \hat{\theta}(X)\right)^2 \middle|_X\right]\right]$

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

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

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

$\hat{\theta}$ 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 $E\left[\left(\theta - \hat{\theta}(X)\right)^2\right]$ $\boxed{\text{CQFD}}$

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

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

$\boxed{\text{Calcul de } E[\theta/X]}$ 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

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

X_i ind

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

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

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

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

proportionnel

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

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

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

quand n gd $\frac{\sigma_b^2}{n} \ll \sigma^2$ $\hat{\theta}_{MAP} \# \bar{X}$

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

quand n petit $\frac{\sigma_b^2}{n} \gg \sigma^2$ $\hat{\theta}_{MAP} \# m$

(n=0)

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

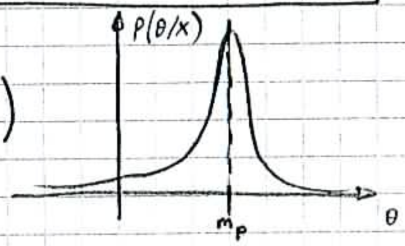
L'estimateur $\hat{\theta}(x)$ qui minimise $E\left[C\left(\theta, \hat{\theta}(x)\right)\right]$ avec

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

Δ "arbitrairement" petit est défini par

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

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



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

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 les mesures où les VA st indépendant

(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

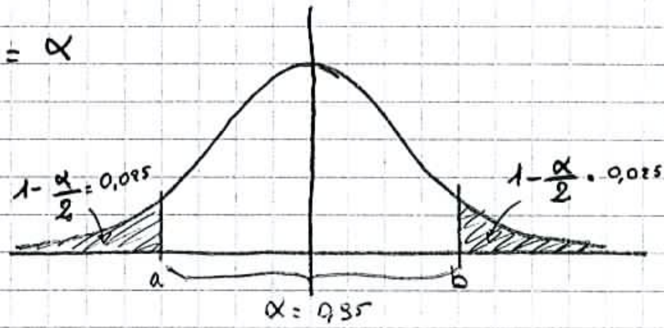
$$\boxed{T(X_1 \dots X_n) = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - m}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \rightarrow \text{Tabulé}}$$

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

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

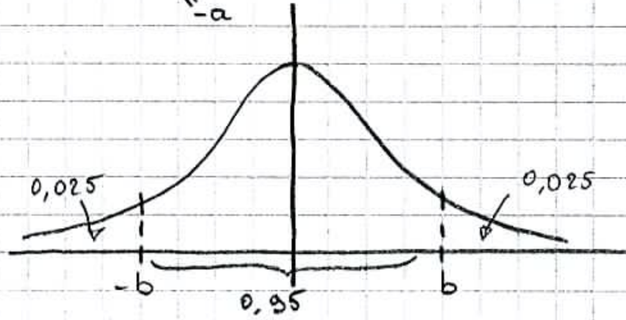
il se détermine grâce à

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

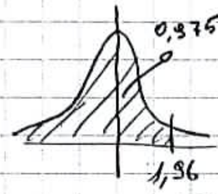


<u>Prenons $\alpha = 0{,}95 = 95\%$</u>

On détermine a et $\underset{=-a}{b}$ à l'aide des tables



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



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

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

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

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

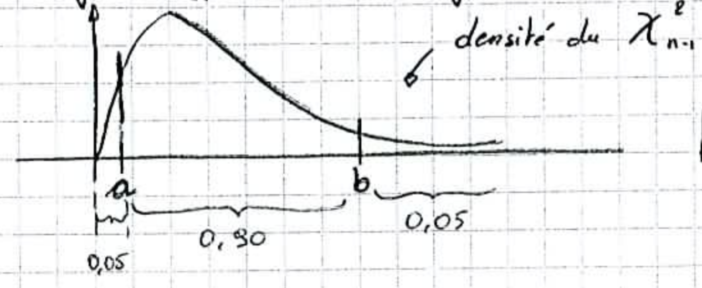
$\left[\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right]$ est l'intervalle de confiance pour m avec $\alpha = 0{,}95$

Ex 2 : $X_i \sim N(m, \sigma^2)$ <u>m inconnue</u>

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 $\begin{cases} a = \_\_\_\_\_\_ \\ b = \_\_\_\_\_\_ \end{cases}$

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

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

<u>Ex 3</u> : $X_i \sim N(m, \sigma^2)$ 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 $\dfrac{U}{\sqrt{\frac{V}{n-1}}} \sim \underbrace{t_{n-1}}_{\text{loi de Student ou du t, à } n-1 \text{ d° de libertés}}$

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

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

VIII Méthode des moindres carrés : Exemple de la régression linéaire

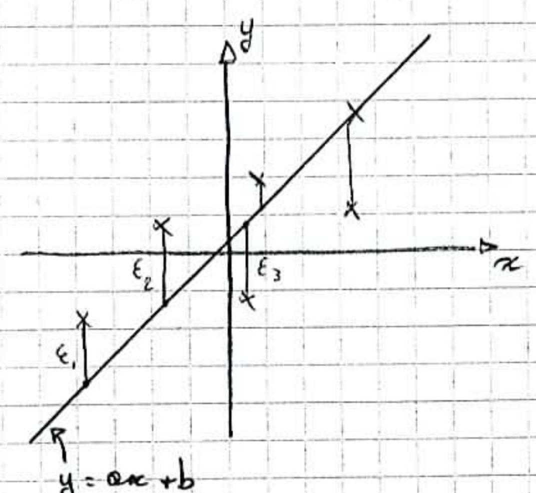
1) Principe

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)



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

x_i : abscisses observées non aléatoires

ε_i : erreur

y_i observée

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) $n\overline{Y} - an\overline{x} - nb = 0 \Rightarrow \boxed{b = \overline{Y} - a\overline{x}}$

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

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

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

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

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

Nous trouvons alors l'approximation affine :

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

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 :

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

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

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 $\mathrm{E}[Y-b]^2$ 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 $Y = m_Y$.

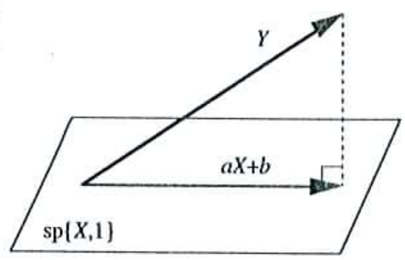
Remarque : Ces résultats peuvent s'obtenir plus simplement de manière géométrique en considérant l'espace de Hilbert $\mathrm{L}^2(\Omega, \mathcal{C}, P)$ des variables aléatoires du second ordre définies sur Ω :

D'après les propriétés des espaces de Hilbert, trouver l'approximation affine $aX+b$ de Y qui minimise $\mathrm{E}[Y-aX-b]^2$ 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 $1_\Omega(\omega)$) :

$$Y-aX-b \perp X \iff \mathrm{E}[(Y-aX-b)X] = 0$$

$$Y-aX-b \perp 1 \iff \mathrm{E}[(Y-aX-b)1] = 0$$

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 $\sigma_Y \neq 0$: $x = \mathrm{E}(X) + \rho_{X,Y}\dfrac{\sigma_X}{\sigma_Y}[y - \mathrm{E}(Y)]$. Soit :