

VII - Méthodes d'Estimation

Principe des méthodes d'estimation

Soit un échantillon aléatoire de mesures $\underline{x} = (x_1, x_2, \dots, x_n)$ extraits d'une population sous-jacente définie par le fdp $f(x | \underline{\theta}_0)$ a k paramètres $\underline{\theta}_0 = (\theta_{0,1} \dots \theta_{0,k})$ dont on veut estimer les valeurs à partir de l'échantillon.

- Estimations ponctuelles :
meilleures valeurs $\hat{\underline{\theta}}$ de $\underline{\theta}_0$ étant donné l'échantillon \underline{x}
- Calcul des matrices de variance-covariance,
et des intervalles de valeurs correspondant à un niveau de confiance donné.

L'information que l'on tire du fait que la distance Terre-Soleil est de

149 597 870 660 m

n'est pas la même, suivant que l'erreur standard sur cette mesure est de

20 m ou 20 000 000 m

Définitions et notations

Vraie valeur du paramètre: θ_0 (inconnue)

Statistique: $t = t(\underline{x}, \theta)$

Estimateur: statistique $t(\underline{x})$ permettant de donner une valeur au paramètre θ

Estimation: valeur prise par l'estimateurs pour l'ensemble \underline{x} observé:

$$\hat{\theta} = t(\underline{x})$$

Vraisemblance: probabilité jointe conditionnelle d'observer l'échantillon \underline{x} pour une valeur de θ

$$\mathcal{L}(\underline{x} | \theta) = \prod_{i=1}^n f(x_i | \theta)$$

Fonction de vraisemblance: $\mathcal{L}(\theta | \underline{x})$ comme fonction de θ pour l'échantillon \underline{x} observé

Approche classique (ou fréquentiste ou de Neyman) de la définition de l'intervalle de confiance pour la mesure d'une variable continue (rappel)

Soient:

- un échantillon de mesures $\underline{x} = (x_1, x_2, \dots, x_n)$ extrait de la population $f(x | \theta_0)$
- estimateur $t(\underline{x})$ de θ_0
- estimation $\hat{\theta} = t(\underline{x})$ de θ_0 pour échantillon \underline{x} observé
- fdp $g(t | \theta)$ de t pour une valeur θ donnée

La connaissance de $g(t | \theta)$ est capitale dans la réalisation d'une mesure expérimentale. Pour qu'une procédure expérimentale ait un sens, il faut qu'il soit possible de connaître a priori la fdp de la valeur de la mesure pour toute valeur possible de la vraie valeur θ_0 . Du moins dans un domaine de valeurs $[\theta_{min}, \theta_{max}]$ pour lequel l'expérience a été conçue.

Si θ_0 n'appartient pas au domaine $[\theta_{min}, \theta_{max}]$, l'expérience sera un échec

- soit parce que $g(t | \theta)$ ne peut plus être calculée de manière fiable,
- soit parce que la variance de $g(t | \theta)$ devient si grande que l'expérience perd toute résolution.

Exemples:

1/ mesure de la moyenne d'un échantillon \underline{x} extrait

d'une fdp poissonniene $f(x | \mu_0) = \frac{1}{x!} \mu_0^x e^{-\mu_0}$

$$\hat{\mu} = t \equiv \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$g(t | \mu) = N\left(t | \mu, \frac{\mu^2}{n}\right) \text{ si } n \text{ grand}$$

2/ mesure d'une grandeur θ de vraie valeur θ_0

avec une erreur standard σ

$\hat{\theta} = t(x) \equiv x =$ résultat de la mesure

$$g(t | \theta) = N(t | \theta, \sigma^2)$$

3/ mesure d'un nombre d'occurences x d'un processus poissonien : $f(x | \mu_0) = \frac{1}{x!} \mu_0^x e^{-\mu_0}$

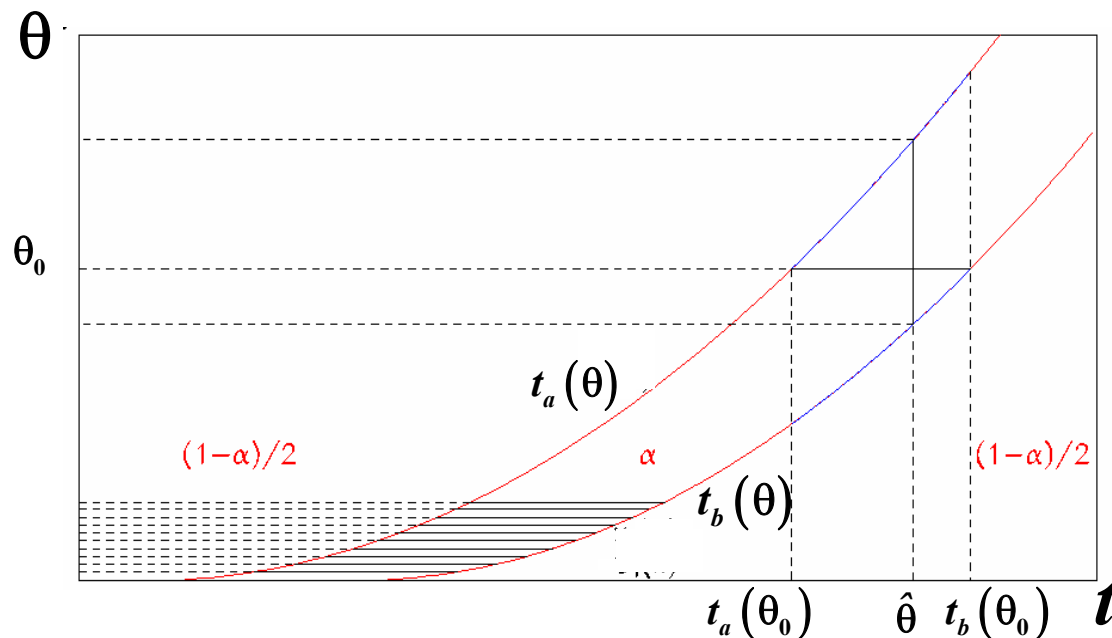
$\hat{\mu} = t(x) \equiv x =$ résultat de la mesure

$$g(t | \mu) = \frac{1}{t!} \mu^t e^{-\mu}$$

Ceinture de confiance de Neyman centrée sur la vraie valeur

Calculer pour un nombre fini de valeurs de θ dans le domaine attendu

les contours $t_a(\theta)$ et $t_b(\theta) \Rightarrow \int_{t_a(\theta)}^{t_b(\theta)} g(t|\theta) dt = \alpha$ et $\int_{-\infty}^{t_a(\theta)} g(t|\theta) dt = \int_{t_b(\theta)}^{\infty} g(t|\theta) dt = \frac{1-\alpha}{2}$

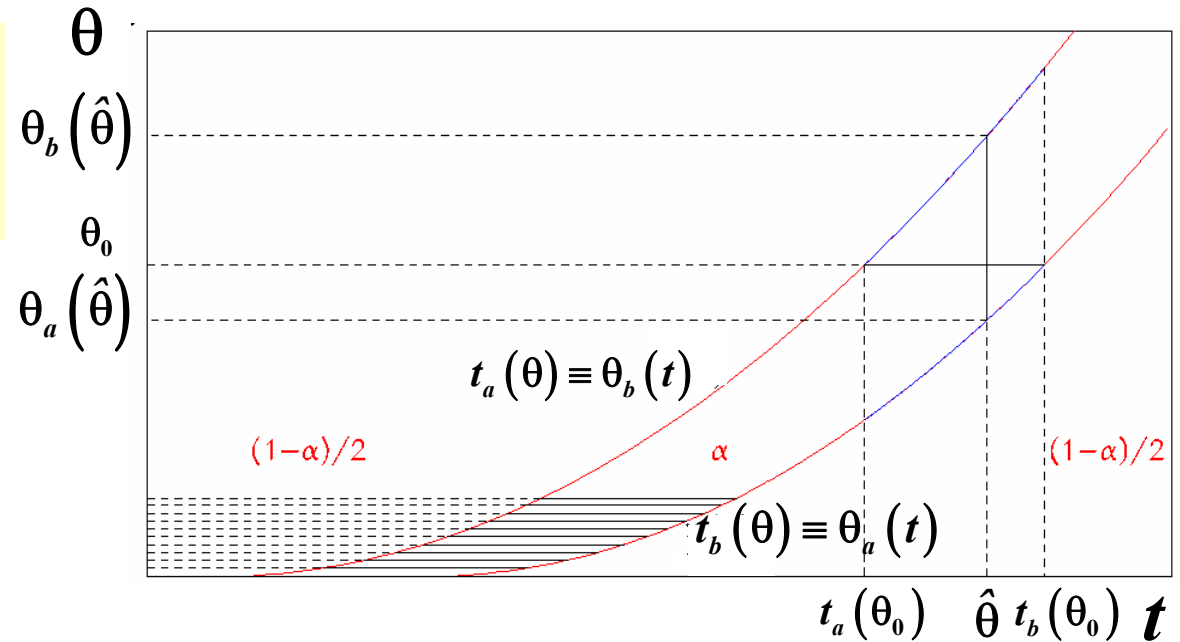


Pour toute mesure $\hat{\theta} = t(\underline{x})$ de θ_0 on a que

$$P(\hat{\theta} \in [t_a(\theta_0), t_b(\theta_0)]) = \alpha$$

Signification de l'intervalle de confiance dans l'approche classique

$t_{a,b}(\theta), \theta_{a,b}(t)$ définissent la **Ceinture de Neyman** dans le plan θ, t



$$\hat{\theta} = t(\underline{x})$$

$$P(\hat{\theta} \in [t_a(\theta_0), t_b(\theta_0)]) = \alpha \quad \Rightarrow$$

$$P(\theta_0 \in [\theta_a(\hat{\theta}), \theta_b(\hat{\theta})]) = \alpha$$

$$\theta_{a,b}(t) = \text{relation inverse de } t_{b,a}(\theta)$$

$\hat{\theta}$ est une variable aléatoire

θ_0 est une constante

$t_a(\theta_0), t_b(\theta_0)$ sont des constantes

$\theta_a(\hat{\theta}), \theta_b(\hat{\theta})$ sont des variables aléatoires

Signification de l'intervalle de confiance dans l'approche classique

Que signifie la relation $P\left(\theta_0 \in \left[\theta_a, \theta_b\right]\right) = \alpha$??

dans $\alpha\%$ des cas, l'affirmation $\theta_0 \in \left[\theta_a, \theta_b\right]$ est aléatoirement vraie

dans $(1-\alpha)\%$ des cas, l'affirmation $\theta_0 \in \left[\theta_a, \theta_b\right]$ est aléatoirement fausse

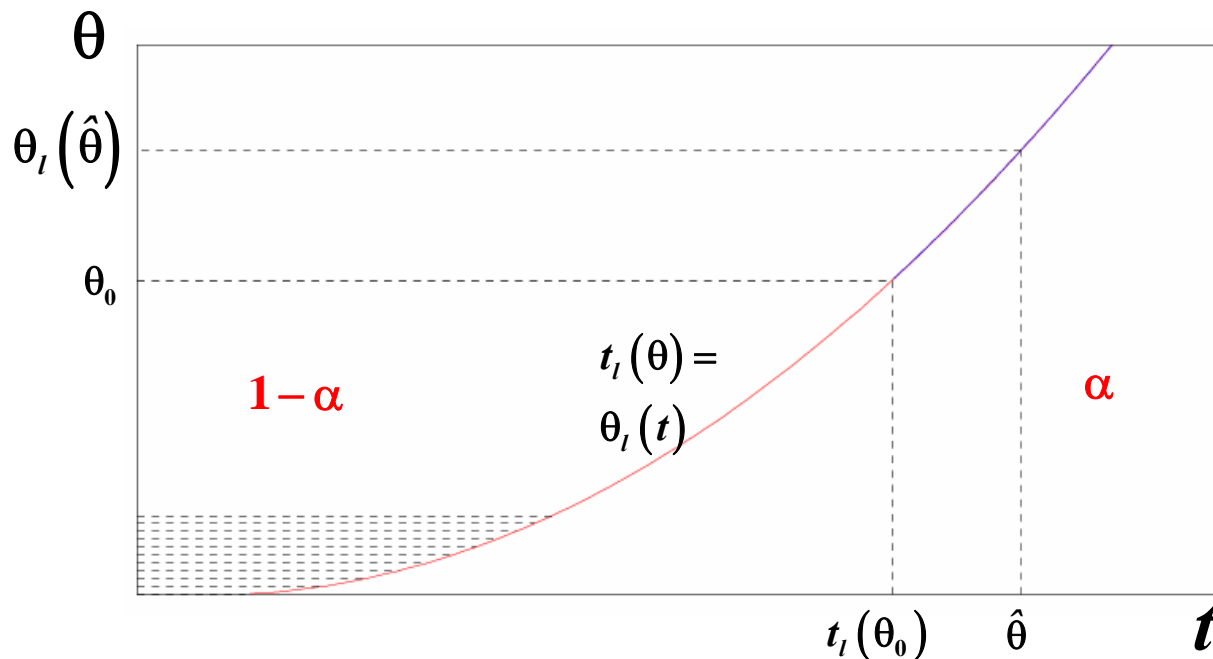
α = niveau de confiance (*C.L.*)

$\left[\theta_a, \theta_b\right]$ = intervalle de confiance au niveau de confiance α

Ceinture de confiance de Neyman pour une limite supérieure

Calculer pour un nombre fini de valeurs de θ dans le domaine attendu

le contour $t_l(\theta) \Rightarrow \int_{-\infty}^{t_l(\theta)} g(t | \theta) dt = 1 - \alpha$ et $\int_{t_l(\theta)}^{+\infty} g(t | \theta) dt = \alpha$



Pour toute mesure $\hat{\theta} = t(\underline{x})$ de θ_0 on a que $P(\hat{\theta} \geq t_l(\theta_0)) = \alpha$

Signification de la limite supérieure dans l'approche classique

$$\hat{\theta} = t(\underline{x})$$

$$P(\hat{\theta} \geq t_l(\theta_0)) = \alpha$$

\Rightarrow

$$P(\theta_0 \leq \theta_l(\hat{\theta})) = \alpha$$

$\theta_l(t) =$ relation inverse de $t_l(\theta)$

$\hat{\theta}$ est une variable aléatoire

$t_l(\theta_0)$ est une constante

θ_0 est une constante

$\theta_l(\hat{\theta})$ est une variable aléatoire

Que signifie la relation $P(\theta_0 \leq \theta_l) = \alpha$??

dans $\alpha\%$ des cas, l'affirmation $\theta_0 \leq \theta_l$ est aléatoirement vraie

dans $(1-\alpha)\%$ des cas, l'affirmation $\theta_0 \leq \theta_l$ est aléatoirement fausse

$\alpha =$ niveau de confiance (C.L.)

$\theta_l =$ limite supérieure au niveau de confiance α

Qualités d'un estimateur

Consistance

$$\lim_{n \rightarrow \infty} \hat{\theta} = \theta_0$$

$$\bar{x} \rightarrow \mu, \quad s^2, S^2 \rightarrow \sigma^2$$

Absence de biais

$$E[\hat{\theta}] = \theta_0 \quad E[\bar{x}] = \mu, \quad E[s^2, S^2] = \sigma^2$$

$$E\left[s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right] = \frac{n-1}{n} \sigma^2 \quad \text{mais} \quad \lim_{n \rightarrow \infty} s'^2 = \sigma^2$$

Si $\tau = \tau(\theta)$ univoque et réciproque: $\hat{\tau} = \tau(\hat{\theta})$

généralement $E[\tau(\hat{\theta})] \neq \tau(E[\hat{\theta}])$

$\Rightarrow \hat{\theta}$ et $\hat{\tau}$ ne sont pas simultanément non biaisés

Variance minimale

Étant donné:

- la forme de la fdp de la population dont est extrait l'échantillon
- la taille de l'échantillon

Meilleure estimation si :

- variance petite
- taille grande

Inégalité de Cramer-Rao :

Étant donné la quantité d'information disponible pour estimer θ :

La variance sur l'estimateur a une valeur minimale :

$$V(t) = E[(t - \theta)^2] \geq \frac{1}{A(\underline{x}, \theta)}$$

Indice de capacité informative ne dépendant que de f à taille d'échantillon donné :

$$A(\underline{x}, \theta) = E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right] = E\left[\left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]$$

$$\text{Estimation de } \tau(\theta) \text{ avec biais } b(\theta) : V(t) = E[(t - \tau(\theta))^2] \geq \frac{\left(\frac{\partial \tau}{\partial \theta} + \frac{\partial b}{\partial \theta}\right)^2}{A(\underline{x}, \theta)}$$

Démonstration de l'inégalité de Cramer-Rao

$$\mathcal{L}(\underline{x} | \theta) = \prod_{i=1}^n f(x_i | \theta)$$

$$E[t] = \int \dots \int t(\underline{x}) \mathcal{L}(\underline{x} | \theta) d\underline{x} = \theta \text{ par définition si } t \text{ est non biaisé} \quad (6)$$

$$\int \dots \int \mathcal{L}(\underline{x} | \theta) d\underline{x} = 1 \text{ de part la normalisation de } f \quad (7)$$

dérivation de (6) et (7) par θ :

$$\int \dots \int t \frac{d\mathcal{L}}{d\theta} d\underline{x} = \int \dots \int t \frac{d \log \mathcal{L}}{d\theta} \mathcal{L} d\underline{x} = E \left[t \frac{d \log \mathcal{L}}{d\theta} \right] = 1 \quad (8)$$

$$\int \dots \int \frac{d\mathcal{L}}{d\theta} d\underline{x} = \int \dots \int \frac{d \log \mathcal{L}}{d\theta} \mathcal{L} d\underline{x} = E \left[\frac{d \log \mathcal{L}}{d\theta} \right] = 0$$

+(9)

$$(8) - (9) \times \theta : E \left[(t - \theta) \frac{d \log \mathcal{L}}{d\theta} \right] = 1$$

$$\text{Inégalité de Schwatz : } E[x^2] E[y^2] \geq E[x \times y] \quad (10)$$

↓

Inégalité de Cramer-Rao :

$$V(t) = E[(t - \theta)^2] \geq E \left[\left(\frac{d \log \mathcal{L}}{d\theta} \right)^2 \right]^{-1} \quad (11)$$

$$\text{Condition d'égalité de Schwartz: } y = ax \quad (12)$$

Efficacité

Variance sur estimation = variance minimum

$$V(t) = \left(-E \left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \right] \right)^{-1}$$

Condition d'efficacité:

Condition d'égalité de Schwartz (11) : relation linéaire entre $\frac{\partial \log \mathcal{L}}{\partial \theta}$ et $(t - \theta)$

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = A(\theta) \times (t - \theta)$$

↑ indépendant de \underline{x}

$$-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} = -\frac{\partial A}{\partial \theta} (t - \theta) + A(\theta)$$

$$E \left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \right] = -\frac{\partial A}{\partial \theta} E[t - \theta] + A(\theta) = A(\theta)$$

↑ =0 si estimateur non biaisé

$$V(t) = \frac{1}{A(\theta)}$$

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = A(\theta) \times (t - \tau(\theta) - b(\theta))$$

généralisation

$$V(t) = \frac{\left(\frac{\partial \tau}{\partial \theta} + \frac{\partial b}{\partial \theta} \right)^2}{A(\theta)}$$

Suffisance

L'ensemble de l'information concernant θ est contenue dans l'estimateur.

Condition de suffisance : factorisation de la vraisemblance

$$\mathcal{L}(\underline{x} | \theta) = \prod_{i=1}^n f(x_i | \theta) = g(\underline{x}) \cdot h(t(\underline{x}) | \theta)$$

$g(\underline{x})$ ne dépend pas de θ et ne peut apporter d'information sur θ , tandis que $h(t(\underline{x}) | \theta)$ ne dépend de \underline{x} qu'au travers de l'estimateur $t(\underline{x})$ qui contient donc toute l'information permettant d'estimer θ .

Exemple: L'estimateur $t(\underline{x}) = \frac{1}{n} \sum_{i=1}^n x_i$ de la moyenne $\hat{\mu}$ est suffisant.

Aucune autre statistique ne peut apporter d'information complémentaire sur $\hat{\mu}$.

Un estimateur efficace est toujours suffisant:

La condition de suffisance $\frac{\partial \log \mathcal{L}}{\partial \theta} = \frac{\partial h(t(\underline{x}) | \theta)}{\partial \theta}$ est contenue dans

la condition d'efficacité $\frac{\partial \log \mathcal{L}}{\partial \theta} = A(\theta) \times (t - \theta)$