UV Statistique pour l'Ingénieur

Cours n° 8

Statistique inférentielle : estimation

méthodes d'estimation par intervalle

A. ROGOZAN

Estimation *par intervalle de confiance* d'un paramètre

- Soit T un estimateur ponctuel du paramètre θ d'une loi supposée connue (choisir l'estimateur T le plus précis)

- Objectif: construire un estimateur par intervalle de confiance $[T_1, T_2]$ de niveau de confiance 1-α

  - le risque (statistique) que le paramètre θ soit en dehors de l'intervalle de confiance estimé est égale à α

  - en répétant un grand nombre de fois l'opération d'échantillonnage (ou l'enquête), les estimations $[t_1, t_2]$ des intervalles de confiances contiendront dans (1-α)% des cas le paramètre θ

Intervalle de confiance d'un paramètre

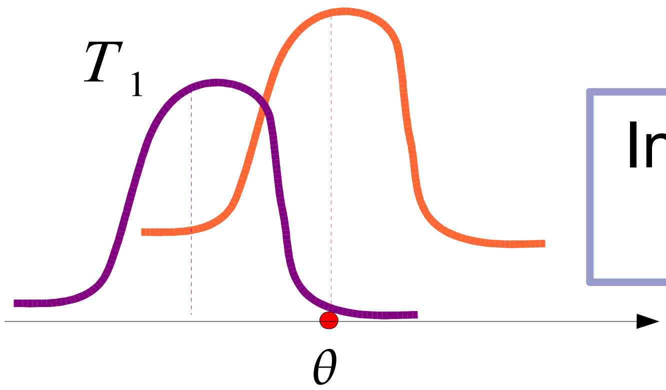
- Soit un intervalle de confiance $[T_1, T_2]$
de niveau de confiance 1-α

  - Intervalle aléatoire : $T_1 = \varphi_1(T)$ et $T_2 = \varphi_2(T)$

  - Il n'y pas d'unicité de l'intervalle de confiance

  - Exemple :

T estimateur sans biais



Intervalle de confiance unilatéral

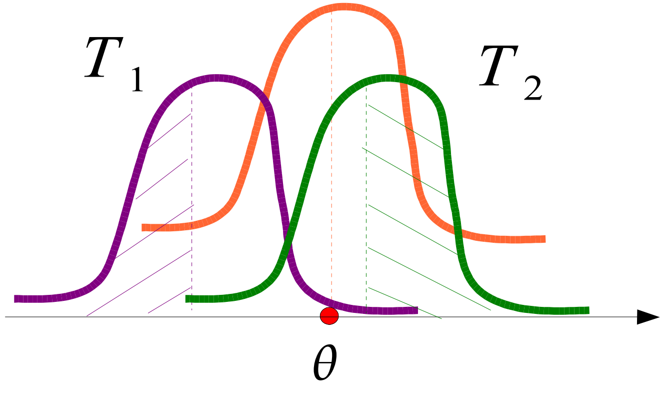
$[T_1, +\infty]$

Intervalle de confiance d'un paramètre

- Soit un intervalle de confiance $[T_1, T_2]$ de niveau de confiance 1-α

$$T_1 = \varphi_1(T)$$
$$T_2 = \varphi_2(T)$$

T estimateur sans biais



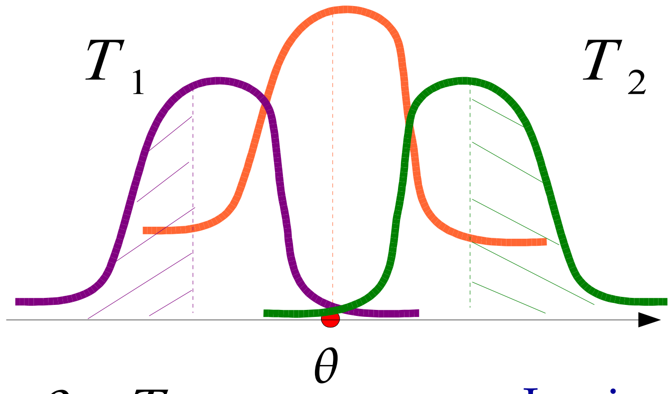
Intervalle de confiance <u>bilatéral</u>
$[T_1, T_2]$ <u>à risques asymétriques</u>

Intervalle de confiance d'un paramètre

- Soit un intervalle de confiance $[T_1, T_2]$ de niveau de confiance 1-α

$$T_1 = \varphi_1(T)$$
$$T_2 = \varphi_2(T)$$

T estimateur sans biais



Intervalle de confiance <u>bilatéral</u>
$[T_1, T_2]$ <u>à risques symétriques</u>

Le risque que $\theta < T_1$
est égal à $\alpha/2$

Le risque que $\theta > T_2$
est égal à $\alpha/2$

Propriétés de l'intervalle de confiance d'un paramètre

- Soit un intervalle de confiance $[T_1, T_2]$
  de niveau de confiance 1-α

  - si le niveau de confiance 1-α augmente,
    alors le risque statistique α doit diminuer

    => écartement de l'intervalle de confiance

  - pour un niveau de confiance 1-α fixé
    si la taille de l'échantillon *n* augmente,
    alors $Var(T)$ diminue

    => T devient plus précis
    => réduction de l'intervalle de confiance

Exemple

- Une enquête sur 400 PC de ce modèle donne un benchmark moyen de 0.23 ?=0,01
- Que décidez-vous ?

- Modélisation : $B \sim N(\mu, \sigma^2 = 10^{-2})$
- Enquête : $\overline{B} \sim N(\mu, \sigma^2/400)$

$$\frac{\overline{B}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

  - fonction de répartition de la loi *N* : $P(-2.58 \leq \frac{\overline{B}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 2.58) = 0.99$

Exemple

- $$P\left(-2.58 \le \frac{\overline{B}-\mu}{\frac{\sigma}{\sqrt{n}}} \le 2.58\right) = 0.99$$ $$P\left(-2.58\frac{\sigma}{\sqrt{n}} \le \overline{B}-\mu \le 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99$$

- on cherche un intervalle de confiance sur μ

$$P\left(\overline{B}-2.58\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{B}+2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99$$

$$\boxed{T_1 = \overline{B}-2.58\frac{\sigma}{\sqrt{n}} \qquad T_2 = \overline{B}+2.58\frac{\sigma}{\sqrt{n}}}$$

- $[T_1, T_2]$ intervalle aléatoire : les bornes sont fonction de l'estimateur $T=\overline{B}$
  - application numérique : intervalle = *[0.217 , 0.243] ... on achète !*

« Fiche recette » pour l'estimation d'un paramètre *par intervalle de confiance*

- Choisir un estimateur T de θ - le meilleur possible - dont on connaît la loi de probabilité en fonction de θ

- Déterminer la fonction pivotale $f(T,\theta)$, dont la loi de probabilité, notée P, ne dépend plus de θ (P = loi pivot)

- Déterminer l_1 et l_2 tels que $P(l_1 \leqslant f(T,\theta) \leqslant l_2) = 1-\alpha$

- En déduire, si possible, $P(T_1 \leqslant \theta \leqslant T_2) = 1-\alpha$
  où T_i sont 2 statistiques fonction de T

  $T_1 = \varphi_1(T)$ et $T_2 = \varphi_2(T)$

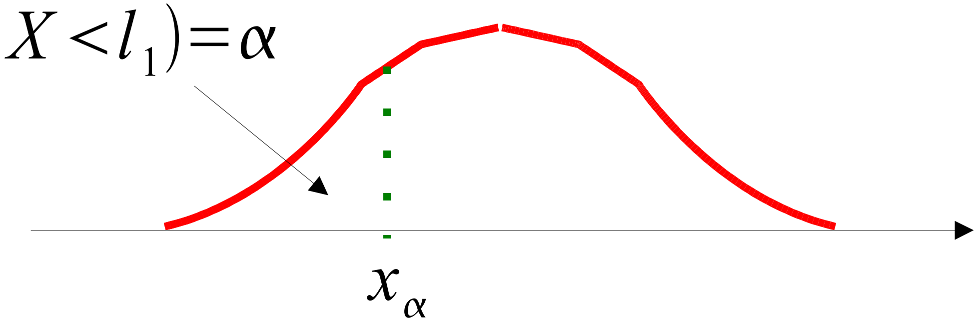
Déterminer les bornes l_1 et l_2 de la loi pivot P

- A partir de la fonction de répartition de la loi pivot P de la variable aléatoire f(T,θ)
  - Intervalle de confiance unilatéral

$$P(l_1 \leqslant f(T,\theta)) = 1-\alpha \quad \Rightarrow \quad P(f(T,\theta) \leqslant l_1) = \alpha$$

$$X = f(T,\theta) \sim P$$

$$P(X < l_1) = \alpha$$



$$\boxed{l_1 = F_P^{-1}(\alpha)}$$

l_1 = fractile d'ordre α de la loi pivot P

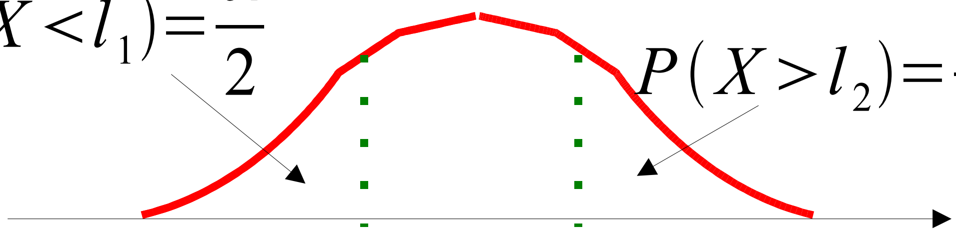
Déterminer les bornes l_1 et l_2 de la loi pivot P

- A partir de la fonction de répartition de la loi pivot P de la variable aléatoire f(T,θ)
  - Intervalle de confiance bilatéral à risques symétriques

$$X = f(T,\theta) \sim P$$

$$P(X < l_1) = \frac{\alpha}{2}$$

$$P(X > l_2) = \frac{\alpha}{2}$$



$x_{\frac{\alpha}{2}}$ $x_{1-\frac{\alpha}{2}}$

$$\boxed{l_1 = F_P^{-1}\left(\frac{\alpha}{2}\right)}$$

l_1 = fractile d'ordre α/2 de la loi pivot P

$$\boxed{l_2 = F_P^{-1}\left(1 - \frac{\alpha}{2}\right)}$$

l_2 = fractile d'ordre 1 - α/2 de la loi pivot P

Quelques lois de fonctions pivotales usuelles

- $f(T = S^2, \theta = \sigma^2)$ $\qquad \chi_p^2$ loi du chi-deux à *p* d.d.l
  - (estimation par intervalle de la variance d'une loi normale)

- $f(T = \bar{X}, \theta = \mu)$ $\qquad \tau_p$ loi de student à *p* d.d.l
  - (estimation par intervalle de l'esperance d'une loi normale, variance inconnue)

$f(T\ ,\ \theta)$ suit $N(0,1)$

(estimation par intervalle d'un paramètre inconnu θ, lorsque $E(X)=f(\theta)$ et $Var(X)=g(\theta)$ inconnues)

Loi de la fonction pivotale $f(T=S^2, \theta=\sigma^2)$

- χ^2_p loi du chi-deux à *p* degrés de liberté

$$\boxed{\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}} \qquad \boxed{(n-1)\frac{S^{*2}}{\sigma^2} \sim \chi^2_{n-1}}$$

- Propriétés :
  - $E(\chi^2_p) = p$
  - $Var(\chi^2_p) = 2p$
  - si p>30, $\sqrt{2\chi^2_p} \sim N(\sqrt{2p-1}, 1)$

Loi du chi-deux χ^2_p à *p* degrés de liberté

- χ^2_p = loi de probabilité d'une *somme de* p *carrés de v.a. normales centrées réduites indépendantes*

- Pourquoi d.d.l.=*n-1* dans $\dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$?

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad\qquad \frac{nS^2}{\sigma^2} = \sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{\sigma^2} = \sum_{i=1}^{n} U_i^2$$

  - les U_i sont des v.a. centrées réduites, mais seulement *n-1* sont indépendantes car $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$

Loi de la fonction pivotale $f(T=\bar{X}, \theta=\mu)$

- τ_p loi de Student à *p* degrés de liberté

$$\boxed{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)} \quad \Rightarrow \quad \boxed{\frac{\bar{X}-\mu}{\sqrt{\frac{S^{*2}}{n}}} \sim \tau_{n-1}}$$

une seule inconnue

plusieurs inconnues

$\mu = E(X)$

$\sigma^2 = Var(X)$

S^{*2} estimateur de σ^2

- Propriétés
  - $E(\tau_p)=0$
  - $Var(\tau_p)=\frac{p}{p-2}$ $avec\ p>2$
  - si *p*>30, $\tau_p \rightarrow N(0, \frac{p}{(p-2)})$

Loi de Student τ_p à *p* degrés de liberté

- Soient
  - *U* une v.a. suivant une loi N(0,1)
  - V une v.a. suivant une loi χ^2_p

  alors $\tau_p = \dfrac{U}{\sqrt{\dfrac{V}{p}}}$ suit une loi de Student à *p* d.d.l

  - vérification pour la loi pivotale :

$$U = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \qquad V = (n-1)\frac{S^{*2}}{\sigma^2} \qquad (d.d.l = n-1) \qquad \frac{\overline{X} - \mu}{\sqrt{\dfrac{S^{*2}}{n}}} \sim \tau_{n-1}$$

Exemples

- ex1 : soit une v.a. $X \sim N(\mu, \sigma^2)$ avec μ inconnu et σ connu
  - proposer un estimateur pour le paramètre μ ;
  - déterminer l'intervalle de confiance bilatéral de niveau de confiance $1-\alpha$, puis celui unilatéral

Exemples

- ex2 : soit une v.a. $X \sim N(\mu, \sigma^2)$ avec μ et σ inconnus
  - proposer un estimateur pour le paramètre σ^2 ;
  - déterminer l'intervalle de confiance bilatéral de niveau de confiance $1-\alpha$

Exemples

- ex3 : soit une v.a. $X \sim N(\mu, \sigma^2)$ avec μ et σ inconnus
  - proposer un estimateur pour le paramètre μ ;
  - déterminer l'intervalle de confiance bilatéral de niveau de confiance $1-\alpha$