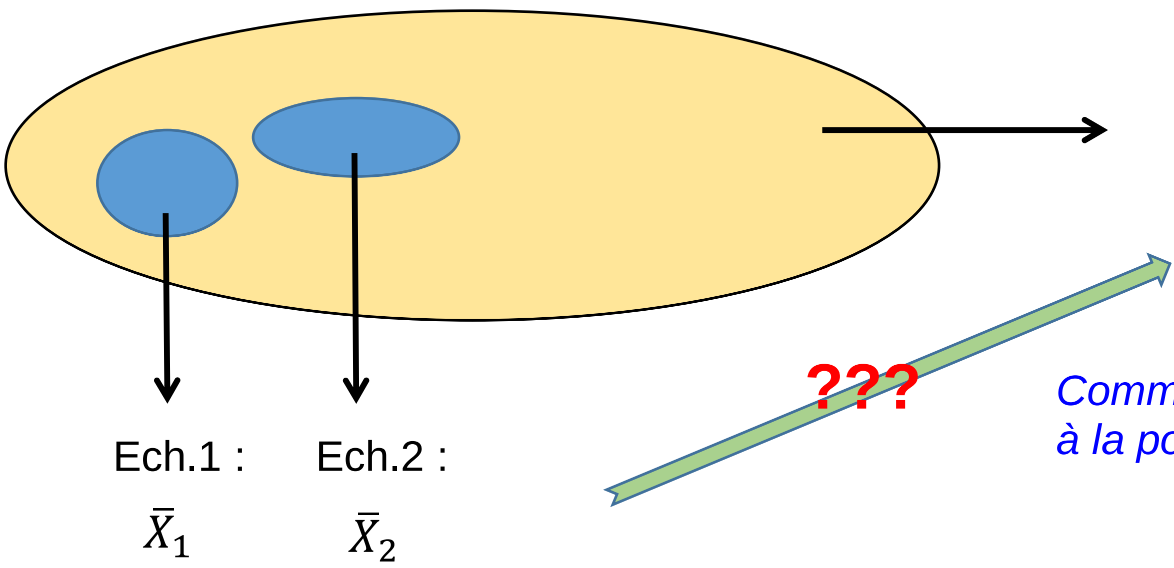
STA401

Statistique et Calcul des Probabilités

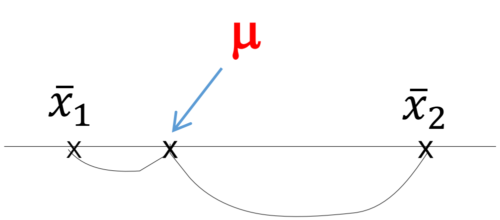
Responsable *: Carole Durand-Desprez*

CHAPITRE 3 : Estimation

1. Estimateurs – Maximum de vraisemblance :



Comment passer de l'échantillon à la population ?



Bonnes propriétés : qualités pour que les valeurs trouvées dans l'échantillon donnent une "bonne indication" sur le modèle de la population.
$\longrightarrow$ "bons" estimateurs

Mesure de l'écart $\longrightarrow$ Intervalle de confiance

Estimateur :

Soit X une v.a. de loi de probabilité P, soit θ un paramètre de cette loi (ex : moyenne, variance …). Soit ($X_1, \dots, X_n$) un échantillon de taille n.

Définition 1 : Un estimateur T_n de θ est une variable aléatoire fonction de l'échantillon.

Ex : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur de l'espérance, mais ce n'est pas le seul !!

Qualités d'un estimateur :

1) Convergent : $\lim_{n\to+\infty} P(|T_n - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$

2) Sans biais : $E(T_n - \theta) = 0 \quad ou \quad E(T_n) = \theta$

3) Efficace : $V(T_n)$ est la plus petite possible : $V(T_n) \leq V(T) \quad \forall T$ estimateur

Exemple : **Qualités de la moyenne empirique $\bar{X}$** :

- C' est un estimateur de E(X) ou μ. Voir chapitre précédent : $E(\bar{X}) = \mu \quad et \quad V(\bar{X}) = \sigma^2/n$
- Sans biais : $E(\bar{X} - \mu) = \frac{1}{n}\sum E(X_i) - E(\mu) = \mu - \mu = 0$
- Convergent : Bienaymé Tchebychev $\rightarrow$ $P(|\bar{X} - \mu| > \epsilon) \leq \frac{V(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$
- Variance asymptotiquement nulle : $V(\bar{X}) = \sigma^2/n \rightarrow 0$

Maximum de vraisemblance :

POUR INFORMATION

Soit X une v.a. de loi de densité f ; soit θ un paramètre de cette loi (ex : moyenne, variance …) ; soit ($x_1, \dots, x_n$) un échantillon de taille n.

Définition 2 : La vraisemblance de l'échantillon ($x_1, \dots, x_n$) pour θ est la fonction :

$$V(X,\theta) = f(x_{1,}\dots,x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$$

Définition 3 : L'estimateur du maximum de vraisemblance $\hat{\theta}$ est solution de :

$$Max_\theta\{V(X,\theta)\} \quad ou\ bien \quad Max_\theta\{\ln(V(X,\theta))\} = Max_\theta\left\{\sum_{i=1}^{n} \ln\left(f(x_{1,}\dots,x_n,\theta)\right)\right\}$$

[La fonction ln étant croissante, la transformation ne change pas le point réalisant le max]

Problème d'optimisation **:** Trouver l'estimateur du max de vraisemblance revient à trouver θ qui annule la dérivée en étant bien un maximum (dérivée seconde négative) :

$$\longrightarrow \quad \frac{\partial \ln(V(X,\theta))}{\partial\theta} = 0 \quad et \quad \frac{\partial^2 \ln(V(X,\theta))}{\partial\theta^2} \leq 0$$

On cherche la valeur de $\boldsymbol{\theta}$ *qui rend l'échantillon (* $x_1, \dots, x_n$ *) le plus probable.*

POUR INFORMATION

Exemple : **Estimateurs du maximum de vraisemblance de μ et σ^2 pour la loi $N(\mu\,;\sigma^2)$**

$$\longrightarrow \quad V(X,\mu) = \prod\left(\frac{1}{\sigma\sqrt{(2\pi)}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\right) = \left(\frac{1}{\sigma\sqrt{(2\pi)}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}$$

$$\longrightarrow \quad \ln\big(V(X,\mu)\big) = \ln\left(\frac{1}{\sigma\sqrt{(2\pi)}}\right)^n - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 = -n\ln(\sigma) + \ln\big(\sqrt{2\pi}\big) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

$$\longrightarrow \quad \frac{\partial\ln\big(V(X,\mu)\big)}{\partial\mu} = \frac{1}{\sigma^2}\sum(x_i-\mu) = \frac{1}{\sigma^2}\sum(x_i) - \frac{n\mu}{\sigma^2} = 0 \quad \Rightarrow \quad \boxed{\hat{\mu} = \frac{\sum x_i}{n}}$$

$\longrightarrow$ $\boxed{\hat{\mu} = \bar{X}}$ Estimateur du maximum de vraisemblance.
Il est convergent, sans biais et asymptotiquement efficace.

Pour μ connue :

$$\longrightarrow \quad \frac{\partial\ln\big(V(X,\mu)\big)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum(x_i-\mu)^2 = 0 \quad \Rightarrow \quad \boxed{\widehat{\sigma^2} = \frac{1}{n}\sum(x_i-\mu)^2}$$

$\longrightarrow$ Estimateur du maximum de vraisemblance.
Il est convergent, sans biais et asymptotiquement efficace.

POUR INFORMATION

Pour μ inconnue :

$$\rightarrow \quad \frac{\partial \ln\left(V(X,\mu)\right)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum (x_i-\mu)^2 = 0$$

$$\sum (x_i-\mu)^2 = \sum \left((x_i-\bar{x}) + (\bar{x}-\mu)\right)^2 = \sum (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 + 2(\bar{x}-\mu)\sum (x_i-\bar{x})$$

$$= \sum (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 + 2(\bar{x}-\mu) * 0 = \sum (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

$$\rightarrow \quad \frac{-n}{\sigma} + \frac{1}{\sigma^3}\left(\sum (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right) = 0 \quad \Rightarrow \quad \widehat{\sigma^2} = \frac{1}{n}\sum (x_i-\bar{x})^2$$

$$\rightarrow \quad \boxed{S^2 = \frac{1}{n}\sum (X_i-\bar{X})^2}$$

Estimateur du max. de vraisemblance de σ^2.
Il est convergent mais il a un biais.

$$E(S^2) = E\left(\frac{1}{n}\sum (X_i)^2 - \bar{X}^2\right) = \frac{1}{n}\sum E\left(X_i^2\right) - E(\bar{X}^2) = \frac{n}{n}(\sigma^2+\mu^2) - (\sigma^2/n + \mu^2) = \frac{n-1}{n}\sigma^2$$

donc $\quad E\left(\frac{n}{n-1}S^2\right) = \sigma^2$

$$\rightarrow \quad \boxed{S'^2 = \frac{1}{n-1}\sum (X_i-\bar{X})^2}$$

Estimateur convergent et sans biais de σ^2

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i$$

$$\widehat{\sigma^2} = S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$$

$$\hat{p} = F = \frac{K}{n}$$

Les estimateurs

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad suit \quad N(0;1)$$

$$\frac{\bar{X}-\mu}{S'/\sqrt{n}} = \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \quad suit \quad T_{n-1}$$

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \quad suit \quad \aleph^2_{n-1}$$

$$\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \quad suit \quad N(0;1) \longrightarrow \quad T.C.L.$$

Lois des estimateurs

$$\forall i,\ X_i \quad suit \quad N(\mu;\sigma^2) \quad alors$$

$$\bar{X} \quad suit \quad N\left(\mu;\frac{\sigma^2}{n}\right)$$

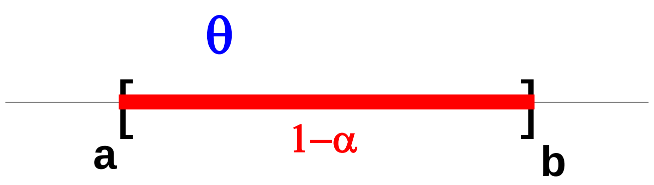
2. Intervalle de confiance

a/ Le principe :

Trouver un intervalle [a;b] qui contient le paramètre θ avec une probabilité $1 - \alpha$:

$P(a \leq \theta \leq b) = 1 - \alpha \quad \Longleftrightarrow \quad \theta \in [a; b]$ avec un niveau de confiance de $1 - \alpha$

→
- Estimer le paramètre
- Loi de l'estimateur
- Lecture de table
- Calcul des bornes a et b



α risque que θ ne soit pas dans l'intervalle $\Longleftrightarrow$ $1-\alpha$ confiance que θ y soit

Si α augmente alors l'intervalle diminue
Si α diminue alors l'intervalle augmente (L'intervalle de risque $\alpha = 0$ est : $\mathbb{R}$)

Amplitude de l'intervalle (ou longueur) : b-a

b/ Intervalle de confiance de μ avec σ^2 connue :

Trouver un intervalle [a;b] contenant μ avec une probabilité 1 - α : P(a < μ < b) = 1 – α

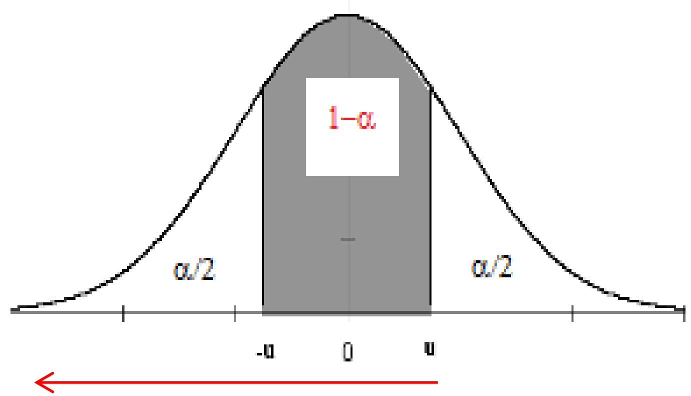
- $\hat{\mu} = \bar{X}$ $\quad et \quad$ $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $\quad suit \quad N(0;1)$ [σ^2 supposée connue]

- Construction :

$$1-\alpha = P(a \le \mu \le b) = P\left(\underbrace{\frac{\bar{X}-b}{\sigma/\sqrt{n}}}_{-u} \le \underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{N(0;1)} \le \underbrace{\frac{\bar{X}-a}{\sigma/\sqrt{n}}}_{u} \right)$$

$$\Leftrightarrow \quad 1-\alpha = P(-u \le Y \le u)$$

$$1-\frac{\alpha}{2} = P(Y \le u)$$



- Lire le quantile $u_{1-\alpha/2}$ sur la table $N(0;1)$

- Calcul de a et b :

$$u_{1-\alpha/2} = \frac{\bar{X}-a}{\sigma/\sqrt{n}} \quad donc \quad a = \bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$-u_{1-\alpha/2} = \frac{\bar{X}-b}{\sigma/\sqrt{n}} \quad donc \quad b = \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$I(\mu;\alpha) = [a;b]$$

c/ *Exemple* : X suit $N(\mu;\sigma^2)$ avec σ^2 connue, $\sigma^2=4$. Echantillon n=9, $\Sigma\, x_i = 900$.

- Moyenne empirique : $\bar{x} = 900/9 = 100$
- Moyenne estimée : $\hat{\mu} = \bar{x} = 100$
- Variance connue (pas empirique s², ni estimée s'²) : $\sigma^2 = 4$

- Intervalle de confiance de μ de niveau 95% :

α = 5%, σ^2 connue, donc $I(\mu;\alpha) = \left[\bar{X} - u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}} ; \bar{X} + u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$

$u_{0,975} = 1{,}96$ donc $I(\mu;0{,}05) = \left[100 \pm 1{,}96 \times \dfrac{2}{\sqrt{9}}\right] = [100 \pm 1{,}33] = [98{,}67;101{,}33]$

La moyenne de la population μ est comprise entre 98,67 et 101,33 avec 95% de confiance.

- La longueur de cet intervalle est : $\boxed{2u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}} = 2 \times 1{,}33 = 2{,}66$

Remarques : Si α augmente alors I(μ) diminue. Si n augmente alors I(μ) diminue

- Pour une longueur L=2, trouver n ? (ou trouver α ?)

a) $2u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}} = L \quad \Rightarrow \quad n = \left(2u_{1-\alpha/2}\dfrac{\sigma}{L}\right)^2$ Ici : n=16. [I diminue, n augmente.]

b) $u_{1-\alpha/2} = \dfrac{L\sqrt{n}}{2\sigma} \longrightarrow$ Lecture de table pour avoir *p* puis déduire α.

Ici, $u_{1-\alpha/2}$= 1,5 ; lecture de $N(0;1)$, p = 0,9332. Or, $p=1-\alpha/2$ $\Rightarrow$ $\alpha = 2(1-p)$. Donc : α = 13,36%

[I diminue, α augmente.]

d/ Construction de l'intervalle de la variance :

Trouver un intervalle [a;b] contenant σ^2 avec une probabilité $1 - \alpha$: $P(a < \sigma^2 < b) = 1 - \alpha$

- $\widehat{\sigma^2} = S'^2 \quad et \qquad \dfrac{nS^2}{\sigma^2} = \dfrac{(n-1)S'^2}{\sigma^2} \quad suit \quad \aleph^2_{n-1}$

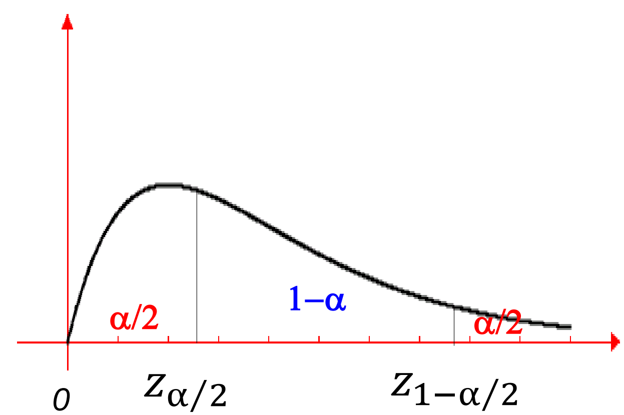
- $1 - \alpha = P(a \leq \sigma^2 \leq b) = P\left(\underbrace{\dfrac{nS^2}{b}}_{z_1} \leq \underbrace{\dfrac{nS^2}{\sigma^2}}_{\aleph^2_{n-1}} \leq \underbrace{\dfrac{nS^2}{a}}_{z_2} \right)$

$$1 - \alpha = P(z_1 \leq Y \leq z_2) \quad \Leftrightarrow \quad P(Y \leq z_1) = \alpha/2 \quad et \quad P(Y \leq z_2) = 1 - \alpha/2$$

- Lire $z_{\alpha/2}$ et $z_{1-\alpha/2}$ sur la table $\aleph^2_{n-1}$

- $z^{(n-1)}_{\alpha/2} = \dfrac{nS^2}{b} \quad et \quad z^{(n-1)}_{1-\alpha/2} = \dfrac{nS^2}{a} \quad donc$

$$\boxed{I(\sigma^2;\alpha) = \left[\frac{nS^2}{z^{(n-1)}_{1-\alpha/2}} \; ; \; \frac{nS^2}{z^{(n-1)}_{\alpha/2}} \right]}$$



Exemple : Pour un échantillon de taille n=10, on calcule s²=1,5. On prend α = 5%

$$z^{(n-1)}_{\alpha/2} = z^{9}_{0,025} = 2,7 \quad et \quad z^{(n-1)}_{1-\alpha/2} = z^{9}_{0,975} = 19,023$$

$$donc \quad I(\sigma^2; 0,05) = \left[\frac{15}{19,023}; \frac{15}{2,7} \right] = [0,7885; 5,555]$$

La variance entre 0,78 et 5,55 …. peu précis

e/ Tous les intervalles de confiance (à savoir retrouver) :

Si σ^2 connue, $u_{1-\alpha/2}$ est le quantile d'ordre 1-α/2 de la loi Normale *N(0;1) :*

$$I(\mu;\alpha) = \left[\bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \; ; \; \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

[moyenne]

Si σ^2 inconnue, $t_{1-\alpha/2}^{n-1}$ est le quantile d'ordre 1-α/2 de la loi Student T_{n-1} :

$$I(\mu;\alpha) = \left[\bar{X} - t_{1-\alpha/2}^{n-1}\frac{S'}{\sqrt{n}} \; ; \; \bar{X} + t_{1-\alpha/2}^{n-1}\frac{S'}{\sqrt{n}}\right]$$

[moyenne]

Si grand échantillon ($n \geq 50$) :

$$I(\mu;\alpha) = \left[\bar{X} - u_{1-\alpha/2}\frac{S}{\sqrt{n}} \; ; \; \bar{X} + u_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

[moyenne]

$$I(p;\alpha) = \left[F - u_{1-\alpha/2}\sqrt{\frac{F(1-F)}{n}} \; ; \; F + u_{1-\alpha/2}\sqrt{\frac{F(1-F)}{n}}\right]$$

[probabilité]

Si $z_{\alpha/2}^{(n-1)}$ et $z_{1-\alpha/2}^{(n-1)}$ sont les quantiles d'ordre α/2 et 1-α/2 de la loi du Khi-deux $\aleph_{n-1}^2$

$$I(\sigma^2;\alpha) = \left[\frac{nS^2}{z_{1-\alpha/2}^{(n-1)}} \; ; \; \frac{nS^2}{z_{\alpha/2}^{(n-1)}}\right]$$

[variance]