Estimation par intervalle de Confiance.

1- Rappel :

2- Fonction Pivotale : C'est une fct statistique qui permet de construire un intervalle de confiance pour un paramètre inconnue

Déf : On concidère $(X_1, \dots, X_n)$ un échantillon de taille n issue d'une variable aléatoire réelle de densité $f(\cdot, \theta)$ et $\hat{\theta}$ un estimateur de θ.

La fonction $\pi(\hat{\theta}, \theta)$ est une fct pivotale de θ ssi la loi de probabilité de π est indépendant de θ.

Exemples :

1-

$(X_1, \dots, X_n)$ iid $\mathcal{N}(m, \sigma^2)$; $f(x, m, \sigma^2)$

$\cdot\ \hat{m} = \bar{X} = \frac{1}{n} \sum x_i \hookrightarrow \mathcal{N}(m, \frac{\sigma^2}{n})$

$E(\hat{m}) = m$; $V(\hat{m}) = \frac{\sigma^2}{n}$

$\cdot\ (\bar{X} - m) \hookrightarrow \mathcal{N}(0, \frac{\sigma^2}{n})$

$\cdot\ \sqrt{n}\, \frac{\bar{X} - m}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$: $\pi(m, \hat{m})$

$\pi(m, \hat{m})$ est une fct pivotale pour m (σ^2 : inconnue)

si la variance est inconnue,

$\pi(m, \hat{m}) = \sqrt{n}\, \frac{\bar{X} - m}{\hat{\sigma}}$ avec $\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

$$\Rightarrow (n-1) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{\sigma^2} \hookrightarrow \chi^2_{(n-1)}$$

$$\Rightarrow \pi(m, \hat{m}) = \sqrt{n}\, \frac{\bar{X} - m}{\hat{\sigma}} = \sqrt{n}\, \frac{\bar{X} - m}{\sigma \sqrt{\frac{\hat{\sigma}^2}{\sigma^2} \frac{(n-1)^2}{n-1}}}$$

$$= \frac{\sqrt{n}\, \frac{\bar{X} - m}{\sigma}}{\sqrt{\frac{(n-1)\hat{\sigma}^2}{\sigma^2} / n-1}} = \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1}}} = T(n-1)$$

Remarque :

$\hat{p} \xrightarrow{\mathcal{L}} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

Fct Pivotale $\pi(\theta, \hat{\theta})$ hia fct chental3ou beha l'IC ta3 θ li houa inconnu, $\hat{\theta}$

$\sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} N(0,1)$: $\pi(p, \hat{p})$ est asymptotiquement pivotale.

2 -

$(X_1, \dots, X_n)$ iid (m, σ^2)

Ce qui est important, c'est que la fct pivotale soit définie en termes d'une combinaison linéaire de l'estimateur $\hat{\theta}$ et d'un ou plusieurs paramètres inconnus

$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

$\pi(\sigma^2, \hat{\sigma}^2) = (n-1) \frac{\hat{\sigma}^2}{\sigma^2} \hookrightarrow \chi^2_{(n-1)}$

3 - $(X_1, \dots, X_n) \hookrightarrow \mathcal{B}_0(p)$

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

$E(\hat{p}) = p$; $V(\hat{p}) = \frac{1}{n} p(1-p)$

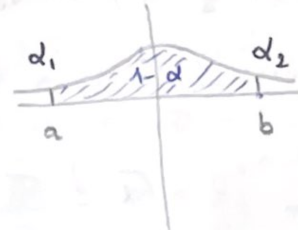
3 - Intervalle de Confiance :

On considère $(X_1, \dots, X_n)$ un échantillon iid issue d'un v.a.r de densité $f(\cdot, \theta)$; $\hat{\theta}$ est un estimateur de θ et $\pi(\hat{\theta}, \theta)$ une fct pivotale pour θ.

L'intervalle de confiance de θ de niveau $1-\alpha$ noté $IC_\theta^{1-\alpha}$ est tq : $P(a < \pi(\theta, \hat{\theta}) < b) = 1 - \alpha$

où a et b des constantes à déterminer.

Exemples :



1- $(X_1, \dots, X_n)$ iid $\hookrightarrow \mathcal{N}(m, \sigma_0^2)$

$$\pi(m, \hat{m}) = \sqrt{n}\,\frac{\bar{X} - m}{\sigma_0} \hookrightarrow \mathcal{N}(0,1)$$

$$IC_m^{1-\alpha} : P\left(a < \sqrt{n}\,\frac{\bar{X}-m}{\sigma_0} < b\right) = 1 - \alpha$$

F : fct de répartition

$\Phi(a) = \alpha_1$

$\Phi(b) = 1 - \alpha_2 = 1 - \alpha + \alpha_1$

si la distribution est symétrique : $a = -b$, $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$

$F(-x) = 1 - F(x)$

$$\Rightarrow P\left(-b < \sqrt{n}\,\frac{\bar{X}-m}{\sigma_0} < b\right) = 1 - \alpha$$

$$\Rightarrow P\left(|\mathcal{N}(0,1)| < b\right) = 1 - \alpha$$

$$\Rightarrow 2\Phi(b) - 1 = 1 - \alpha$$

$$\Rightarrow \Phi(b) = 1 - \frac{\alpha}{2}$$

$$\Rightarrow b = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = z_{1-\frac{\alpha}{2}}$$

$$\left(\begin{array}{l} \alpha = 5\% \Rightarrow 1 - \frac{\alpha}{2} = 0{,}975 \\ \qquad\quad \Rightarrow b = 1{,}96 \end{array}\right)$$

$$-z_{1-\frac{\alpha}{2}} < \sqrt{n}\,\frac{\bar{X}-m}{\sigma_0} < z_{1-\frac{\alpha}{2}}$$

$$\Rightarrow -z_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}} + \bar{X} < m < z_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}} + \bar{X}$$

$$\Rightarrow IC_m^{1-\alpha} = \bar{X} \pm \frac{\sigma_0}{\sqrt{n}}\, z_{1-\frac{\alpha}{2}}$$

invnormale $(1 - \frac{\alpha}{2})$

$F^{-1}(1 - \frac{\alpha}{2})$

2 -

$$(X_1, \dots, X_n) \hookrightarrow \mathcal{N}(m, \sigma^2)$$

$$T = \sqrt{n}\,\frac{\bar{X} - m}{\hat{\sigma}} \hookrightarrow st(n-1)$$

$$St^{-1}_{(n-1)}\left(1 - \frac{\alpha}{2}\right)$$

$$IC_m^{1-\alpha} = \bar{X} \pm \frac{\hat{\sigma}}{\sqrt{n}}\, t^{(n-1)}_{1-\frac{\alpha}{2}}$$

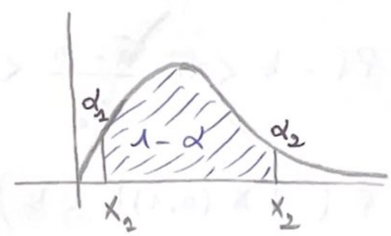
$n = 20$, $\bar{X} = 12$

$\alpha = 5\%$, $t_{1-\frac{\alpha}{2}} = 2{,}09$, $\hat{\sigma} = 1$

$$IC_m^{1-\alpha = 95\%} = 12 \pm \frac{1}{\sqrt{20}} \times 2{,}09$$

3 -

$$(X_1, \dots, X_n) \rightarrow \mathcal{N}(m, \sigma^2)$$



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

$$\pi(\sigma^2, \hat{\sigma}^2) = (n-1)\frac{\hat{\sigma}^2}{\sigma^2} \rightarrow \chi^2_{(n-1)}$$

$$IC_{\sigma^2}^{1-\alpha} : \quad P\left(X_1 < n-1\,\frac{\hat{\sigma}^2}{\sigma^2} < X_2\right) = 1 - \alpha \quad ; \quad \alpha_1 + \alpha_2 = \alpha$$

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2}$$

$$P(\chi^2_{(n-1)} < X_1) = \alpha_1 = \frac{\alpha}{2}$$

$$P(\chi^2_{(n-1)} > X_2) = \alpha_2 = \frac{\alpha}{2}$$

3- Intervalle de confiance pour le rapport des variances.

On considère 2 échantillons $(X_1, \dots, X_n)$; $(Y_1, \dots, Y_n)$ iid issus respectivement des v.a.r X et Y d'esperence m_1 et m_2 et de variance σ_1^2 et σ_2^2. On considère le rapport $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$.

Construire un intervalle de confiance de niveau $(1-\alpha)$ pour λ.

Procédure :

- fonction pivotale : $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$, $\hat{\lambda} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

$$\hat{\sigma}_2^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$$

$$\rightarrow (n_1 - 1) \frac{\hat{\sigma}_1^2}{\sigma_1^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma_1^2} \longrightarrow \chi^2_{(n-1)}$$

$$\rightarrow (n_2 - 1) \frac{\hat{\sigma}_2^2}{\sigma_2^2} = \frac{\sum (Y_i - \bar{Y})^2}{\sigma_2^2} \longrightarrow \chi^2_{(n-1)}$$

$$\hat{\lambda} = \frac{(n_1-1)\frac{\hat{\sigma}_1^2}{\sigma_1^2} / (n_1-1)}{(n_2-1)\frac{\hat{\sigma}_2^2}{\sigma_2^2} / (n_2-1)} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

Fisher - snédécor

$$F = \frac{(n_1-1)\frac{\hat{\sigma}_1^2}{\sigma_1^2} / (n-1)}{(n_2-1)\frac{\hat{\sigma}_2^2}{\sigma_2^2} / (n-1)} = \frac{\chi^2_{(n_1-1)} / (n_1-1)}{\chi^2_{(n_2-1)} / (n_2-1)} \longrightarrow F(n_1-1, n_2-1)$$

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \longrightarrow F(n_1-1, n_2-1)$$ une fct pivotale pour $\frac{\sigma_1^2}{\sigma_2^2} = \lambda$

$P(F_1 < F < F_2) = 1-\alpha$

$\Rightarrow P(F < F_1) = \alpha_1$

et

$P(F(n_1-1, n_2-1) < F_2) = 1-\alpha+\alpha_1$

$$F_1 < \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < F_2$$

$$\Rightarrow \frac{1}{F_2} < \frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} < \frac{1}{F_1}$$

$$\Rightarrow \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \cdot \frac{1}{F_2} < \lambda < \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \cdot \frac{1}{F_1} \Rightarrow CI_\lambda^{1-\alpha} = \left[\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \cdot \frac{1}{F_2}, \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \cdot \frac{1}{F_1}\right]$$

4- Intervalle de Confiance de différence d'espérence mathématique.

On considère 2 échantillons $(X_1, \dots, X_n)$ et $(Y_1, \dots, Y_n)$ issus resp de 2 v.a.r (normale) X et Y d'espérence m_1 et m_2 et de variance σ_1^2 et σ_2^2.

On s'intéresse à la différence $d = m_1 - m_2$

$\hat{d} = \hat{m}_1 - \hat{m}_2$.

$$\left.\begin{array}{l} \hat{m}_1 = \bar{X} \hookrightarrow \mathcal{N}\left(m_1, \frac{\sigma_1^2}{n_1}\right) \\ \hat{m}_2 = \bar{Y} \hookrightarrow \mathcal{N}\left(m_2, \frac{\sigma_2^2}{n_2}\right) \end{array}\right\} d = \hat{m}_1 - \hat{m}_2 \hookrightarrow \left(m_1 - m_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$H = \frac{\overbrace{\bar{X}-\bar{Y}}^{\hat{d}} - \overbrace{(m_1 - m_2)}^{d}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \longrightarrow \mathcal{N}(0,1)$$

1- σ_1^2 et σ_2^2 connus :

$$\Rightarrow IC_{m_1-m_2}^{1-\alpha} = (\bar{X}-\bar{Y}) \pm \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} \cdot z_{1-\frac{\alpha}{2}} \quad \left(z_{1-\frac{\alpha}{2}} = \phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)$$

f.t de rép de la loi normale

2- σ_1^2 et σ_2^2 inconnus et égales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

$$T = \frac{\bar{X} - \bar{Y} - (m_1 - m_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\mathcal{L} = \prod_{i=1}^{n_1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x_i - m_1)^2}{\sigma^2}\right) \cdot \prod_{j=1}^{n_2} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(y_i - m_2)^2}{\sigma^2}\right)$$

$$l = \sum_{i=1}^{n_1} \left[-\frac{1}{2} \frac{(x_i - m_1)^2}{\sigma^2} - \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2\right] + \sum_{j=1}^{n_2} \left[-\frac{1}{2} \frac{(y_i - m_2)^2}{\sigma^2} - \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2\right]$$

$$= -\frac{n_1 + n_2}{2} \ln \sigma^2 - \frac{n_1 + n_2}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left[\sum_{i=1}^{n_1} (x_i - m_1)^2 + \sum_{j=1}^{n_2} (y_j - m_2)\right]$$

<u>Cas 1</u> : $\frac{\partial l}{\partial m_1} = 0$; $\hat{m}_1 = \bar{X}$ et $\frac{\partial l}{\partial m_2} = 0$; $\hat{m}_2 = \bar{Y}$

$$\frac{\partial l}{\partial \sigma^2} = 0 \quad ; \quad -\frac{n_1 + n_2}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^2}\left(\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2\right) = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sigma^2}{n_1 + n_2} \left[\sum_i \frac{(x_i - \bar{x})^2}{\sigma^2} + \sum_j \frac{(y_j - \bar{y})^2}{\sigma^2}\right]$$

$$E(\hat{\sigma}^2_{m_2}) = \frac{\sigma^2}{n_1 + n_2} \left((n_1 - 1) + (n_2 - 1)\right)$$

$$\hat{\sigma}_2 = \frac{(n_1 - 1) \sum \frac{(x_i - \bar{x})^2}{n_1 - 1} + (n_2 - 1) \sum \frac{(y_i - \bar{y})^2}{n_2 - 1}}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)\hat{\sigma}_2^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}$$

$$\Rightarrow T = \frac{(\bar{x} - \bar{y}) - (m_1 - m_2)}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow St(n_1 + n_2 - 2)$$

$$IC^{1-\alpha}_{m_1 - m_2} : \bar{x} - \bar{y} \pm \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t^{1-\frac{\alpha}{2}}_{(n_1 + n_2 - 2)}$$

Si $n_1 = n_2 = n$ (Echantillon pairée).

$d_i = x_i - y_i \longrightarrow N\left(m_1 - m_2 \,,\, \frac{\sigma_1^2 + \sigma_2^2}{2\sigma^2} \right)$

$\hat{d} = \hat{m}_1 - \hat{m}_2 = \bar{x} - \bar{y}$

$\hat{\sigma}^2 = P$

$$\mathcal{L} = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp\left(\frac{-1}{2\, 2\sigma^2} \Sigma\left(d_i - \underbrace{(m_1 - m_2)}_{\theta}\right)^2 \right)$$

$$\hat{\theta} = \bar{d} = \hat{m}_1 - \hat{m}_2 = \bar{x} - \bar{y}$$

$$\frac{\partial l}{\partial \sigma^2} = 0 \Rightarrow \frac{-n}{2\hat{\sigma}^2} + \frac{1}{4\hat{\sigma}^4} \Sigma\left(d_i - (\bar{x} - \bar{y})\right)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{2n} \Sigma\left(d_i - (\bar{x} - \bar{y})\right)^2$$

⚠ $\Sigma\left(d_i - (\bar{x} - \bar{y})\right)^2 = \Sigma\left((x_i - \bar{x}) - (y_i - \bar{y})\right)^2$

$= \Sigma (x_i - \bar{x})^2 + \Sigma (y_i - y)^2 + 2\Sigma (x_i - \bar{x})(y_i - \bar{y})$

5 - Intervalle de confiance pour un écart type σ :

Si $n_1 = n_2 = n$. Echantillon pairés.

$$d_i = X_i - Y_i \sim N\left(m_1 - m_2, \frac{\sigma_1^2 + \sigma_2^2}{2\sigma^2}\right)$$

$$\bar{d} = \hat{m}_1 - \hat{m}_2 = \bar{X} - \bar{Y}$$

$\hat{\sigma}^2 = ?$

$$\mathcal{L} = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp -\frac{1}{2 \cdot 2\sigma^2} \sum (d_i - \underbrace{(m_1 - m_2)}_{\theta})^2$$

$$\hat{\theta} = \bar{d} = \hat{m}_1 - \hat{m}_2 = \bar{X} - \bar{Y}$$

$$\ell = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \cdot \ln\sqrt{\pi} - \frac{1}{4\sigma^2} \sum (d_i - (m_1 - m_2))^2$$

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \Rightarrow -\frac{n}{2\hat{\sigma}^2} + \frac{1}{4\hat{\sigma}^4} \sum (d_i - (\bar{x} - \bar{Y}))^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{2n} \sum (d_i - (\bar{x} - \bar{Y}))^2$$

⚠

$$\sum (d_i - (\bar{x} - \bar{Y}))^2 = \sum ((x_i - \bar{x}) - (Y_i - \bar{Y}))^2$$

$$= \sum (x_i - \bar{x})^2 + \sum (Y_i - \bar{Y})^2 - 2 \sum (x_i - \bar{x})(Y_i - \bar{Y})$$

5- Intervalle de Confiance pour un écart type σ:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\hat{\sigma}^2} \quad ? \text{ fonction pivotale}$$

Prop EMV

On considère un modèle statistique regulier.

$$\hat{\theta}_{MV} \xrightarrow{\mathcal{L}} N(\theta, I_n(\theta)^{-1})$$

$$(x_1 \text{ —— } x_n) \overset{iid}{\sim} N(m, \sigma^2)$$

$$I_n(\sigma) = -E\left(\frac{\partial^2 \ell}{\partial^2 \sigma^2}\right) =$$

$$\ell = n \ln \sigma - cte - \frac{1}{2\sigma^2} \sum (x_i - m)^2$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum (x_i - m)^2$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^2} \sum (x_i - m)^2$$

$$E\left(\frac{\partial^2 \ell}{\partial^2 \sigma^2}\right) = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} n \cdot \sigma^2$$

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

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

$$\hat{\sigma}_{mv} \xrightarrow{\mathcal{L}} N\left(\sigma, \frac{\sigma^2}{2n}\right)$$

$$\Pi = \sqrt{2n}\, \frac{\hat{\sigma}_{mv} - \sigma}{\sigma} \rightsquigarrow N(0,1)$$

$$IC_{\sigma}^{1-\alpha} = ?$$