Chapitre 8

Estimation

Estimateur

On appelle **estimateur** T d'un paramètre θ, toute fonction aléatoire des valeurs observées dans l'échantillon pour estimer θ

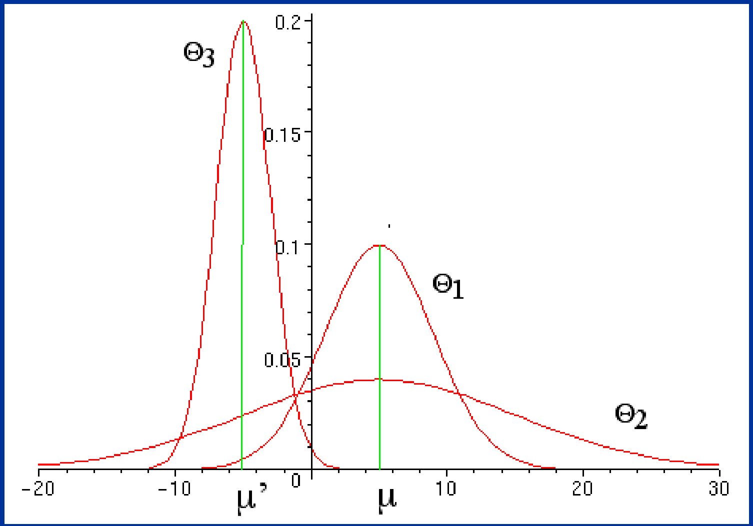
L'estimateur est une variable aléatoire dont la distribution est appelée **distribution d'échantillonnage** du paramètre θ

Estimateur

$E(T)=\theta$ Si $E(T)=\theta$, l'estimateur est sans biais $B(T) = E(T)-\theta=0$

$V(T)$ minimale

Un **estimateur** sans biais et de variance minimale est **efficace**



T_1 et T_2 sont sans biais

T_3 est biaisé

T_1 est plus efficace que T_2

Estimateur

Notion d'estimation

On appelle **estimation** de θ, la valeur numérique de l'estimateur.

Elle est notée $\hat{\theta}$

Estimateur

1. Estimateur de l'espérance μ d'une variable aléatoire

$$T = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\mathbf{E(\overline{X}) = \mu \;\; et \;\; V(\overline{X}) = \sigma^2 \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 = \frac{\sigma^2}{n}}$$

L'estimation de l'espérance est : $\hat{\theta} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$

Estimateur

2. Estimateur de la variance σ^2 d'une variable aléatoire

- μ est connue

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$E(T) = \sigma^2$$

L'estimation de la variance est : $\hat{\theta} = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

Estimateur

2. Estimateur de la variance σ^2 d'une variable aléatoire

- μ est inconnue

$$T = \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n}{n-1} S^2$$

$$E(T) = \sigma^2$$

L'estimation de la variance est : $\hat{\sigma}^2 = \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{n}{n-1} s^2$

Estimateur

3. Estimateur d'une proportion

$$T = \frac{X}{n} = \Psi$$

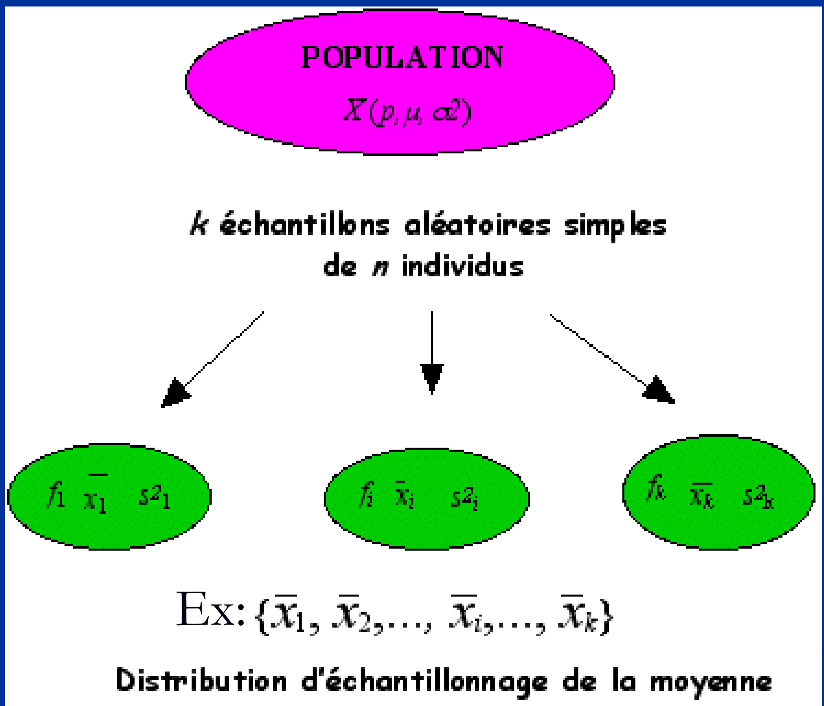
$$E(\Psi) = p \quad \text{et} \quad V(\Psi) = \frac{pq}{n}$$

L'estimation d'une proportion est : $\hat{\theta} = \hat{p} = \frac{x}{n}$

Distribution d'échantillonnage

1. Définition

- D'un point de vue empirique



La distribution associée à ces

k estimations constitue la

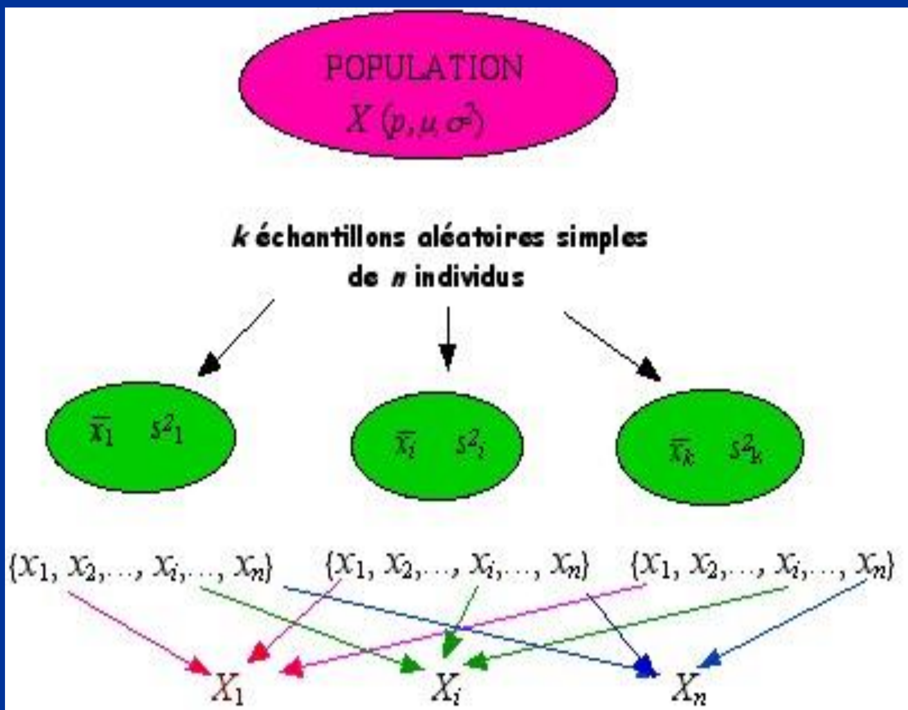
distribution d'échantillonnage

de l'estimateur du paramètre θ

Distribution d'échantillonnage

1. Définition

- D'un point de vue théorique



Les n valeurs $x_1 \ldots x_i \ldots x_n$ peuvent être considérées comme des observations de **n variables aléatoires $X_1 \ldots X_i \ldots X_n$**

On peut définir une nouvelle variable aléatoire **$T = f(X_1, \ldots, X_i, \ldots X_n)$**

Exemple: $T = X_1 + \ldots + X_i + \ldots + X_n$

Distribution d'échantillonnage

2. Distribution d'échantillonnage de la moyenne

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{avec } E(\overline{X}) = \mu \ \text{ et } V(\overline{X}) = \frac{\sigma^2}{n}$$

$$\text{Si } \ X \sim N(\mu, \sigma) \ \text{ alors } \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\text{Si } \ X \sim \text{loi quelconque } \ \text{ alors } \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ si n grand } (n \geq 30)$$

$$\text{Si } \ X \sim \text{loi quelconque } \ \text{ alors } \overline{X} \sim \text{? si n petit } (n < 30)$$

Distribution d'échantillonnage

3. Distribution d'échantillonnage d'une proportion

$$\Psi = \frac{X}{n} \quad \text{avec} \quad E(\Psi) = p \quad \text{et} \quad V(\Psi) = \frac{pq}{n}$$

$$\text{Si } X \sim B(n, p) \text{ alors } \Psi \sim N\left(p, \sqrt{\frac{pq}{n}}\right) \text{ si n grand } (n \geq 30) \text{ et } np > 5$$

$$\text{Si } X \sim B(n, p) \text{ alors } \Psi \sim \text{ ? si n petit } (n < 30)$$

Intervalle de confiance

1. Définition

Connaissant la distribution de l'estimateur T, on a une certaine confiance (1-α) qu'un intervalle autour de $\hat{\theta}$, contienne θ

Cet intervalle, noté $\left[\hat{\theta} - i ; \hat{\theta} + i\right]$ ou plus généralement $\left[T - i ; T + i\right]$ et appelé intervalle de confiance de θ, est tel que : $p(\hat{\theta} - i \leq \theta \leq \hat{\theta} + i) = 1 - \alpha$

L'intervalle de confiance indique la précision d'une estimation

Intervalle de confiance

2. Intervalle de confiance d'une moyenne

Il est noté $[\overline{X}-i;\overline{X}+i]$ tel que : $p(\overline{X}-i \leq \mu \leq \overline{X}+i) = 1-\alpha$

- **X~N(μ,σ) et σ² connue**

$$IC = \left[\overline{X}-\varepsilon_{\alpha}\frac{\sigma}{\sqrt{n}};\overline{X}+\varepsilon_{\alpha}\frac{\sigma}{\sqrt{n}}\right]$$

- **X~N(μ,σ) et σ² inconnue**

$$IC = \left[\overline{X}-t_{\alpha}^{(n-1)}\frac{\hat{\sigma}}{\sqrt{n}};\overline{X}+t_{\alpha}^{(n-1)}\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

Intervalle de confiance

2. Intervalle de confiance d'une moyenne

Il est noté $\left[\overline{X}-i;\overline{X}+i\right]$ tel que : $p(\overline{X}-i \leq \mu \leq \overline{X}+i) = 1-\alpha$

- X~ loi quelconque et n ≥ 30

$$\mathbf{IC} = \left[\overline{X}-\varepsilon_\alpha \frac{\hat{\sigma}}{\sqrt{n}};\overline{X}+\varepsilon_\alpha \frac{\hat{\sigma}}{\sqrt{n}}\right]$$

- X~ loi quelconque et n < 30

?

Intervalle de confiance

3. Intervalle de confiance d'une proportion

Il est noté $[\Psi - i ; \Psi + i]$ tel que : $p(\Psi - i \leq p \leq \Psi + i) = 1 - \alpha$

- $X \sim B(n,p)$ avec $\Psi = X/n$ et $n \geq 30$

$$IC = \left[\frac{X}{n} - \varepsilon_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}} ; \frac{X}{n} + \varepsilon_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

- $X \sim B(n,p)$ avec $\Psi = X/n$ et $n < 30$

? Recours à des lois exactes