Cours Statistiques
n° 4.

Estimation : obtenir une valeur (ou que faire se peut approchée) d'une caractéristique (ou d'une grandeur) statistique de la population en exploitant un échantillon.

Estimateur : θ de la caractéristique qu'on cherche

θ la valeur inconnue de la caractéristique de la population.

$\hat{\theta} = g(X_1, \dots, X_n)$ estimateur de θ

$\hat{\theta} = g(x_1, \dots, x_n)$ estimation de θ.

Moyenne μ_X de la population : $\mu_X = \lim\limits_{n \to \infty} \frac{1}{n} \sum\limits_{k=1}^{n} x_k$

$\theta : \mu_X \qquad \theta = \mu_X$

$\hat{\theta}_n = g(X_1, \dots X_n) = \frac{1}{n} \sum\limits_{k=1}^{n} X_k$ estimateur de la v.a

$\hat{\theta} = g(x_1 \dots x_n) = \frac{1}{n} \sum\limits_{k=1}^{n} x_k$

Critères de qualité de l'estimateur

1° Estimateur sans biais

$E(\hat{\theta}) = \theta \qquad B_\theta(\hat{\theta}) = \hat{\theta} - \theta$

$E(\hat{\theta}_n) = \theta \qquad B_\theta(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta$

Si on prend plusieurs échantillons et on calcule

$\hat{\theta}_1, \dots, \hat{\theta}_N$

$\frac{1}{N} \sum\limits_{p=1}^{N} \hat{\theta}_p \xrightarrow[N \to \infty]{} \theta$

2° $V(\hat{\theta}_n)$ doit être minimale
$V(\hat{\theta})$ soit minimale :
Estimateur sans biais de variance minimale

3° Estimateur convergent
$\lim_{n \to \infty} V(\hat{\theta}_n) = 0$

Précision ou risque d'un estimateur
$R_\theta(\hat{\theta}_n) = E[(\hat{\theta}_n - \theta)^2]$
$= V(\hat{\theta}_n) + (E(\hat{\theta}_n) - \theta)^2$
$R_\theta(\hat{\theta}_n) = \underbrace{V(\hat{\theta}_n)}_{\text{minimale}} + \underbrace{B_\theta^2(\hat{\theta}_n)}_{0}$
Le risque est petit → la précision grande

Exemples d'estimateurs
Espérance E de la population $E(X) = \mu_X = \theta$
Espérance empirique $\overline{X} = m_X = \frac{1}{n}\sum_{k=1}^{n} X_k = \hat{\theta}$

tirage $(x_1 \ldots x_n)$ $\hat{\theta}_n = \overline{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$

$E(\overline{X}_n) = E(X) = \mu_X = \theta$ ⇒ sans biais
$V(\overline{X}_n) = \frac{\sigma_X^2}{n}$ variance minimale (démo en TD)

Variance $\theta = V(X) = \sigma_X^2$
1° La moyenne de la population est connue μ_X
$\hat{\theta} = T_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_X)^2$

$(x_1 \ldots x_n)$ $\hat{\theta} = t_n^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_X)^2$

$$E(\hat{\theta}) = E(T_n^2) = \theta = \sigma_X^2 \Rightarrow \text{sans biais}$$

$$V(\hat{\theta}) = V(T_n^2) = \frac{\mu_X^4 - \sigma_X^4}{n} \xrightarrow[n \to \infty]{} 0$$

2° La moyenne de la population est inconnue. On l'estime par $\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k$

$$\hat{\theta} = S_n^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X}_n)^2$$

$$(x_1 \dots x_n) : \hat{\theta} = S_n^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2$$

$$E(\hat{\theta}) = E(S_n^2) = \frac{n-1}{n} \sigma_X^2 \rightarrow \text{avec biais}$$

$$V(\hat{\theta}) = V(S_n^2) = \frac{n-1}{n^3} \left[(n-1) \mu_X^4 - (n-3) \sigma_X^4 \right]$$

Variance corrigée

$$\hat{\theta} = S_n^2 = \frac{1}{n-1} \sum (X_k - \bar{X})^2$$

$$E(\hat{\theta}) = E(S_n^2) = \sigma_X^2 \text{ sans biais}$$

$$V(\hat{\theta}) = V(S_n^2) = \frac{1}{n} \left(\mu_X^4 - \frac{n-3}{n-1} \sigma_X^4 \right)$$

Corrélation entre deux va idd X et Y

$$\hat{\theta} = r_n(X, Y) = \frac{\sum_{k=1}^{n} (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\sum_{k=1}^{n} (X_k - \bar{X}_n) \cdot \sum_{k=1}^{n} (Y_k - \bar{Y}_n)} \in [-1; 1]$$