

Théorie de l'estimation

GASMI Abdelkader

December 9, 2024

ESTIMATEUR

A partir de l'examen d'un échantillon, on essaye d'obtenir une information quantitative sur des paramètres de la population à estimer (en général m et σ).

1. L'estimation est dite *ponctuelle* si on estime un paramètre de la population par un seul nombre.
2. L'estimation est dite *par intervalle* si on estime un paramètre inconnu θ par une construction d'un intervalle $]a, b[$ qui contient ce paramètre avec une probabilité α fixée:

$$P(a < \theta < b) = \alpha$$

a et b sont dite limites de confiance et α seuil (ou niveau) de confiance.

T est un estimateur de θ si T converge en moyenne quadratique vers θ :

$$E(T) \longrightarrow \theta$$

$$V(T) \longrightarrow 0 \quad \text{quand} \quad n \longrightarrow \infty$$

Un estimateur T de θ est dit sans biais si :

$$E(T) = \theta.$$

$\bar{X}$ est, par exemple un estimateur ponctuel non biaisé de m . Car

$$E(\bar{X}) = m$$

s^2 est un estimateur biaisé de σ^2 , car:

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

Si on la variance utilise $S'^2 = \frac{n-1}{n}S^2$ on a:

$$E(S'^2) = \sigma^2,$$

et S'^2 est un estimateur sans biais de σ^2 .

ESTIMATEURS PONCTUELS - MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

Soit un paramètre θ (en général μ ou σ) d'une population à estimer. Il faut trouver un estimateur T , à partir d'un échantillon; qui reflète le mieux le paramètre considéré.

► $E(T) = \theta$.

Par exemple $\bar{X}$ et S'^2 sont deux estimateurs sans biais de μ et σ^2 respectivement.

► $P(|\theta - T| \leq \epsilon) \longrightarrow 1$ quand $n \longrightarrow \infty$

$\bar{X}$ converge en probabilité vers μ .

On a $V(\bar{X}) \longrightarrow 0$ quand $n \longrightarrow \infty$

- La taille n de l'échantillon étant fixé, parmi les estimateurs T de θ on choisit celui dont la variance est la plus petite.

La méthode de maximum de vraisemblance consiste à choisir comme estimateur de θ la valeur particulière de θ qui maximise la fonction de vraisemblance:

$$L(x_1, x_2, \dots, x_n, \theta).$$

l'estimateur $L(x_1, x_2, \dots, x_n)$ et la fonction des observations qui maximisent $L(x_1, x_2, \dots, x_n, \theta)$ par rapport à tous les θ possibles cette estimateur T et solutions de l'équation:

$$\frac{\partial L(x_1, x_2, \dots, x_n, \theta)}{\partial \theta}, \quad \text{ou} \quad \frac{\partial \ln L(x_1, x_2, \dots, x_n, \theta)}{\partial \theta}.$$

INTERVALLE DE CONFIANCE AVEC LA DISTRIBUTION NORMALE

Intervalle de confiance de la moyenne.

on veut estimer la moyenne m d'une population normale à l'aide d'un échantillon aléatoire. (Si la taille de l'échantillon n est telle que $n \geq 30$ la distribution d'échantillonnage de la moyenne est normale quelle que soit la distribution de la population)

$\bar{X}$ suit une loi normale $N(m, \sigma_{\bar{X}})$ avec:

► $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ (tirage non exhaustif) ou

► $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ (tirage exhaustif) N : la taille de la population.

On a $\frac{\bar{X} - m}{\sigma_{\bar{X}}}$ suit une loi normale $N(0, 1)$

On cherche un intervalle centre sur m avec une population égale à α :

$$P\left(\frac{|\bar{X} - m|}{\sigma_{\bar{X}}} \leq u_{\alpha}\right) = \alpha$$

donc $m \in]\bar{X} - u_{\alpha}\sigma_{\bar{X}}, \bar{X} + u_{\alpha}\sigma_{\bar{X}}[$ avec une probabilité égale à α
Dans le cas où σ (écart type de la population) est inconnue on utilise l'estimation S' ou S de σ
(L'échantillon étant par hypothèse de taille et $n > 30$)
Si $\alpha = 0.95$ on a $u_{\alpha} = 1,96$

Intervalle de confiance d'une fréquence

Dans une population on estime que la proportion p d'un certain caractère à partir de la proportion F trouvée dans un échantillon de taille n

On a:

$$p \in]f - u_{\alpha}\sigma_f, f + u_{\alpha}\sigma_f[$$

avec

$$\sigma_f = \sqrt{\frac{p(1-p)}{n}} \quad (\text{tirage non exhaustif})$$

ou

$$\sigma_f = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \quad (\text{tirage exhaustif})$$

N : la taille de la population.

Si n est grand on estime p par f pour le calcul des limites de confiance.

Ces résultats sont dus au fait que Z et le nombre d'éléments de l'échantillon qui présentent le caractère étudié on a :

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

Si la taille de l'échantillon est assez grande, Z suit une loi normale ainsi que $f = \frac{Z}{n}$ suit une loi normale $N(\mu_f, \sigma_f)$ avec :

$$\mu_f = p \quad \text{et} \quad \sigma_f = \sqrt{\frac{p(1-p)}{n}} \quad (\text{tirage non exhaustif})$$

d'où $\frac{f - \mu_f}{\sigma_f} = \frac{f - p}{\sigma_f}$ suit une loi normale $N(0, 1)$

et $p \in]f - u_\alpha \sigma_f, f + u_\alpha \sigma_f[$

f étant un estimateur sans biais de p

$$\left(E(f) = \mu_f = E\left(\frac{Z}{n}\right) = \frac{1}{n}E(Z) = \frac{1}{n}np = p \right)$$

On estime p par f dans le calcul des limites de confiance d'où:

$$p \in \left[f - u_\alpha \sqrt{\frac{f(1-f)}{n}}, f + u_\alpha \sqrt{\frac{f(1-f)}{n}} \right]$$

Intervalle de confiance de différence des moyennes

On extrait un échantillon de taille n_i de la population P_i est ($i = 1, 2$) de moyenne m_i et d'ecart-type σ_i .

Si le tirage est non exhaustif ou on a:

$$m_1 - m_2 = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_{\bar{X}_1 - \bar{X}_2}$$

$$\mu_{\bar{X}_1 - \bar{X}_2} \in \left[\bar{X}_1 - \bar{X}_2 - u_\alpha \sigma_{\bar{X}_1 - \bar{X}_2}, \bar{X}_1 - \bar{X}_2 + u_\alpha \sigma_{\bar{X}_1 - \bar{X}_2} \right]$$

avec

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$