PROCESSUS STOCHASTIQUE ESTIMATION

PR. M. DJAMEL MOUSS

Laboratoire d'Automatique et Productique LAP

Genie Industriel

ESTIMATION

INTRODUCTION

Pour bien comprendre le sens du mot "estimation", nous allons citer des définitions différentes

Définition / (01) : Action d'estimer, de déterminer une valeur.

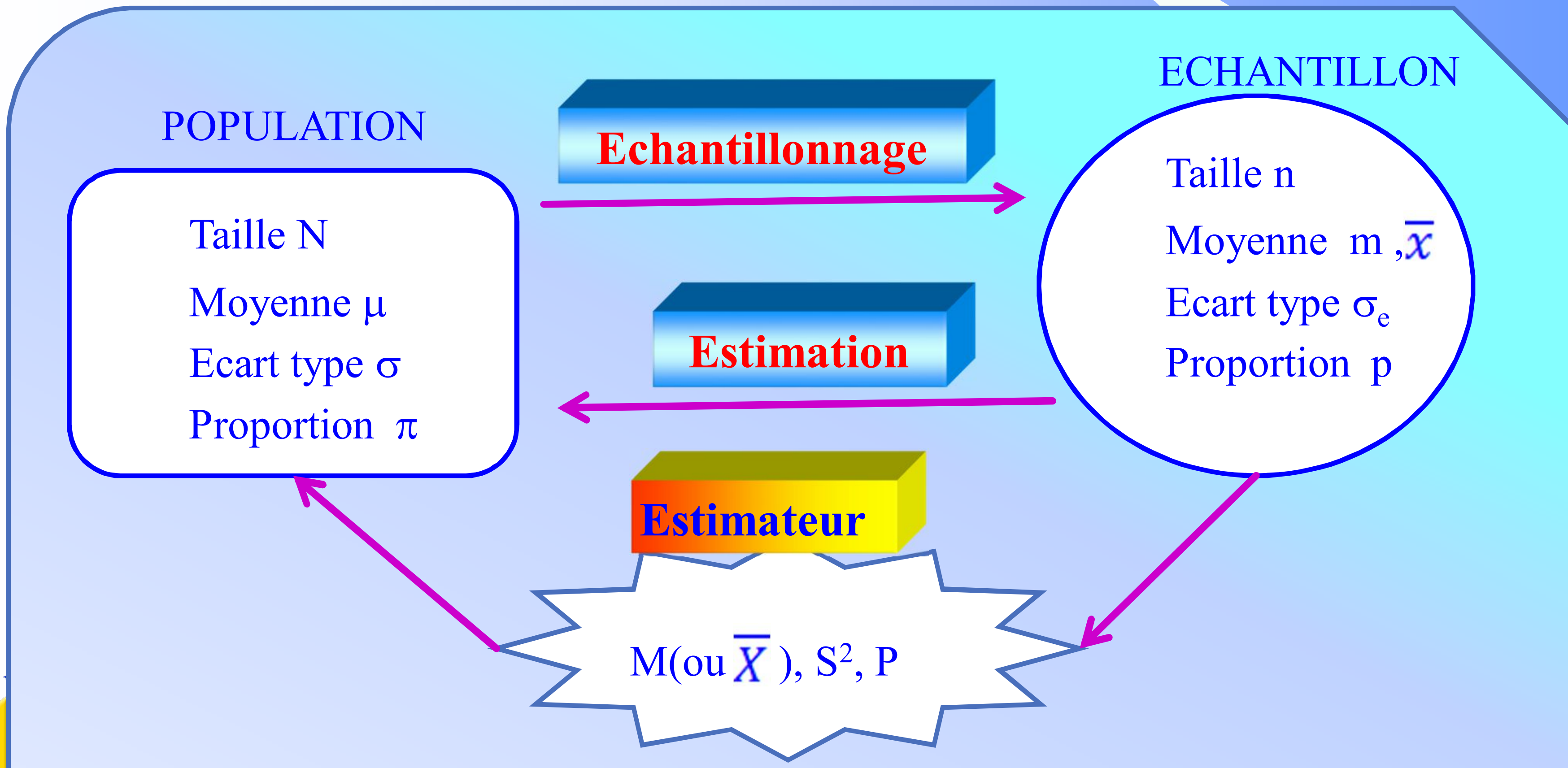
Définition / (02) : Recherche de la valeur d'un ou de plusieurs paramètres d'une loi statistique à partir d'observations ou de sondages sur un ou plusieurs échantillons d'une population

Définition / (03) : Une estimation est une valeur particulière prise par un estimateur.

Définition / (04) : Estimer le paramètre θ consiste à donner une valeur approchée à ce paramètre à partir d'un sondage de la population.

DEFINITION

L’estimation est le procédé par lequel on détermine les valeurs inconnues des paramètres de la population à partir des données de l’échantillon. Pour cela, on utilise des distributions théoriques, c’est à dire des variables aléatoires dont on connait les lois de probabilité.



Estimateur : Un estimateur est une statistique permettant d'évaluer un paramètre inconnu relatif à une loi de probabilité (comme son espérance ou sa variance). Il peut par exemple servir à estimer certaines caractéristiques d'une population totale à partir de données obtenues sur un échantillon.

Soient $X_1, X_2, \ldots, X_i, , \ldots, X_n$ n réalisations indépendantes de la variable aléatoire X (discrète ou continue) et θ un paramètre associé à la loi de probabilité suivi par X, un estimateur du paramètre θ est une variable aléatoire Θ fonction des X_i $\Theta = f(X_1, X_2, \ldots, X_i, , \ldots, X_n,)$

Laboratoire d'Automatique et Productique LAP

Estimateur et estimation :

X ,
$\Theta = E(X)$?

$(x_1, \ldots\ldots\ldots, x_n)$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\bar{x}$ est une approximation ou estimation de Θ

X ,
$\Theta = E(X)$?

$(x_1, \ldots\ldots\ldots, x_n)$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$(x'_1, \ldots\ldots\ldots, x'_n)$

$$\overline{x'} = \frac{1}{n}\sum_{i=1}^{n} x'_i$$

Les deux ***estimations*** $\bar{x}$ et $\overline{x'}$ de θ sont deux réalisation de la statistique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ appelée ***estimateur*** de θ

L’estimation d’un paramètre inconnu, noté θ est fonction des observations résultant d’un échantillonnage aléatoire simple de la population.
L’estimateur est donc une nouvelle variable aléatoire construite à partir des données expérimentales et dont la valeur se rapproche du paramètre que l’on cherche à connaître.

L’estimateur de θ est une variable aléatoire dont la distribution de probabilité s’appelle la ***distribution d’échantillonnage*** du paramètre Θ . L’estimateur admet donc une espérance une espérance $E(\Theta)$ et une variance $V(\Theta)$

Propriété d'un estimateur

Convergence

L'estimateur Q doit tendre vers la valeur réelle du paramètre q lorsque le nombre d'individus étudié augmente. On dit que ***« l'Estimateur est convergent »*** .

$$\forall \varepsilon > 0 \;\; P(|\Theta - \theta| > \varepsilon) \to 0 \;\; \text{Lorsque } n \to \infty$$

$$\lim_{n \to \infty} \Theta = \theta$$

Biais

Le biais d'un estimateur noté B(Θ) est la différence moyenne entre sa valeur et celle du paramètre qu'il estime. Il doit être nul pour avoir un bon estimateur

$$B(\Theta) = E(\Theta - \theta)$$

Un estimateur est dit ***sans biais*** si

$$B(\Theta) = 0 \Leftrightarrow E(\Theta) = \theta$$

Laboratoire d'Automatique et Productique LAP

L'erreur Quadratique Moyenne

L'erreur quadratique moyenne (EQM) d'un estimateur Θ est définit par

$$EQM(\Theta) = E[(\Theta - \theta)^2]$$

L'erreur quadratique moyenne est une mesure de la précision d'un estimateur

On a:

$$EQM(\Theta) = V(\Theta) + [Biais(\Theta)]^2$$

Soit Θ1 et Θ2 deux estimateurs:

Le problème est de savoir lequel des deux estimateur est le meilleur c'est-a dire quel est le plus efficace ?

Le plus efficaces des estimateur est celui ayant la plus petite EQM

Distribution d'échantillonnage

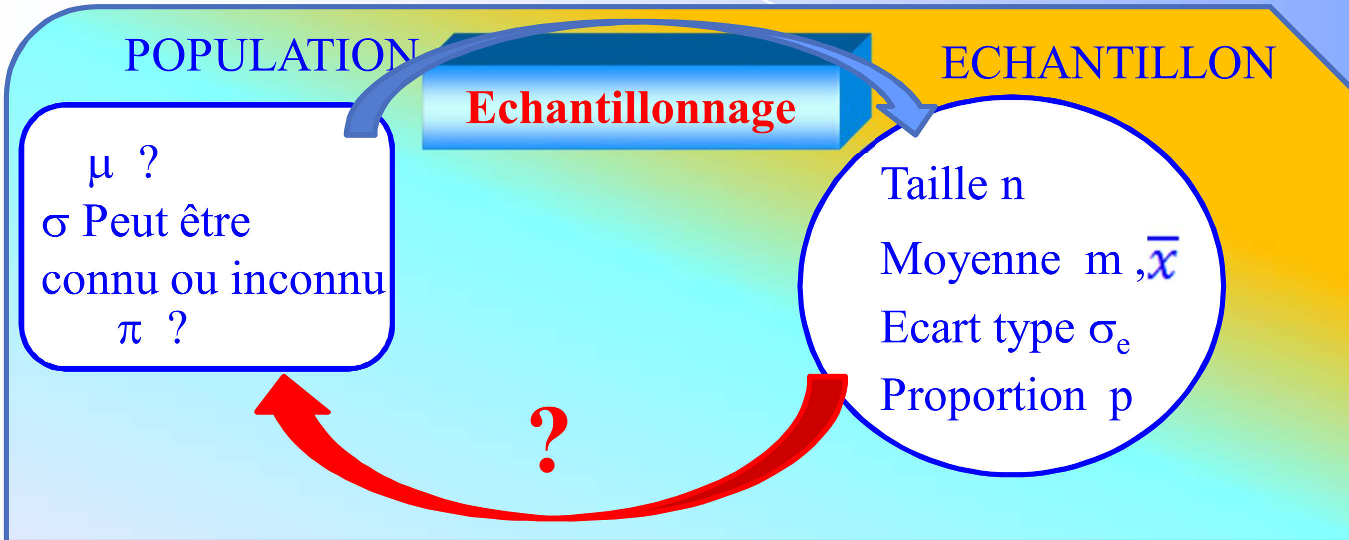
La distribution d'échantillonnage est l'étude de la de probabilité de l'échantillon en fonction de la distribution de la variable parente lorsque la taille de l'échantillon augmente.

Pour résoudre les problèmes d'estimation de paramètres inconnus, il faut tout d'abord étudier les distributions d'échantillonnage, c'est à dire la loi de probabilité suivie par l'estimateur.

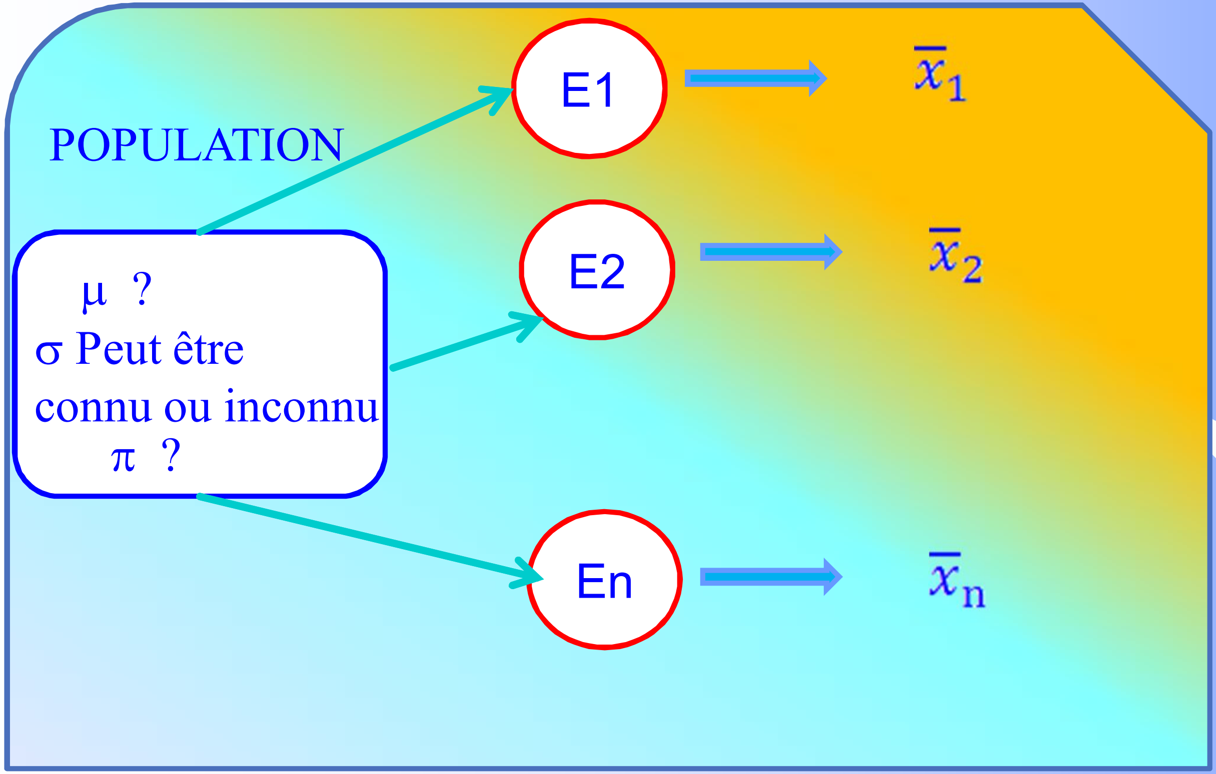
NOTATION

	Population	Echantillon
Taille	N	n
Moyenne	μ	m - $\bar{x}$
Ecart type	σ	σ_e - s
Variance	σ^2	σ_e^2 – s^2
Proportion	π	p

Distribution d’échantillonnage- Position du problème



Comment à partir des valeurs d’un caractère de l’échantillon estimer celui de la population



Les valeurs $\overline{x}_1$, $\overline{x}_2$,…$\overline{x}_n$ sont différentes. Laquelle prendre en considération ?

Exemple

Une population est constituée de 5 étudiants en statistique Leur professeur s'intéresse au temps hebdomadaire consacré au sport par chaque étudiant. Un questionnaire lui permet d'établir le tableau suivant.

Etudiants	Temps de Sport (H)
A	7
B	3
C	6
D	10
E	4
Total	30

En considérant ces 5 étudiants comme la population , on aura alors

$$\mu = 6.0$$
$$\sigma = 2.45$$

Le professeur décide de faire une étude statistique et choisit un échantillon de taille 3. Quelle relation existe-t-il entre cette moyenne d'échantillon et la véritable moyenne de la population ?

Toutes les possibilités sont regroupées dans le tableau ci-dessous.

N°	Echantillon	Valeur de l'Etude	Moyenne Echantillon
1	A-B-C	7-3-6	5.33
2	A-B-D	7-3-10	6.67
3	A-B-E	7-3-4	4.67
4	A-C-D	7-6-10	7.67
5	A-C-E	7-6-4	5.67
6	A-D-E	7-10-4	7
7	B-C-D	3-6-10	6.33
8	B-C-E	3-6-4	4.33
9	B-D-E	3-10-4	5.67
10	C-D-E	6-10-4	6.67
T			60.00

On remarque facilement la relation entre la moyenne de la population et la moyenne d'échantillonnage

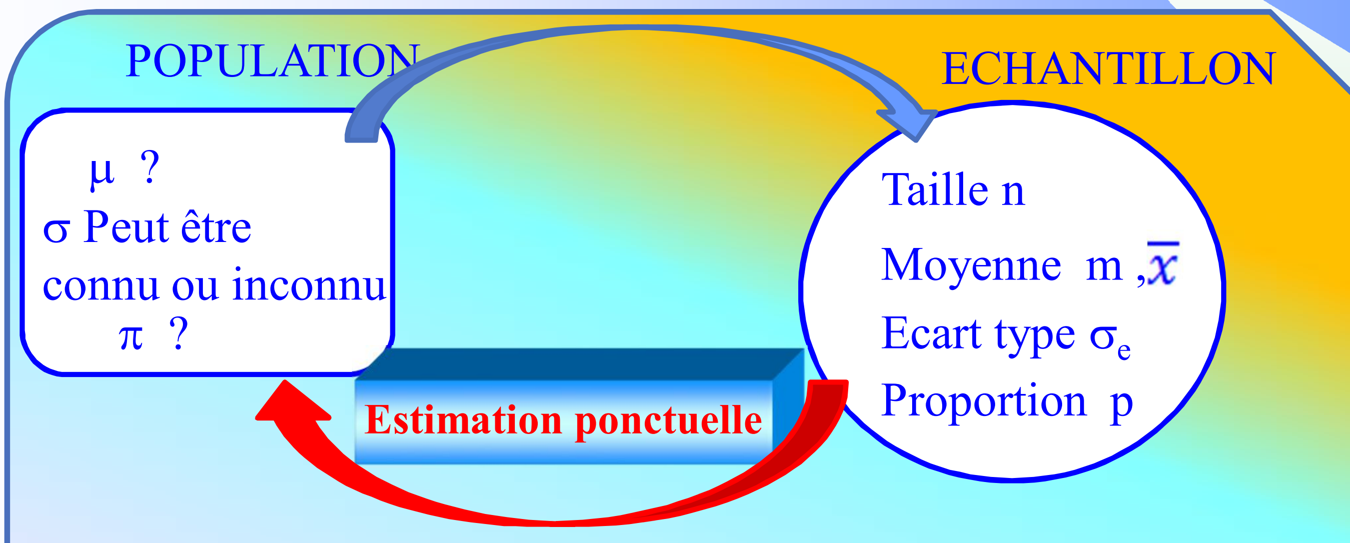
$$\mu_{\overline{X}} = \mu$$

Une synthèse est donnée sur le tableau suivant en considérant les 2 aspects qui sont le tirage avec remise et sans remise

	Echantillonnage Aléatoire Simple	
	Avec Remise	Sans Remise
Nombre d'échantillon	N^n	C_N^n
Distribution d'échantillonnage de la moyenne	$\mu_{\overline{X}} = \mu$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$	$\mu_{\overline{X}} = \mu$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
Distribution d'échantillonnage de la variance	σ^2	$\frac{n}{n-1} S^2$
Distribution d'échantillonnage de l'écart-type	σ	$\sqrt{\frac{n-1}{n}} \sigma$
Distribution d'échantillonnage de la fréquence	$E(f_n) = \pi = p$ $V(f_n) = \frac{p(1-p)}{n}$	$E(f_n) = \pi = p$ $V(f_n) = \frac{p(1-p)}{n} \frac{N-n}{N-1}$

Estimation Ponctuelle

A partir d'un échantillon, on essaye d'obtenir une information auantitative sur des paramètre de la population (en général la moyenne, la proportion et l'écart-type)



Moyenne $\mu = \bar{x}$

Ecart-Type $\sigma = \sqrt{\frac{n}{n-1}}\,\sigma_e$

Proportion $\pi = p$

$$\sigma_p = \sqrt{\frac{n}{n-1}}\sqrt{\frac{p(1-p)}{n}} \longrightarrow \sigma_p = \sqrt{\frac{p(1-p)}{n-1}} \qquad si\ n \leq 30$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \qquad si\ n > 30$$

Estimation par Intervalle de confiance

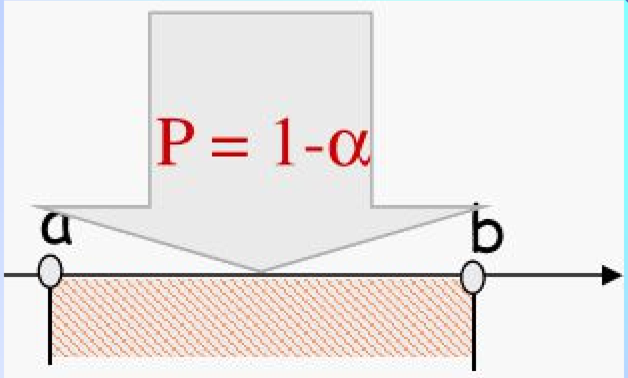
Plutôt que d'estimer un paramètre à l'aide d'un seul nombre, il arrive fréquemment que l'on fasse l'estimation en donnant un intervalle de valeurs, Un intervalle de confiance est défini de telle sorte que l'on puisse affirmer avec un degré de confiance fixé que le paramètre visé se trouve dans cet intervalle.

L'estimation par intervalle associe à un échantillon aléatoire un intervalle [a,b] qui recouvre le paramètre X à estimer avec une certaine probabilité

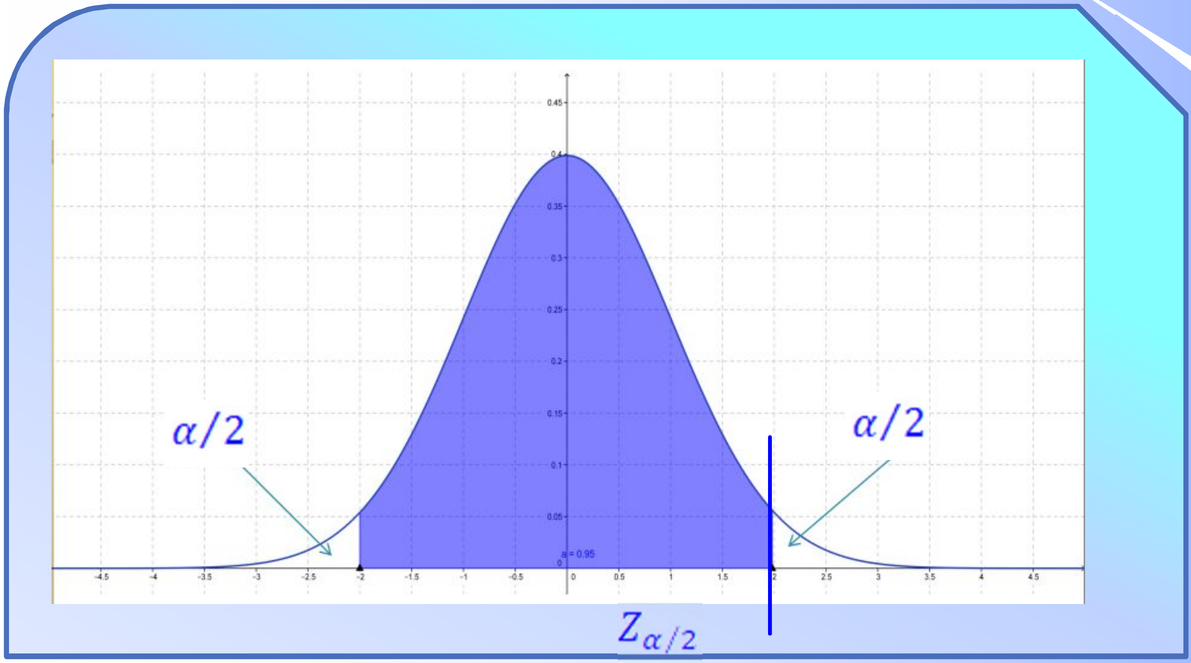
Cet intervalle est appelé Intervalle de confiance (noté IC_α) car la probabilité, que le paramètre à estimer X se trouve compris entre a et b, est égale à $1-\alpha$ qui est le degré de confiance

$$P(X \in [a,b]) = 1 - \alpha$$

α:est appelé risque; généralement égal à 0.05 ou 0.01



Le risque que le paramètre à estimer X n'appartienne pas l'intervalle Ica $P(X \notin [a, b]) = \alpha$ et comme peut être supérieur à b ou inférieur à a, nous allons diviser ce risque a des 2 cotés de l'intervalle



$Z_{\alpha/2}$ est obtenu par lecture inverse de la table de la loi Normale

Estimation d'une Moyenne par IC

Grand échantillon (n ≥ 30)

Premier Cas: σ^2 Variance de la population connue

$$IC_\alpha : \mu \in \left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad , \quad \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Deuxième Cas: σ^2 Variance de la population inconnue

$$IC_\alpha : \mu \in \left[\bar{x} - Z_{\alpha/2}\frac{S}{\sqrt{n}} \quad , \quad \bar{x} + Z_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Avec

$$S = \sqrt{\frac{n}{n-1}}\,\sigma_e$$

$Z_{\alpha/2}$ est obtenu à l'aide la table de la loi normale centrée réduite

Estimation d'une Moyenne par IC

Petit échantillon (n < 30)

Premier Cas: σ^2 Variance de la population connue

$$IC_\alpha : \mu \in \left[\bar{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad , \quad \bar{x} + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Deuxième Cas: σ^2 Variance de la population inconnue

$$IC_\alpha : \mu \in \left[\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} \quad , \quad \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Avec

$$S = \sqrt{\frac{n}{n-1}}\,\sigma_e$$

$t_{\alpha/2}$ est obtenu à l'aide la table de la loi de Student pour α et ddl = n-1

Estimation d'une Proportion par IC

$$IC_\alpha : \pi \in \left[p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \quad , p + Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right]$$

Avec la condition

$$n \geq 30 \quad et \;\; min\,\{np\,; n(1-p)\} \; > \; 5$$

Intervalle de confiance pour une somme de Moyenne

Supposons que nous disposons de deux (ou plus) échantillons pour mener l'étude d'un paramètre X. Soit:

$\overline{X}_1, S_1^2$:la moyenne et la variance de statistique X dans l'échantillon de taille n_1

$\overline{X}_2, S_2^2$:la moyenne et la variance de statistique X dans l'échantillon de taille n_2

Problème: Si on pose $Z = X_1 \pm X_2$. Quel est l' IC_α de la variable Z

Laboratoire d'Automatique et Productique LAP

Grand échantillon $(n \geq 30)$

Premier Cas: σ_1^2 et σ_2^2 Variance des populations connues

$$IC_\alpha : \mu_Z \in \left[\overline{x}_1 \pm \overline{x}_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \overline{x}_1 \pm \overline{x}_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

Deuxième Cas: σ_1^2 et σ_2^2 Variance des populations inconnues

$$IC_\alpha : \mu_Z \in \left[\overline{x}_1 \pm \overline{x}_2 - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}, \overline{x}_1 \pm \overline{x}_2 + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}} \right]$$