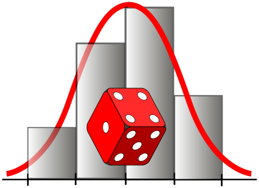
Estimation paramétrique par intervalle de confiance



Module : Techniques d'estimation pour l'ingénieur

Plan

1. Introduction
2. Principe de l'estimation par intervalle de confiance
3. Construction de l'intervalle de confiance
    - Intervalle de confiance de la moyenne
    - Intervalle de confiance de a proportion
    - Intervalle de confiance de la variance

Activité introductive

Dans une société musicale qui exécute des œuvres vocales, supposons qu'elle cherche à estimer la moyenne μ_{pop} de 40 chanteurs à partir d'un échantillon de 5 observations de cette chorale.
Si on estime μ_{pop} par la moyenne d'échantillon $\mu_{\text{éch}}$, donnée par :

$$\mu_{\text{éch}} = \frac{1}{5}\sum_{i=1}^{5} X_i$$

on ne peut pas raisonablement croire que $\mu_{\text{éch}} = \mu_{pop}$ exactement. On fera une petite erreur d'estimation.

D'ailleurs, comme l'échantillon est aléatoire, la valeur de $\mu_{\text{éch}}$ que vous auriez obtenue sur un autre échantillon aurait probablement été différente, quoique tout aussi pertinente.

On dit qu'elle ne prend pas en compte les fluctuations d'échantillonnage.

Voici quelques exemples d'échantillons ; on remarque effectivement que la valeur de $\mu_{\text{éch}}$ fluctue d'un échantillon à l'autre :

	X_1	X_2	X_3	X_4	X_5	$\mu_{\text{éch}}$
échantillon 1	1.89	1.79	1.74	1.90	1.74	1.812
échantillon 2	1.74	1.95	1.76	1.75	1.71	1.782
échantillon 3	1.84	1.84	1.88	1.85	1.89	1.86
échantillon 4	1.84	1.84	1.75	1.83	1.75	1.802
échantillon 5	1.59	1.68	1.79	1.89	1.79	1.748

? Comment avoir confiance en cette estimation ponctuelle?

Réponse : pour estimer μ_{pop}, vous ne pouvez pas simplement donner la valeur de $\mu_{\text{éch}}$, mais vous devez l'accompagner de **marges d'erreurs**. En d'autres termes , il est nécessaire de lui associer un intervalle qui contient, avec une certaine probabilité, la vraie valeur du paramètre dans la population.

L'objet de ce chapitre est de comprendre comment déterminer ces **marges d'erreurs**, ou, en termes mathématiques, comment construire un **intervalle de confiance d'un paramètre inconnu** .

Principe de l'estimation par intervalle de confiance

Principe

- L'estimation par intervalle de confiance consiste à associer à un échantillon un intervalle aléatoire I_θ dont on a de fortes chances de croire qu'il contient la vraie valeur du paramètre inconnu θ.
- Un intervalle de confiance I_θ, de **risque** $\alpha \in [0,1]$ ou bien de **niveau de signification** $1-\alpha$, pour un paramètre θ, est un intervalle aléatoire de la forme $[a_\alpha, b_\alpha]$, où $(a_\alpha, b_\alpha) \in \mathbb{R}^2$, définit comme suit:

$$\mathbb{P}(\theta \in I_\theta) = \mathbb{P}(a_\alpha \leq \theta \leq b_\alpha) = 1-\alpha$$

Principe de l'estimation par intervalle de confiance

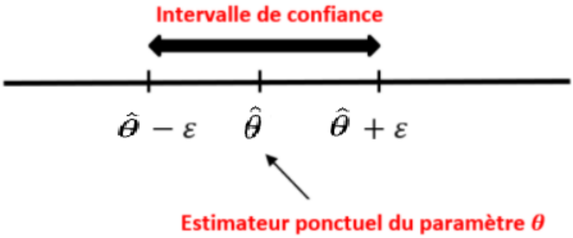
- Le risque α : est la probabilité que le paramètre θ n'appartienne pas à l'intervalle I_θ, autrement dit c'est la probabilité que l'on se trompe en affirmant que $\theta \in I_\theta$. C'est donc une probabilité d'erreur, qui doit être assez petite. Les valeurs usuelles de α sont 10%, 5%, 1%, …, etc.
- Le niveau de signification $1 - \alpha$: est la probabilité que le paramètre θ appartient à l'intervalle I_θ.

Principe de l'estimation par intervalle de confiance

↳ Il semble alors logique de chercher un intervalle de confiance I_θ pour θ de la forme $[\hat{\theta} - \varepsilon \ , \ \hat{\theta} + \varepsilon]$, où $\hat{\theta}$ est un estimateur ponctuel sans biais de θ. Selon la caractérisation ci-dessus ceci revient alors à déterminer les **marges d'erreurs** $\varepsilon > 0$ de sorte que :

$$\mathbb{P}(\hat{\theta} - \varepsilon < \theta < \hat{\theta} + \varepsilon) = \mathbb{P}(\widehat{\theta}_1 < \theta < \widehat{\theta}_2) = 1 - \alpha$$

Avec $\widehat{\theta}_1 = \hat{\theta} - \varepsilon$ et $\widehat{\theta}_2 = \hat{\theta} + \varepsilon$.



Principe de l'estimation par intervalle de confiance

Pour toute la suite du cours :

- $(X_1, \ldots, X_n)$, $n > 0$, un échantillon de taille n.
- Pour un risque α donné, on construira un intervalle de confiance I_θ dans le cas où l'inconnu θ est :
  1. La moyenne $\mu \in \mathbb{R}$ d'une population
  2. La variance $\sigma^2 \in \mathbb{R}$ d'une population
  3. la proportion p d'un caractère qualitatif relatif à une population.

esprit Se former autrement