

Module 7 - Intervalles de confiance

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Introduction aux intervalles de confiance

Introduction : estimation ponctuelle imparfaite

Un échantillon donne lieu à UNE estimation du paramètre. Quelle confiance peut-on avoir en cette estimation ?

Illustration.

Concept d'estimation par intervalle de confiance

Un intervalle de confiance :

- complète l'estimation ponctuelle en l'entourant d'une **marge d'erreur**.
- donne une idée de la précision obtenue avec l'échantillon.
- est associé à un **niveau de confiance** (probabilité que l'intervalle contienne la vraie valeur du paramètre).

Définition formelle

- $X_1, \dots, X_n$: échantillon aléatoire de taille n ;
- θ : paramètre d'intérêt (comme μ ou σ^2);

Intervalle de confiance de niveau $1 - \alpha$ pour un paramètre θ

Deux statistiques $[C_1, C_2]$ telles que

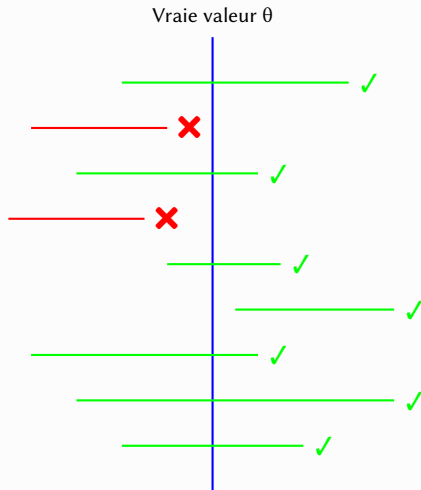
$$P(C_1 \leq \theta \leq C_2) = 1 - \alpha$$

- $1 - \alpha$: niveau de confiance de l'intervalle (fixé arbitrairement par le chercheur).
- Souvent $1 - \alpha = 0,95$. Parfois $1 - \alpha = 0,90$ ou $0,99$ ou autre.

Intervalles de confiance à 95%

Interprétation

En moyenne, 95% des intervalles de confiance à 95% **contiennent** la vraie valeur du paramètre.



✓ : l'intervalle contient θ
✗ : l'intervalle ne contient pas θ

Rappel sur les lois échantillonnales

Lorsque $X_1, \dots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$, on connaît les lois échantillonnales des statistiques suivantes :

Statistique	Loi échantillonnale
$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$	
$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$	
$U = \frac{(n-1)S^2}{\sigma^2}$	

4 cas étudiés

Nous construirons 4 intervalles de confiance :

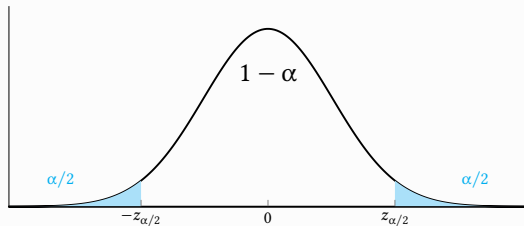
Cas	Paramètre	Situation
1	μ	$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue
2	μ	$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ avec σ^2 inconnue
3	μ	$X_1, \dots, X_n \sim$ loi inconnue, mais n est grand
4	σ^2	$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

Intervalles de confiance pour une moyenne

Cas 1 pour μ : loi normale, variance **CONNUE**

- On suppose que $X_1, \dots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, où σ^2 est une valeur connue.
- On déduit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, donc que $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
- Soit $z_{\alpha/2}$, la valeur telle que $P(Z > z_{\alpha/2}) = \alpha/2$. On a donc que

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.$$



Cas 1 pour μ : loi normale, variance connue

- On cherche $[C_1, C_2]$ tels que $P(C_1 \leq \mu \leq C_2) = 1 - \alpha$
- On sait que $P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$

Exemple 1

Soit X , le diamètre (en cm) des disques de frein produits par une machine, où $X \sim \mathcal{N}(\mu, 0,0009)$. On choisit un échantillon de 9 disques au hasard et on calcule un diamètre moyen de $\bar{x}$ cm.

a) Donnez un intervalle de confiance de niveau 95% pour μ .

Illustration

Exemple 1

b) Donnez l'IC si $\bar{x} = 12,05$.

QUIZ!

Déterminez si les énoncés suivants (liés à l'exemple 1) sont vrais ou faux.

- 95% des disques de cette compagnie ont un diamètre entre 12,03 et 12,07 cm.

[Illustration](#)

- 95% des intervalles calculés avec cette formule (à partir d'échantillons différents) contiendraient le vrai diamètre moyen.
- On a confiance à 95% que l'intervalle (12,03, 12,07) cm contient la vraie valeur de μ .
- Si on prenait un nouvel échantillon de taille $n = 9$, la probabilité que sa moyenne $\bar{X}$ soit entre 12,03 et 12,07 cm est de 95%. [Illustration](#)

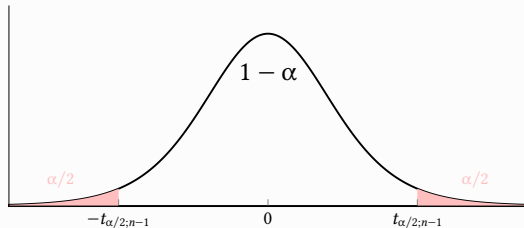
Exemple 1 : taille d'échantillon

c) Quelle est la taille minimale de l'échantillon à collecter pour que la marge d'erreur sur l'estimation du diamètre moyen ne dépasse pas 0,01 cm, si on conserve un niveau de confiance de 95% ?

Cas 2 pour μ : loi normale, variance INCONNUE

- On suppose que $X_1, \dots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, où σ^2 est une valeur inconnue, donc à estimer par S^2 .
- On déduit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, donc $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- Soit $t_{\alpha/2; n-1}$, la valeur telle que $P(T > t_{\alpha/2; n-1}) = \alpha/2$. Ainsi,

$$P(-t_{\alpha/2; n-1} \leq T \leq t_{\alpha/2; n-1}) = 1 - \alpha.$$



Cas 2 pour μ : loi normale, variance inconnue

- On cherche $[C_1, C_2]$ tels que $P(C_1 \leq \mu \leq C_2) = 1 - \alpha$
- On sait que $P\left(-t_{\alpha/2; n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2; n-1}\right) = 1 - \alpha$

Exemple 2

Retournons à l'exemple des disques. Supposons que σ^2 ne soit pas connue et qu'on l'estime par la variance échantillonnale des 9 diamètres observés, soit $s^2 = 0,0009 \text{ cm}^2$.

Donnez un intervalle de confiance à 95% pour μ . Comparez-le avec l'intervalle obtenu en supposant la variance connue.

Cas 3 pour μ : loi inconnue, n grand

- On suppose que $X_1, \dots, X_n$ sont i.i.d. d'une loi inconnue.
- Peut-on construire des intervalles de confiance ?
- L'idée : utiliser le **théorème central limite** et le fait que pour n grand :

$$s^2 \simeq \sigma^2 \text{ et } \bar{X} \approx \mathcal{N}(\mu, s^2/n) \Rightarrow$$

- Si n est grand, on peut construire des intervalles de confiance dont le niveau sera **approximativement** du niveau désiré.

Notez qu'il est possible qu'on soit dans une situation où σ^2 est supposée connue. Il faut donc simplement utiliser σ à la place de s dans tout ce qui suit.

Cas 3 pour μ : loi inconnue, n grand

- On cherche $[C_1, C_2]$ tels que $P(C_1 \leq \mu \leq C_2) = 1 - \alpha$
- On sait que $P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$
- On déduit que $P\left(\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right) \approx 1 - \alpha$

Exemple 3

Pour estimer le salaire moyen annuel de jeunes diplômés dans un certain domaine, on a prélevé un échantillon de taille $n = 100$ et obtenu $\bar{x} = 34\,224 \$$ et $s = 4\,668 \$$.

Donnez un intervalle de confiance de niveau 90% pour le salaire moyen annuel des jeunes diplômés dans ce domaine.

La marge d'erreur et l'erreur-type

Les trois intervalles de confiance pour μ sont symétriques :

$$\bar{X} \pm \text{marge d'erreur}$$

$$\bar{X} \pm \text{quantile} \times \text{erreur-type}$$

Longueur de l'intervalle = $2 \times$ marge d'erreur

QUIZ!

Déterminez si les énoncés suivants sont vrais ou faux.

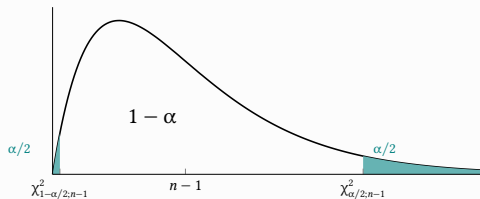
La marge d'erreur d'un IC sur la moyenne μ augmente (donc la précision diminue) si :

- la variabilité des mesures augmente.
- la taille d'échantillon augmente.
- le niveau de confiance augmente.

Intervalles de confiance pour une variance

Cas 4 pour σ^2 : loi normale, μ et σ^2 inconnues

- On suppose que $X_1, \dots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, où μ et σ^2 sont inconnues.
- On sait que $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$
- Soit $\chi_{\alpha/2; n-1}^2$, la valeur telle que $P(U > \chi_{\alpha/2; n-1}^2) = \alpha/2$. Ainsi,



$$P(\chi_{1-\alpha/2; n-1}^2 \leq U \leq \chi_{\alpha/2; n-1}^2) = 1 - \alpha.$$

Cas 4 pour σ^2 : loi normale, μ et σ^2 inconnues

- On cherche $[C_1, C_2]$ tels que $P(C_1 \leq \sigma^2 \leq C_2) = 1 - \alpha$
- On sait que $P\left(\chi_{1-\alpha/2; n-1}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{\alpha/2; n-1}^2\right) = 1 - \alpha$

Intervalle pour l'écart-type σ

$$P(C_1 \leq \sigma^2 \leq C_2) = 1 - \alpha$$

est équivalent à

$$P(\sqrt{C_1} \leq \sqrt{\sigma^2} \leq \sqrt{C_2}) = 1 - \alpha.$$

Pour avoir un IC(σ), on prend la racine carrée des bornes de IC(σ^2).

Exemple 4

On veut calibrer la machine qui fabrique les disques de frein pour qu'elle soit la plus stable possible. On veut d'abord estimer la variation du diamètre des disques produits, avec un intervalle de confiance à 95%.

Un échantillon de taille 25 a donné une variance échantillonnale de $0,015 \text{ cm}^2$. On suppose que les diamètres sont i.i.d. de loi normale.

En extra

Seriez-vous capable de construire un intervalle de confiance pour un ratio de variances ?

Réponses aux exemples

1. b) $IC(\mu)$ à 95% : [12,030; 12,070]
 c) $n = 35$
2. $IC(\mu)$ à 95% : [12,028; 12,072]
3. $IC(\mu)$ à 90% : [33 456,114; 34 991,886]
4. $IC(\sigma)$ à 95% : [0,096; 0,170]