

# Quelques rappels sur les intervalles de confiance

## I/ Généralités

Soient :  $X$  une variable aléatoire de loi paramétrée par  $\theta$  et  $X_1, \dots, X_n$   $n$  variables i.i.d selon la loi de  $X$ .

### 1) Principe d'un intervalle de confiance

Plutôt que d'estimer ponctuellement la vraie valeur inconnue du paramètre  $\theta$ , on recherche un intervalle recouvrant «très vraisemblablement» cette vraie valeur.

Définition : On appelle intervalle de confiance de niveau de confiance  $1 - \alpha$  du paramètre  $\theta$  tout intervalle  $IC$  tel que :  $P(IC \ni \theta) = 1 - \alpha$  pour  $\alpha \in [0,1]$  fixé.

Les bornes de l'intervalle de confiance  $IC$  dépendent de l'échantillon, elles sont donc aléatoires.

Par abus de langage, on note souvent  $P(\theta \in IC) = 1 - \alpha$ .

Remarquons que si  $\alpha$  augmente (ou que si  $n$  augmente), l'amplitude de l'intervalle de confiance diminue.

### 2) Vocabulaire

La probabilité  $\alpha$  pour que l'intervalle de confiance ne contienne pas la vraie valeur peut être répartie différemment de part et d'autre des bornes de l'intervalle de confiance. Ecrivons donc  $\alpha = \alpha_1 + \alpha_2$  où  $\alpha_1$  et  $\alpha_2$  mesurent respectivement les risques à gauche et à droite de dépasser un seuil plancher ou plafond.

- L'intervalle de confiance est dit bilatéral quand  $\alpha_1 \neq 0$  et  $\alpha_2 \neq 0$ . Si  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ , l'intervalle est dit symétrique. Il est dissymétrique sinon.
- L'intervalle de confiance est dit unilatéral si  $\alpha_1 \alpha_2 = 0$  :
  - quand on veut assurer une valeur minimale au paramètre à estimer, on considère  $\alpha_1 = \alpha$  et  $\alpha_2 = 0$ , l'intervalle de confiance est alors de la forme :  $IC = [a, +\infty[$ .
  - quand on ne veut absolument pas dépasser un seuil maximal, on prend  $\alpha_1 = 0$  et  $\alpha_2 = \alpha$  et on obtient alors un intervalle de confiance de la forme :  $IC = ]-\infty, b]$ .

### 3) Construction

Pour construire un intervalle de confiance, on utilise une variable aléatoire dont on connaît la distribution de probabilité.

Définition : une fonction pivotale pour le paramètre  $\theta$  est une fonction des observations  $(X_1, \dots, X_n)$  et du paramètre  $\theta$  dont la loi ne dépend pas du paramètre  $\theta$ .

On recherche dans la suite des fonctions pivotales particulières adaptées aux cas étudiés.

## II/ Intervalles de confiance pour l'espérance

On envisage deux cas :

- la variable aléatoire mesurée est normale et le nombre de réalisations est quelconque,
- la variable aléatoire mesurée n'est pas normale et le nombre de réalisations est important. Dans ce cas, la distribution de la moyenne empirique tend vers une loi normale d'après le théorème central limite. On parlera d'intervalle de confiance asymptotique.

Dans la suite on considère  $X \sim N(m, \sigma^2)$  et  $X_1, \dots, X_n$   $n$  variables i.i.d selon la loi de  $X$ .

On définit la moyenne empirique  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  et la variance empirique modifiée  $S_n'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ .

### 1) Cas où la variance est connue

Après centrage et réduction de la moyenne empirique, on obtient :  $\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \sim N(0,1)$

On a :  $P\left(-u \leq \sqrt{n} \frac{\bar{X}_n - m}{\sigma} \leq u\right) = 1 - \alpha$  où  $u$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi  $N(0,1)$ .

Ce qui revient à :  $P\left(\bar{X}_n - u \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + u \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ .

Quand la variance est connue, l'intervalle de confiance bilatéral symétrique pour l'espérance d'une loi normale s'écrit donc au niveau  $1 - \alpha$  sous la forme suivante :

$$\boxed{IC(m) = \left[ \bar{x}_n - u \frac{\sigma}{\sqrt{n}}, \bar{x}_n + u \frac{\sigma}{\sqrt{n}} \right]} \quad \bar{x}_n \text{ est la réalisation de } \bar{X}_n \text{ sur l'échantillon.}$$

Remarque : si  $\alpha = 5\%$ , le fractile d'ordre 0,975 de la loi normale centrée réduite correspond à 1,96.  
si  $\alpha = 10\%$ , le fractile d'ordre 0,95 de la loi normale centrée réduite vaut environ 1,64.

### 2) Cas où la variance est inconnue

On a :  $\sqrt{n} \frac{\bar{X}_n - m}{S_n'} \sim St(n-1)$  (loi de Student à  $n-1$  degrés de libertés).

d'où  $P\left(-t \leq \sqrt{n} \frac{\bar{X}_n - m}{S_n'} \leq t\right) = 1 - \alpha$  où  $t$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi  $St(n-1)$

et donc  $P\left(\bar{X}_n - t \frac{S_n'}{\sqrt{n}} \leq m \leq \bar{X}_n + t \frac{S_n'}{\sqrt{n}}\right) = 1 - \alpha$ .

Quand la variance est inconnue, l'intervalle de confiance bilatéral symétrique pour l'espérance d'une loi normale s'écrit donc au niveau  $1 - \alpha$  sous la forme suivante :

$$\boxed{IC(m) = \left[ \bar{x}_n - t \frac{s_n'}{\sqrt{n}}, \bar{x}_n + t \frac{s_n'}{\sqrt{n}} \right]} \quad \bar{x}_n \text{ et } s_n' \text{ sont les réalisations respectives de } \bar{X}_n \text{ et } S_n' \text{ sur l'échantillon.}$$

Remarque : quand  $n \rightarrow \infty$ , on approxime la loi de Student par la loi normale centrée réduite. On retrouve alors le cas précédent.

### 3) Cas particulier : intervalle de confiance pour une proportion

Soient  $X_1, \dots, X_n$  i.i.d. selon  $B(p)$  et  $X = \sum_{i=1}^n X_i \sim B(n, p)$ . Notons  $F_n = \frac{X}{n}$  estimateur sans biais de  $p$ .

- Dans le cas de grands échantillons :

En approchant une loi binomiale vers une loi normale, on a :  $\sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \xrightarrow[n \rightarrow \infty]{\text{loi}} N(0,1)$ .

Ce qui permet d'écrire :  $P\left(-u \leq \sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \leq u\right) = 1 - \alpha$  où  $u$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi  $N(0,1)$ .

Et donc l'intervalle de confiance bilatéral symétrique pour une proportion  $p$  au niveau  $1 - \alpha$  s'obtient en

résolvant l'inéquation :  $\left| \sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \right| \leq u$

Ce qui donne en notant  $f_n$  la réalisation de  $F_n$  sur l'échantillon:

$$IC(p) = \left[ \frac{f_n + \frac{u^2}{2n} - \frac{u}{\sqrt{n}} \sqrt{\frac{u^2}{4n} + f_n(1-f_n)}}{1 + \frac{u^2}{n}}, \frac{f_n + \frac{u^2}{2n} + \frac{u}{\sqrt{n}} \sqrt{\frac{u^2}{4n} + f_n(1-f_n)}}{1 + \frac{u^2}{n}} \right]$$

Pour une taille d'échantillon importante, on considère l'approximation suivante :

$$IC(p) = \left[ f_n - u \sqrt{\frac{f_n(1-f_n)}{n}}, f_n + u \sqrt{\frac{f_n(1-f_n)}{n}} \right]$$

Cette approximation est parfaitement justifiée sur le plan théorique.

En effet, d'après le théorème de Slutsky, on a :  $\sqrt{F_n(1-F_n)} \xrightarrow{p} \sqrt{p(1-p)}$ .

On en déduit donc que :  $\sqrt{n} \frac{F_n - p}{\sqrt{F_n(1-F_n)}} \xrightarrow[n \rightarrow \infty]{\text{loi}} N(0,1)$ .

D'où :  $P\left(-u \leq \sqrt{n} \frac{F_n - p}{\sqrt{F_n(1-F_n)}} \leq u\right) = 1 - \alpha$  où  $u$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de la loi  $N(0,1)$ .

Quand  $n$  est grand, l'intervalle de confiance bilatéral symétrique pour une proportion s'écrit donc au niveau  $1 - \alpha$  sous la forme indiquée :

$$IC(p) = \left[ f_n - u \sqrt{\frac{f_n(1-f_n)}{n}}, f_n + u \sqrt{\frac{f_n(1-f_n)}{n}} \right] \quad f_n \text{ est la réalisation de } F_n \text{ sur l'échantillon.}$$

- Sinon, construction d'intervalles de confiance « exacts » :

On construit ces intervalles en considérant la fonction de répartition de la loi binomiale. Si la probabilité de recouvrement de l'intervalle ne vaut pas exactement  $1 - \alpha$ , on prend l'intervalle ayant la plus petite probabilité de recouvrement parmi ceux ayant une probabilité de recouvrement supérieure à  $1 - \alpha$ .

### III/ Intervalles de confiance pour la variance d'une loi normale

Soient  $X \sim N(m, \sigma^2)$  et  $X_1, \dots, X_n$   $n$  variables i.i.d selon la loi de  $X$ .

#### 1) Cas où l'espérance est connue

Soit  $S_n^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ . On a  $n \frac{S_n^{*2}}{\sigma^2} \sim \chi^2(n)$

D'où  $P\left(\chi^2_{\frac{\alpha_1}{2}} \leq n \frac{S_n^{*2}}{\sigma^2} \leq \chi^2_{1-\frac{\alpha_2}{2}}\right) = 1 - \alpha$  où  $\chi^2_{\alpha_1}$  est le fractile d'ordre  $\alpha_1$  de la loi  $\chi^2(n)$ ,  
et  $\chi^2_{1-\alpha_2}$  est le fractile d'ordre  $1 - \alpha_2$  de la loi  $\chi^2(n)$ .

Quand l'espérance est connue, l'intervalle de confiance bilatéral pour la variance d'une loi normale s'écrit donc au niveau  $1 - \alpha$  sous la forme suivante :

$$IC(\sigma^2) = \left[ n \frac{S_n^{*2}}{\chi^2_{1-\frac{\alpha_2}{2}}}, n \frac{S_n^{*2}}{\chi^2_{\frac{\alpha_1}{2}}} \right]$$

$S_n^*$  est la réalisation de  $S_n^*$  sur l'échantillon.

Remarque : cet intervalle n'est pas centré car la loi du khi-deux n'est pas symétrique.

#### 2) Cas où l'espérance est inconnue

On considère la variance empirique modifiée  $S_n'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  comme fonction pivotale pour  $\sigma^2$ .

On sait que  $\frac{(n-1)S_n'^2}{\sigma^2} \sim \chi^2(n-1)$ .

On a donc  $P\left(\chi^2_{\frac{\alpha_1}{2}} \leq (n-1) \frac{S_n'^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha_2}{2}}\right) = 1 - \alpha$  où  $\chi^2_{\alpha_1}$  est le fractile d'ordre  $\alpha_1$  de la loi  $\chi^2(n-1)$ ,  
et  $\chi^2_{1-\alpha_2}$  le fractile d'ordre  $1 - \alpha_2$  de la loi  $\chi^2(n-1)$ .

Quand l'espérance est inconnue, l'intervalle de confiance bilatéral pour la variance d'une loi normale s'écrit donc au niveau  $1 - \alpha$  sous la forme suivante :

$$IC(\sigma^2) = \left[ (n-1) \frac{S_n'^2}{\chi^2_{1-\frac{\alpha_2}{2}}}, (n-1) \frac{S_n'^2}{\chi^2_{\frac{\alpha_1}{2}}} \right]$$

$S_n'$  est la réalisation de  $S_n'$  sur l'échantillon.