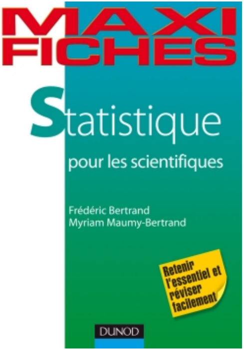
Étude de cas

Myriam Maumy-Bertrand

IRMA, UMR 7501, Université de Strasbourg

04 octobre 2013

Ce chapitre s'appuie essentiellement sur le livre :



3ème partie

Intervalles de confiance

Sommaire

Il est souvent plus réaliste et plus intéressant de fournir un renseignement du type

$$\theta_1 < \theta < \theta_2$$

plutôt que d'écrire sèchement

$$\widehat{\theta}_n = c.$$

Définition

Fournir un tel intervalle $]\theta_1 ; \theta_2[$ s'appelle donner une estimation par intervalle de confiance de θ ou une estimation ensembliste de θ.

Voir le livre « Maxi Fiches », page 124.

Sommaire

La méthode des intervalles de confiance est la suivante : soit $\widehat{\theta}_n$ un estimateur de θ dont nous connaîssons la loi de probabilité pour chaque valeur de θ.

Définition

Étant donné une valeur θ_0 du paramètre θ, nous déterminons un **intervalle de probabilité bilatéral de niveau** $(1-\alpha)$ *pour l'estimateur $\widehat{\theta}_n$, c'est-à-dire deux bornes θ_1^n et θ_2^n telles que*

$$\mathbb{P}\left(\theta_1^n < \widehat{\theta}_n < \theta_2^n | \theta = \theta_0\right) \geqslant 1 - \alpha.$$

Voir le livre « Maxi Fiches », page 124.

Remarque

Ces deux bornes dépendent évidemment de la valeur θ_0.

Remarque

Nous pourrions également construire des **intervalles unilatéraux** pour lesquels $\theta_1^n = -\infty$ ou $\theta_2^n = +\infty$.

Remarque

Nous choisissons dans la plupart des cas un intervalle de probabilité à risque symétrique $\alpha/2$ et $\alpha/2$.

Voir le livre « Maxi Fiches », page 124.

Nous adoptons la règle de décision suivante.
Soit $\widehat{\theta}_n(x_1, \ldots, x_n) = \widehat{\theta}_n(obs)$ la valeur observée de $\widehat{\theta}_n$:

- si $\widehat{\theta}_n(obs) \in]\theta_1^n; \theta_2^n[$, nous conservons θ_0 comme valeur possible du paramètre θ ;
- si $\widehat{\theta}_n(obs) \notin]\theta_1^n; \theta_2^n[$, nous éliminons θ_0.

Nous répétons cette opération pour toutes les valeurs de θ.

Voir le livre « Maxi Fiches », page 124.

Sommaire

Définition

Nous appelons **intervalle de confiance de niveau de confiance** $(1-\alpha)$ **(coefficient de confiance) du paramètre** θ *tout intervalle* $]\theta_1;\theta_2[$ *tel que :* $\mathbb{P}(\theta \in]\theta_1;\theta_2[) = 1-\alpha$ *pour* $\alpha \in [0;1]$ *fixé.*

Voir le livre « Maxi Fiches », page 124.

Propriétés

1. $]\theta_1 ; \theta_2[$ est un intervalle aléatoire car il dépend de l'estimateur $\widehat{\theta}_n$.
2. $]\theta_1 ; \theta_2[$ s'obtient par :

$$\left\{ \begin{array}{lcl} \theta_1 & = & \left(\theta_2^n\right)^{-1}(\widehat{\theta}_n(obs)) \\ \theta_2 & = & \left(\theta_1^n\right)^{-1}(\widehat{\theta}_n(obs)) \end{array} \right. .$$

3. Si nous augmentons le niveau $1-\alpha$, nous augmentons la longueur de l'intervalle de probabilité.

Voir le livre « Maxi Fiches », page 125.

Remarque

Si la taille de l'échantillon notée n augmente, comme l'estimateur $\widehat{\theta}_n$ est supposé convergent, la variance de l'estimateur notée $\mathrm{Var}\left(\widehat{\theta}_n\right)$ diminue et par conséquent l'intervalle $]\theta_1\,;\theta_2[$ diminue également.

Sommaire

$\widehat{\mu}_n$ est le meilleur estimateur de μ et $\widehat{\mu}_n$ suit une loi $\mathcal{N}\left(\mu; \dfrac{\sigma^2}{n}\right)$.

Définition

L'intervalle de probabilité de $\widehat{\mu}_n$ à $1-\alpha$ est :

$$\mu - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \widehat{\mu}_n < \mu + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},$$

où u_p est le quantile d'ordre p pour la loi gaussienne centrée et réduite.

Voir le livre « Maxi Fiches », page 126.

Définition

L'intervalle de confiance de $\widehat{\mu}_n$ à $1-\alpha$ est :

$$\widehat{\mu}_n(x_1,\ldots,x_n) - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \widehat{\mu}_n(x_1,\ldots,x_n) + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},$$

où $u_{1-\alpha/2} = 1,96$ si $1-\alpha = 0,95$.

Voir le livre « Maxi Fiches », page 126.

Nous utilisons le fait que $T_{n-1} = \sqrt{n-1}\,\dfrac{\widehat{\mu}_n - \mu}{S_n}$ suit une loi de Student à $(n-1)$ degrés de liberté.

Définition

L'intervalle de probabilité pour T_{n-1} à $1-\alpha$ est :

$$-t_{n-1;1-\alpha/2} < \sqrt{n-1}\,\frac{\widehat{\mu}_n - \mu}{S_n} < t_{n-1;1-\alpha/2},$$

où $t_{n-1;1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ pour la loi de Student à $(n-1)$ degrés de liberté.

Voir le livre « Maxi Fiches », page 127.

Définition

L'intervalle de confiance pour μ à $1-\alpha$ est :

$$\widehat{\mu}_n(obs) - t_{n-1;1-\alpha/2}\frac{S_n(obs)}{\sqrt{n-1}} < \mu < \widehat{\mu}_n(obs) + t_{n-1;1-\alpha/2}\frac{S_n(obs)}{\sqrt{n-1}}$$

ou bien

$$\widehat{\mu}_n(obs) - t_{n-1;1-\alpha/2}\frac{S_{n,c}(obs)}{\sqrt{n}} < \mu < \widehat{\mu}_n(obs) + t_{n-1;1-\alpha/2}\frac{S_{n,c}(obs)}{\sqrt{n}},$$

où $t_{n-1;1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ pour la loi de Student à $(n-1)$ degrés de liberté.

Voir le livre « Maxi Fiches », page 127.

Remarque

Le théorème de la limite centrée a pour conséquence que les intervalles précédents sont valables pour estimer μ d'une loi quelconque lorsque la taille n de l'échantillon est assez grande.

Sommaire

Nous utilisons le fait que

1. $\widehat{\sigma^2}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$ est le meilleur estimateur de la variance σ^2 lorsque la moyenne μ est connue,
2. $\dfrac{n\widehat{\sigma^2}_n}{\sigma^2}$ suit une loi du khi-deux à n degrés de liberté comme la somme de n carrés de loi $\mathcal{N}(0;1)$ indépendantes.

Voir le livre « Maxi Fiches », page 128.

Définition

k_1 et k_2 sont les bornes de l'intervalle de probabilité pour $\dfrac{n\widehat{\sigma^2}_n}{\sigma^2}$ si :

$$\mathbb{P}\left(k_1 < \frac{n\widehat{\sigma^2}_n}{\sigma^2} < k_2\right) = 1 - \alpha.$$

Remarque

Par exemple, nous pouvons prendre la borne k_1 égale au quantile d'ordre $\alpha/2$ pour la loi du Khi-deux à n degrés de liberté et la borne k_2 égale au quantile d'ordre $1 - \alpha/2$ pour la loi du Khi-deux à n degrés de liberté.

Voir le livre « Maxi Fiches », page 128.

Définition

L'intervalle de confiance pour σ^2 à $1-\alpha$ est égal à :

$$\frac{n\widehat{\sigma^2}_n(x_1,\ldots,x_n)}{k_2} < \sigma^2 < \frac{n\widehat{\sigma^2}_n(x_1,\ldots,x_n)}{k_1}.$$

Remarque

Par exemple, nous pouvons prendre la borne k_1 égale au quantile d'ordre $\alpha/2$ pour la loi du Khi-deux à n degrés de liberté et la borne k_2 égale au quantile d'ordre $1-\alpha/2$ pour la loi du Khi-deux à n degrés de liberté.

Voir le livre « Maxi Fiches », page 128.

Nous utilisons le fait que

1. $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_n)^2$,
2. $\dfrac{nS_n^2}{\sigma^2}$ suit une loi du khi-deux à $(n-1)$ degrés de liberté comme la somme de $(n-1)$ carrés de loi $\mathcal{N}(0;1)$ indépendantes.

Voir le livre « Maxi Fiches », page 128.

Définition

l_1 et l_2 sont les bornes de l'intervalle de probabilité pour $\frac{nS_n^2}{\sigma^2}$ si :

$$\mathbb{P}\left(l_1 < \frac{nS_n^2}{\sigma^2} < l_2\right) = 1 - \alpha.$$

Définition

L'intervalle de confiance pour σ^2 à $1 - \alpha$ est égal à :

$$\frac{nS_n^2(x_1, \ldots, x_n)}{l_2} < \sigma^2 < \frac{nS_n^2(x_1, \ldots, x_n)}{l_1}.$$

Voir le livre « Maxi Fiches », page 129.

Sommaire

- Étant donné une population INFINIE (ou finie si le tirage s'effectue avec remise) où une proportion π_A des individus possède un certain caractère A, il s'agit de trouver un intervalle de confiance pour π_A à partir de $\widehat{\pi}_{n,A}$, proportion trouvée dans un échantillon de taille n.
- Nous savons que $n\widehat{\pi}_{n,A}$ suit une loi binomiale $\mathcal{B}(n, \pi_A)$. Deux cas alors se présentent :
  1. Si n est petit nous utiliserons les tables de la loi binomiale.
  2. Si n est grand nous savons que $\widehat{\pi}_{n,A}$ suit une loi gaussienne de paramètres π_A et $\dfrac{\pi_A(1-\pi_A)}{n}$.

Voir le livre « Maxi Fiches », page 130.

Définition

L'intervalle de probabilité symétrique « approché » est :

$$\pi_A - u_{1-\alpha/2}\sqrt{\frac{\pi_A(1-\pi_A)}{n}} < \widehat{\pi}_{n,A} < \pi_A + u_{1-\alpha/2}\sqrt{\frac{\pi_A(1-\pi_A)}{n}}.$$

Définition

L'intervalle de confiance « approché » est alors :

$$\widehat{\pi}_{n,A}(obs) - u_{1-\alpha/2}\sqrt{\frac{\widehat{\pi}_{n,A}(obs)(1-\widehat{\pi}_{n,A}(obs))}{n}} < \pi_A$$
$$< \widehat{\pi}_{n,A}(obs) + u_{1-\alpha/2}\sqrt{\frac{\widehat{\pi}_{n,A}(obs)(1-\widehat{\pi}_{n,A}(obs))}{n}}.$$

Voir le livre « Maxi Fiches », page 131.