L'estimation par intervalles

Frédéric Bertrand[1]

[1]IRMA, Université de Strasbourg
Strasbourg, France

ESIEA 4ème Année 01-03-2012

Ce chapitre s'appuie essentiellement sur trois livres :

1. Le livre de Pascal Ardilly,
   administrateur de l'INSEE,
   « Les techniques de sondage »,
   Éditions Technip, 2006.
2. Le livre de Jean-Jacques Droesbeke,
   professeur à l'Université Libre de Bruxelles,
   « Éléments de statistiques »,
   Université de Bruxelles, Ellipses, 2001.
3. Le livre de Gilbert Saporta,
   professeur au CNAM à Paris
   « Probabilités, Analyse de données et Statistique »,
   Éditions Technip, 2006.

Sommaire

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

$$a < \theta < b$$

plutôt que d'écrire sèchement

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

Définition

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

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 de θ, nous déterminons un intervalle de probabilité de niveau $1-\alpha$ pour $\widehat{\theta}_n$, c'est-à-dire deux bornes θ_1 et θ_2 telles que

$$\mathbb{P}\left[\theta_1 < \widehat{\theta}_n < \theta_2 | \theta = \theta_0\right] = 1 - \alpha.$$

Remarque

Ces bornes dépendent évidemment de θ_0.

Remarque

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

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

Nous répétons cette opération pour toutes les valeurs de θ.

Définition

$]a; b[$ *est qualifié d'un intervalle de confiance de niveau* $(1 - \alpha)$ *(coefficient de confiance).*

Remarque

$]a; b[$ est un intervalle aléatoire car il dépend de $\widehat{\theta}_n(obs)$.

$]a, b[$ s'obtient par :

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

Remarque

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

Remarque

Si n augmente, comme $\widehat{\theta}_n$ est supposé convergent, $\mathrm{Var}\left[\widehat{\theta}_n\right]$ diminue donc $]\theta_1 ; \theta_2[$ diminue.

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.

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 $1-\alpha = 0,95$.

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.

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

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

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

Définition

k_1 et k_2 sont les bornes de l'intervalle de probabilité d'une loi $\chi^2(n)$ si :

$$\mathbb{P}\left[k_1 < \frac{n\,S_n^{*2}}{\sigma^2} < k_2\right] = 1 - \alpha.$$

Définition

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

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

Nous utilisons le fait que

1. $$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_n)^2,$$

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.

Définition

l_1 et l_2 sont les bornes de l'intervalle de probabilité d'un $\chi^2(n-1)$ 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 est alors :

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

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 $\dfrac{\pi_A(1-\pi_A)}{n}$.

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 :

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