Étude de quelques estimateurs

Frédéric Bertrand[1]

[1] IRMA, Université de Strasbourg
Strasbourg, France

ESIEA 4ème Année 08-02-2010

Ce chapitre s'appuie essentiellement sur deux livres :

1. le livre de Jean-Jacques Droesbeke,
   professeur à l'Université Libre de Bruxelles,
   « Éléments de statistiques »,
   Université de Bruxelles, Ellipses, 2001.
2. le livre de Pascal Ardilly,
   administrateur de l'INSEE,
   « Les techniques de sondage »,
   Éditions Technip, 2006.

Sommaire

Nous rappelons la définition suivante :

Définition

L'estimateur $\widehat{\mu}_n$ d'un échantillon aléatoire de taille n d'une moyenne μ est égal à :

$$\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Soient μ et σ^2 l'espérance et la variance de la variable aléatoire X. Nous rappelons que nous avons montré la propriété suivante :

Propriété

Soient $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n. Nous avons

$$\mathbb{E}\left[\widehat{\mu}_n\right] = \mu \quad \textit{et} \quad \mathrm{Var}\left[\widehat{\mu}_n\right] = \frac{\sigma^2}{n}.$$

Propriété

Le théorème de la limite centrée (en abrégé le TCL) nous permet d'écrire :

$$\sqrt{n}\,\frac{\widehat{\mu}_n - \mu}{\sigma} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1) \quad \textit{lorsque} \quad n \to +\infty.$$

Remarque

Il s'agit de la convergence en loi évoquée ici que nous allons rappeler ci-dessous.

Définition

Une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$, de fonction de répartition F_n, converge en loi vers une variable aléatoire X de fonction de répartition F_X continue, si pour tout $x \in \mathbb{R}$, nous avons

$$\lim_{n\to+\infty} F_n(x) = F(x).$$

Remarque

Si l'échantillon aléatoire est réalisé avec un tirage sans remise, alors la variance de l'estimateur $\widehat{\mu}_n$ en est modifiée par ce qui est appelé le coefficient d'exhaustivité.

Proposition

Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n effectué avec un tirage sans remise. Nous avons les deux résultats suivants (Ardilly, p :196-198) :

$$\mathbb{E}\left[\widehat{\mu}_n\right] = \mu \quad \textit{et} \quad \mathrm{Var}\left[\widehat{\mu}_n\right] = \frac{N-n}{N-1}\frac{\sigma^2}{n}.$$

Sommaire

Nous rappelons la définition suivante :

Définition

L'estimateur S_n^2 d'un échantillon aléatoire de taille n d'une variance σ^2 est égal à :

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

Nous rappelons, dans le cas d'un tirage avec remise, la propriété suivante :

Propriété

$$\mathbb{E}\left[S_n^2\right] = \frac{n-1}{n}\sigma^2.$$

Cette propriété montre que $\mathbb{E}\left[S_n^2\right] \neq \sigma^2$.

Remarque

S_n^2 est donc un estimateur biaisé de la variance σ^2 de la population U.

Sommaire

C'est parce que l'estimateur S_n^2 est biaisé que nous avons introduit précédemment un autre estimateur noté $S_{n,c}^2$ défini par :

Définition

L'estimateur corrigé $S_{n,c}^2$ d'un échantillon aléatoire de taille n d'une variance est égal à :

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

Calculons l'espérance $S^2_{n,c}$ en détaillant cette fois-ci les calculs.

$$
\begin{aligned}
\mathbb{E}\left[S^2_{n,c}\right] &= \mathbb{E}\left[\frac{n}{n-1}S^2_n\right] \\
&= \frac{n}{n-1} \times \mathbb{E}\left[S^2_n\right] \\
&= \frac{n}{n-1} \times \frac{n-1}{n} \times \sigma^2, \text{car les } X_i \text{ indépendantes} \\
&= \sigma^2.
\end{aligned}
$$

Propriété

$S^2_{n,c}$ est donc un estimateur sans biais de la variance σ^2 de la population U.

Remarque

Si l'échantillon aléatoire est réalisé avec un tirage sans remise, alors l'espérance et la variance de l'estimateur $S^2_{n,c}$ en sont modifiées.

Proposition

Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n effectué avec un tirage sans remise. Nous avons les deux résultats suivants (Ardilly, p :43-49) :

$$\mathbb{E}\left[S_{n,c}^2\right] = \frac{N}{N-1}\sigma^2 = \sigma_c^2$$

$$\begin{aligned}\mathrm{Var}\left[S_{n,c}^2\right] &= \frac{(N-n)N}{n(n-1)(N-1)^2(N-2)(N-3)} \\ &\times\Big\{\mu_4(N-1)\left[N(n-1)-(n+1)\right] \\ &-\sigma^4\left[N^2(n-3)+6N-3(n+1)\right]\Big\}.\end{aligned}$$

Sommaire

Pour appliquer le théorème de la limite centrée, nous devons alors calculer la variance de S_n^2.

Propriété

Soient $(X_1, \ldots, X_n)$ un échantillon aléatoire de taille n. Nous avons

$$\mathrm{Var}\left[S_n^2\right] = \frac{n-1}{n^3}\left[(n-1)\mu_4 - (n-3)\sigma^4\right],$$

où μ_4 désigne le moment centré d'ordre 4 de X, c'est-à-dire :

$$\mu_4 = \mathbb{E}\left[(X-\mu)^4\right].$$

Remarque

Si la taille n de l'échantillon tend vers plus l'infini, alors nous avons la simplification suivante :

$$\operatorname{Var}\left[S_n^2\right] \sim \frac{\mu_4 - \sigma^4}{n}.$$

Le théorème de la limite centrée donne lieu au :

Théorème

$$\frac{S_n^2 - \dfrac{n-1}{n}\sigma^2}{\sqrt{\mathrm{Var}\left[S_n^2\right]}} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1) \text{ lorsque } n \to +\infty,$$

ce qui peut s'écrire avec l'approximation précédente :

$$\sqrt{n}\,\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1) \text{ lorsque } n \to +\infty.$$

Sommaire

Nous supposons maintenant que X suit une loi $\mathcal{N}(\mu, \sigma^2)$:

Proposition

$\widehat{\mu}_n$ étant une combinaison linéaire de variables gaussiennes est aussi une variable gaussienne. Donc

$$\widehat{\mu}_n \quad \textit{suit une loi} \quad \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Remarque

Il s'agit ici d'une loi exacte contrairement au cas précédent où nous avions une approximation.

Théorème

$$\frac{nS_n^2}{\sigma^2} \text{ suit une loi du } \chi^2_{n-1}.$$

Remarque

La démonstration de ce théorème n'est pas présentée dans ce cours. Le lecteur pourra la consulter dans le livre du Professeur Saporta intitulé : « Probabilités, Analyse des données et Statistique », Seconde édition, Éditions Technip, 2006, pages 282.

Théorème

Les deux estimateurs $\widehat{\mu}_n$ et S_n^2 sont indépendants.

Remarques

1. La démonstration de ce théorème n'est pas présentée dans ce cours. Le lecteur pourra la consulter dans le livre du Professeur Saporta intitulé : « Probabilités, Analyse des données et Statistique », Seconde édition, Éditions Technip, pages 282.
2. Ce théorème est connu sous le nom du théorème de Fisher.

Nous pouvons démontrer la réciproque du dernier théorème : il s'agit donc d'une propriété caractéristique.

Théorème

Si les deux estimateurs $\widehat{\mu}_n$ et S_n^2 sont indépendants alors la variable aléatoire X suit une loi normale de paramètres μ et σ^2.

Corollaire

Puisque

$$\sqrt{n}\,\frac{\widehat{\mu}_n-\mu}{\sigma} \text{ suit la } \mathcal{N}(0,1) \text{ et } \frac{nS_n^2}{\sigma^2} \text{ suit une } \chi^2_{n-1}$$

nous avons

$$T_{n-1}=\frac{\sqrt{n}\,\frac{\widehat{\mu}_n-\mu}{\sigma}}{\sqrt{\frac{nS_n^2}{(n-1)\sigma^2}}}=\sqrt{n-1}\,\frac{\widehat{\mu}_n-\mu}{S_n}$$

où T_{n-1} est une variable de Student à $(n-1)$ degrés de liberté.

Remarque

Ce dernier résultat est extrêmement utile car il ne dépend pas de la variance σ^2 de la population U et servira donc chaque fois que la variance σ^2 est inconnue.

Remarque

Nous constaterons l'utilité de ce résultat spécialement lorsque nous allons construire des intervalles de confiance pour l'estimateur $\widehat{\mu}_n$ et que nous ne connaîtrons pas la valeur de la variance σ^2 de la population U.