

Statistique inférentielle I - Estimation

Nathalie Cheze¹

July 9, 2007

1 Introduction

1.1 Notion d'échantillon

Soit un ensemble de taille N , appelé population. On veut étudier la population relativement à un caractère statistique X .

S'il est possible d'interroger tous les éléments de la population, les propriétés de X sont parfaitement connues. Une telle situation est rare et l'étude de X sera réalisée, en général, à partir de données partielles de X .

La notion d'échantillon va formaliser le choix de ces données partielles. Un échantillon de la population est dit de taille n s'il est constitué de n unités distinctes ou non, tirées parmi les N éléments de la population. On lui associe alors l'échantillon des réalisations de X que l'on note (x_1, x_2, \dots, x_n) . C'est sur cet échantillon que seront effectués les calculs.

Il n'est pas à propos dans ce chapitre d'étudier comment a été établi le plan de sondage fournissant l'échantillon. Ceci fait l'objet de la théorie des sondages. Le tirage le plus usité est le tirage aléatoire avec remise. Il consiste à tirer au hasard l'échantillon, unité par unité. Lorsqu'un élément est tiré, il n'est pas éliminé et participe au tirage suivant. Dans la suite, on suppose que l'échantillon a été obtenu de cette manière.

1.2 Caractéristiques d'un échantillon

Supposons que l'on ait tiré un échantillon de taille n . On peut calculer pour la variable quantitative X , certaines caractéristiques, à partir de l'observation (x_1, x_2, \dots, x_n) . On note:

- la moyenne de l'échantillon des données : $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$
- la variance "biaisée" de l'échantillon des données: $s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$
- la variance "sans biais" de l'échantillon des données: $\hat{s}_n^2 = s_n^2 \times \frac{n}{n-1}$.

La valeur de ces paramètres dépend évidemment de l'échantillon étudié. A chaque échantillon de données, correspond une valeur de \bar{x}_n , s_n^2 et \hat{s}_n^2 . L'étude de variabilité des ces valeurs,

¹Université Paris 10, MODALX, 200 av. de la République, 92000 Nanterre, Nathalie.Cheze@u-paris10.fr

suivant l'échantillon choisi, est un problème important que nous abordons maintenant.

1.3 Fluctuations d'échantillonnage

On considère une population de taille N , pour laquelle on veut étudier un caractère X . On peut prélever plusieurs échantillons différents de taille n produisant ainsi des caractéristiques d'échantillonnage (par exemple, la moyenne ou la variance), différentes selon l'échantillon et qui ne correspondent en général pas aux valeurs des paramètres de la population. **Toute la modélisation statistique repose sur l'aléatoire dans le choix de l'échantillon.** Pour cela, nous allons considérer un cadre probabiliste général, dit modèle d'échantillonnage.

1.3.1 Modèle d'échantillonnage

Définition 1.

On appelle **statistique** toute variable aléatoire (en abrégé v.a.).

Définition 2.

Un n -échantillon est une suite (X_1, X_2, \dots, X_n) de n v.a. indépendantes et de même loi, à valeurs dans l'espace E (en général $E = \mathbf{Z}, E = \mathbb{R}$, ou $E = \mathbb{R}^d$).

Remarque: on obtient un n -échantillon en répétant n fois la même expérience dans les mêmes conditions et de manière indépendante les unes des autres. Cela correspond au tirage avec remise.

Exemples de statistiques.

- Soit un n -échantillon (X_1, X_2, \dots, X_n) , $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est une statistique.
- Si on pose $E(X_i) = \theta$, alors $\bar{X}_n - \theta$ n'est pas une statistique (car elle n'est pas calculable à partir des observations).

1.3.2 Exemples fondamentaux

Considérons un n -échantillon (X_1, X_2, \dots, X_n) et notons

- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, la statistique qui, à chaque échantillon (x_1, x_2, \dots, x_n) associe sa moyenne \bar{x}_n ,

- $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2$, la statistique qui, à chaque échantillon (x_1, x_2, \dots, x_n) associe sa variance "biaisée" s_n^2 ,

- $\hat{S}_n^2 = \frac{n}{n-1} S_n^2$, la statistique qui, à chaque échantillon (x_1, x_2, \dots, x_n) associe sa variance "sans biais" \hat{s}_n^2 .

Résumons sous forme de tableau les symboles utilisés pour désigner les paramètres de la population et les caractéristiques de l'échantillon.

	paramètres de la population	caractéristiques de l'échantillon
taille	N	n
moyenne	μ	\bar{x}_n
variance	σ^2	s_n^2

- Propriétés de la moyenne empirique.

En posant $E(X_i) = \mu$ et $Var(X_i) = \sigma^2, \forall i = 1, \dots, n$, on montre que la moyenne \bar{X}_n vérifie:

- a) $E(\bar{X}_n) = \mu$
 b) $Var(\bar{X}_n) = \frac{\sigma^2}{n}$.

Preuve

a) $E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$, d'après les propriétés de l'espérance mathématique.

b) $Var(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$, d'après les propriétés de la variance et le fait que les v.a. (X_i) sont indépendantes entre elles.

L'interprétation statistique de ces propriétés, qui permet d'illustrer la notion de fluctuations d'échantillonnage, est la suivante:

si on fait la moyenne sur tous les échantillons possibles de taille n dans la population de taille N , obtenus par un échantillonnage avec remise, des \bar{x}_n associées à chaque échantillon, on retrouve la vraie valeur μ .

Illustrons cette propriété sur un exemple trivial.

Considérons une population composée de $N = 3$ enfants (E_1, E_2, E_3) âgés respectivement de 6, 2 et 10 ans. On s'intéresse à la variable $X = \text{âge}$.

On a donc $\mu = 6$ et $\sigma^2 = 10.67$.

Supposons maintenant que l'on ne peut pas avoir l'information $X = \text{âge}$, sur toute la population des trois enfants, mais seulement sur un échantillon de deux enfants.

Si on fait un échantillonnage avec remise, on obtient neuf échantillons qui sont les suivants: $(E_1, E_1), (E_1, E_2), (E_1, E_3), (E_2, E_1), (E_2, E_2), (E_2, E_3), (E_3, E_1), (E_3, E_2)$ et (E_3, E_3) .

On peut calculer \bar{x}_n pour chacun des échantillons et on obtient respectivement: 6, 4, 8, 4, 2, 6, 8, 6 et 10.

Evidemment, on obtient des résultats différents pour chaque échantillon, illustrant les fluctuations d'échantillonnage, des différences d'autant plus importantes que la taille de l'échantillon est petite.

Pour comprendre la propriété a), il suffit de faire la moyenne de ces neuf valeurs. On obtient 6 qui est effectivement la valeur de μ , on retrouve ainsi le résultat énoncé.

Pour vérifier la propriété b), il suffit de calculer la variance de ces neuf valeurs et le résultat obtenu est égal à σ^2/n .

Ce chapitre est volontairement détaillé car il est indispensable de bien comprendre la différence entre population et échantillons, paramètres et caractéristiques pour pouvoir aborder la statistique inférentielle développée dans le quatrième chapitre.

2 Lois de probabilités classiques en statistique

L'objectif de ce chapitre est de décrire les lois des variables aléatoires réelles introduites dans la suite de cet article. Les variables aléatoires X , présentées ci-dessous, sont toutes à valeurs réelles, et ont une loi de densité f ne présentant pas de primitive simple. Pour obtenir les quantités de la forme $P(a \leq X \leq b)$, il existe des tables statistiques qui donnent un certain nombre de valeurs de cette fonction. Pour chacun des exemples, nous donnons la table statistique associée et quelques exemples de calculs de probabilité.

2.1 Loi normale unidimensionnelle

La loi normale est une des lois fondamentales en statistique.

Définition.

La variable aléatoire X suit une loi normale centrée réduite, notée $N(0, 1)$, si sa densité f est donnée par:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \forall x \in \mathbb{R}.$$

Définition.

La variable aléatoire X suit une loi normale d'espérance m et de variance σ^2 , notée $N(m, \sigma^2)$, si sa densité f est donnée par:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}.$$

Remarques

- La représentation graphique de la loi normale $N(m, \sigma^2)$ donne la courbe de Gauss (en cloche), symétrique par rapport à m .
- Soit la variable aléatoire X qui suit une loi normale $N(m, \sigma^2)$, alors la v.a. $U = \frac{X - m}{\sigma}$ suit une loi normale $N(0, 1)$, pour laquelle il existe une table statistique qui permet d'obtenir facilement les quantiles. Grâce à ces propriétés de stabilité par transformation affine, on peut obtenir la probabilité $P(X \leq a)$ en écrivant $P(X \leq a) = P(U \leq \frac{a - m}{\sigma})$ et lire la probabilité associée au quantile $\frac{a - m}{\sigma}$ dans la table de loi normale centrée réduite. Nous donnerons ci-après quelques exemples de lecture de la table statistique de la loi $N(0, 1)$.
- Soient X et Y deux v.a. indépendantes suivant respectivement les lois $N(m_X, \sigma_X^2)$ et $N(m_Y, \sigma_Y^2)$. La v.a. $X + Y$ suit une loi normale $N(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$. On peut généraliser ce résultat à un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$ et obtient ainsi que

la v.a. \bar{X}_n suit une loi normale $N(m, \sigma^2/n)$.

Exemples de calculs de quantiles

Considérons une variable aléatoire U de loi $N(0, 1)$. On cherche à calculer, soit la probabilité $P(U \leq u)$, u quantile donné, soit le quantile u , la probabilité étant donnée. La table donne les valeurs de $F(u) = P(U \leq u)$, $u \geq 0$.

- Que vaut $P(U \leq 1,91)$? On décompose 1,91 en 1,9 et 0,01, on se place à l'intersection de la ligne $u = 1,9$ et la colonne $u = 0,01$ et on lit la probabilité associée 0,9719.
- Que vaut $P(U \leq -1,91)$? La table ne donne pas les quantiles négatifs. Mais, du fait que la densité est symétrique, on peut écrire $P(U \leq -1,91) = P(U \geq 1,91)$. La table donne les probabilités associées à l'évènement $U \leq u$ et non $U \geq u$. De plus, on a $P(U \geq 1,91) = 1 - P(U < 1,91)$ et ainsi, $P(U \leq -1,91) = 1 - 0,9719 = 0,0281$.
- Que vaut $P(0,5 < U \leq 1,01)$? On voit facilement que l'évènement $(0,5 < U \leq 1,01)$ peut se décomposer en $(U \leq 1,01) / (U \leq 0,5)$ et ainsi, $P(0,5 < U \leq 1,01) = P(U \leq 1,01) - P(U \leq 0,5) = 0,8438 - 0,6915 = 0,1523$.
- Que vaut u tel que $P(U \leq u) = 0,8315$? Il faut chercher la valeur 0,8315 à l'intérieur de la table et lire le quantile associé en additionnant les valeurs de u associées à la ligne et à la colonne de 0,8315, c'est-à-dire $u = 0,9 + 0,06 = 0,96$.
- Que vaut u tel que $P(U \leq u) = 0,2358$? Cette valeur ne se trouve pas à l'intérieur de la table. La probabilité 0,2358 étant inférieure à 0,5, cela signifie que le quantile est nécessairement négatif. On peut écrire $P(U \leq u) = 0,2358 = 1 - P(U \leq -u)$ et $-u$ tel que $P(U \leq -u) = 1 - 0,2358 = 0,7642$, c'est-à-dire $-u = 0,72$ et $u = -0,72$.

Considérons maintenant une variable aléatoire X de loi $N(20, 36)$.

- Que vaut $P(X \leq 22,4)$? D'après les remarques précédentes, on peut écrire $P(X \leq 22,4) = P(U \leq \frac{22,4 - 20}{6})$ où U est une variable de loi $N(0, 1)$. Ainsi, $P(X \leq 22,4) = P(U \leq 0,4) = 0,6554$.

2.2 Loi du Khi-deux

Définition.

On considère un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(0, 1)$.

On dit que la variable aléatoire $Z = \sum_{i=1}^n X_i^2$ suit une loi du khi-deux à n degrés de liberté, notée $\chi^2(n)$. Son espérance est donnée par $E(Z) = n$, sa variance par $Var(Z) = 2n$ et sa densité f par:

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2} x^{n/2-1}, \quad \forall x \in \mathbb{R}^+$$

où $\Gamma(p) = \int_0^{+\infty} e^{-x} x^{p-1} dx$.

Remarques

- Il existe une table statistique pour obtenir facilement les quantiles de cette loi . Nous donnerons ci-après quelques exemples de lecture de la table statistique de la loi $\chi^2(n)$.

- Soient X et Y deux v.a. indépendantes suivant respectivement les lois $\chi^2(n)$ et $\chi^2(m)$. La v.a. $X+Y$ suit une loi $\chi^2(n+m)$. La preuve est immédiate si l'on revient à la définition.

Considérons maintenant un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$.

- La v.a. $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - m)^2$ suit une loi $\chi^2(n)$.

Le résultat suivant est important.

Théorème: La variable aléatoire $\frac{n-1}{\sigma^2} \hat{S}_n^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ suit une loi du $\chi^2(n-1)$.

Exemples de calculs de quantiles

Considérons une variable aléatoire X de loi $\chi^2(10)$. On cherche à calculer, soit la probabilité $P(X \leq \chi^2)$, χ^2 quantile positif donné, soit le quantile χ^2 , la probabilité étant donnée. La table donne les valeurs de $P = P(X > \chi^2)$.

- Que vaut χ^2 tel que $P(X > \chi^2) = 0,10$? On se place à l'intersection de la ligne degré de liberté $\nu = 10$ et $P = 0,10$ et on obtient $\chi^2 = 15,987$.
- Que vaut χ^2 tel que $P(X \leq \chi^2) = 0,10$? La table donne la probabilité $P(X > \chi^2)$ et non $P(X \leq \chi^2)$, il suffit donc d'écrire $P(X \leq \chi^2) = 1 - P(X > \chi^2)$ et trouver χ^2 tel que $P(X > \chi^2) = 0,9$, c'est-à-dire que $\chi^2 = 4,865$.
- Que vaut $P(X > 2,558)$? Il faut trouver $\chi^2 = 2,558$ à l'intérieur du tableau sur la ligne $\nu = 10$ et on obtient $P(X > 2,558) = 0,99$.

2.3 Loi de Student

Définition: Soient X et Y deux variables aléatoires indépendantes suivant respectivement une loi normale $N(0, 1)$ et une loi du khi-deux $\chi^2(n)$, $n \in \mathbb{N}^*$.

On dit que la variable aléatoire T définie par $T = \frac{X}{\sqrt{Y/n}}$ suit une loi de Student à n degrés de liberté, notée $St(n)$. Son espérance est donnée par $E(T) = 0$, sa variance n'est définie que pour $n > 2$, par $Var(T) = \frac{n}{n-2}$ et sa densité f par:

$$f(x) = \frac{1}{\sqrt{n} B(1/2, n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad \forall x \in \mathbb{R}$$

où $B(p, q) = \int_0^{+\infty} x^{p-1} (1-x)^{q-1} dx$.

Remarques

- Il existe une table statistique pour obtenir facilement les quantiles de cette loi symétrique. Nous donnerons ci-après quelques exemples de lecture de la table statistique de la loi $St(n)$.

- La v.a. T converge en loi vers une variable aléatoire de loi $N(0, 1)$ lorsque n tend vers l'infini.

- On considère un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$.

Par définition, la variable aléatoire $\frac{\sqrt{n}}{S_n^*} (\bar{X}_n - m)$ suit une loi de Student $St(n - 1)$.

Exemples de calculs de quantiles

Considérons une variable aléatoire X de loi $St(10)$. On cherche à calculer, soit la probabilité $P(X \leq t)$, t quantile positif donné, soit le quantile t , la probabilité étant donnée. La table donne les valeurs de $P = P(|X| \geq t)$.

- Que vaut t tel que $P(-t < X \leq t) = 0,10$? On se place à l'intersection de la ligne degré de liberté $\nu = 10$ et $P = P(|X| \geq t) = 0,90$ et on obtient $t = 0,1289$.

- Que vaut t tel que $P(X > t) = 0,10$? On ne peut pas lire directement dans la table, mais du fait de la symétrie de la densité, on peut écrire $P(|X| \geq t) = 2P(X > t) = 0,20$ et ainsi, on cherche t tel que $P = 0,20$, c'est-à-dire $t = 1,3722$.

- Que vaut $P(X > 0,5415)$? A l'intérieur de la table, à la ligne degré de liberté $\nu = 10$, on voit que $P(|X| \geq 0,5415) = 0,60$, on en déduit que $P(X > 0,5415) = \frac{1}{2}(P(|X| \geq 0,5415)) = 0,3$.

2.4 Loi de Fisher-Snedecor

Définition: Soient X et Y deux variables aléatoires indépendantes suivant respectivement une loi du khi-deux $\chi^2(n)$ et une loi du khi-deux $\chi^2(m)$. On dit que la variable aléatoire U définie par $U = \frac{X/n}{Y/m}$ suit une loi de Fisher-Snedecor à n et m degrés de liberté, notée $F(n, m)$. Son espérance est définie pour $m > 2$ et est donnée par $E(U) = \frac{m}{m-2}$. Sa variance est définie pour $m > 4$ et est donnée par $Var(U) = \frac{2m^2(n+m_2)}{m(m-4)(m-2)^2}$ et sa densité f par:

$$f(x) = \frac{1}{B(n/2, m/2)} n^{n/2} m^{m/2} \frac{x^{n/2-1}}{(m + nx)^{\frac{n+m}{2}}}, \forall x \in \mathbb{R}^+$$

Remarques

- Par définition, la v.a. $V = \frac{1}{U}$ suit une loi de Fisher $F(m, n)$. Nous donnerons ci-après quelques exemples de lecture de la table statistique de la loi $F(m, n)$.

- On considère deux échantillons de taille n et m respectivement, indépendants entre eux et notés (X_1, \dots, X_n) et (Y_1, \dots, Y_m) , de lois $N(m_X, \sigma_X^2)$ et $N(m_Y, \sigma_Y^2)$.

Par définition, la variable aléatoire $\frac{\sum_{i=1}^m (Y_i - \bar{Y}_m)^2 / (m - 1) \sigma_Y^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n - 1) \sigma_X^2}$ suit une loi de Fisher $F(m - 1, n - 1)$.

Exemples de calculs de quantiles

Considérons une variable aléatoire X de loi $F(3, 2)$. Les tables présentées ici permettent de calculer le quantile F , pour la probabilité $P(X \geq F) = 0,05$ et $P(X \geq F) = 0,01$ d'une loi de Fisher $F(\nu_1, \nu_2)$. Il existe évidemment des tables associées à d'autres probabilités.

- Que vaut F tel que $P(X \leq F) = 0,95$? Cela revient à chercher $P(X > F) = 0,05$. Il suffit de se placer à l'intersection de la ligne degré de liberté du dénominateur égale à 2 et de la colonne degré de liberté du numérateur égale à 3 dans la première table et, on obtient $F = 19,164$.
- Que vaut F tel que $P(X \leq F) = 0,99$? Il suffit de se placer à l'intersection de la ligne degré de liberté du dénominateur égale à 2 et de la colonne degré de liberté du numérateur égale à 3 dans la deuxième table et, on obtient 99,164.
- Que vaut F tel que $P(X \geq F) = 0,95$? La table donne seulement les probabilité $P(X > F) = 0,95$. On a vu, dans les remarques précédentes que la v.a. $\frac{1}{X}$ suit une loi $F(2, 3)$. Ainsi, on a $P(X \geq F) = P(\frac{1}{X} \leq \frac{1}{F}) = 0,95$ et $\frac{1}{F} = 9,552$, $F = \frac{1}{9,552} = 0,1047$.

3 Estimation

3.1 Introduction

Le problème de la statistique est le suivant; on observe une variable aléatoire, dont la loi est - complètement ou partiellement - inconnue. Il s'agit, à partir de l'échantillon des observations, d'obtenir le plus d'informations possibles sur cette loi.

Par exemple, supposons que l'on fabrique des pièces sur une machine. Toutes les pièces fabriquées ont la même probabilité θ inconnue d'être défectueuses. Ce nombre θ dépend du réglage de la machine, le réglage est d'autant meilleur que θ est proche de 0; mais comme le réglage ne peut être parfait, on n'a jamais $\theta = 0$. Avant de lancer le cycle de fabrication, on veut vérifier si la machine est "bien réglée", c'est-à-dire si θ est suffisamment petit. Pour cela, on fabrique un certain nombre n de pièces qui servent à tester le réglage. L'observation consiste à compter le nombre X de pièces défectueuses parmi ces n pièces. On peut alors se poser deux types de problèmes.

- 1) Trouver "la" valeur de θ : cela s'appelle **estimer** le paramètre θ . Dans notre exemple, il est naturel de prendre pour **estimateur** de θ la proportion $\hat{\theta}_n = X/n$ de pièces défectueuses.
- 2) S'assurer que la vraie valeur de θ ne dépasse pas un seuil critique θ_0 fixé à l'avance (sinon, il faut refaire le réglage de la machine): cela s'appelle **tester** le fait que $\theta \leq \theta_0$.

Ces deux problèmes sont de natures mathématiques assez différentes. Ils ont cependant en commun le fait **qu'on ne peut pas arriver à une conclusion certaine**. Dans le cas 1), il est "vraisemblable" que la valeur exacte de θ soit proche de l'estimation $\hat{\theta}_n$ (au moins si n est assez grand), mais tout-à-fait invraisemblable qu'elle lui soit exactement égale (voir le paragraphe 2.3.). Dans le cas 2), on peut décider que $\theta \leq \theta_0$ si la proportion X/n est suffisamment petite, mais on ne sera jamais sûr que la vraie valeur de θ soit effectivement plus petite que le seuil θ_0 .

Dans tous les cas, on cherche, à partir de la connaissance d'un n -échantillon d'observations, à obtenir des renseignements sur la valeur inconnue (mais non aléatoire) du paramètre θ :

- 1) Soit on veut déterminer la valeur de θ , ou d'une fonction $g(\theta)$, g connue; il s'agit alors **d'estimation**.
- 2) Soit on veut savoir si θ se trouve dans une partie Θ_0 de l'ensemble Θ , ou dans son complémentaire: il s'agit alors d'un **test statistique**.

3.2 Estimation d'un paramètre unidimensionnel

Dans la suite, on considère un n -échantillon (X_1, X_2, \dots, X_n) de loi P_θ . On cherche à estimer la quantité inconnue θ à partir de l'échantillon (X_1, X_2, \dots, X_n) .

3.2.1 Estimation ponctuelle

A- Estimateurs sans biais, efficacité

Estimer θ veut dire qu'au vu des observations (x_1, x_2, \dots, x_n) , on "décide" que la valeur de θ vaut un certain nombre $T(x_1, x_2, \dots, x_n)$.

Définition 4.

On appelle **estimateur** de θ une statistique $T(X_1, \dots, X_n)$. La valeur de T en l'échantillon observé (x_1, x_2, \dots, x_n) est **l'estimation** du paramètre.

Le problème de l'estimation consiste à optimiser le choix de l'estimateur et pour cela, il faut introduire un critère de qualité. Pour mesurer l'erreur commise en remplaçant θ par T , on utilise en général le risque quadratique défini par

Définition 5.

Le **risque quadratique** de l'estimateur T de θ est défini par

$$R_T(\theta) = E_\theta[(T - \theta)^2],$$

Si S et T sont deux estimateurs de θ , on dit que S est **meilleur** que T au sens du risque quadratique si $R_S(\theta) \leq R_T(\theta)$ pour tout $\theta \in \Theta$.

La détermination d'un **meilleur estimateur** est un problème mathématique extrêmement difficile, car la classe de tous les estimateurs est très vaste. Dans la plupart des cas, on se restreint à une classe particulière d'estimateurs, et notamment à la classe suivante.

Définition 6.

a) L'estimateur T de θ est dit **sans biais** (ou non-biaisé) si

$$E_\theta(T) = \theta, \quad \forall \theta \in \Theta$$

Dans ce cas, le risque quadratique $R_T(\theta)$ est égal à la variance de T sous P_θ .

b) L'estimateur $T = T(X_1, \dots, X_n)$ de θ est dit asymptotiquement **sans biais** si

$$\lim_{n \rightarrow +\infty} E_\theta(T) = \theta, \quad \forall \theta \in \Theta$$

c) On dit que T est un **meilleur estimateur sans biais** de $g(\theta)$ s'il est sans biais et s'il est meilleur que tout autre estimateur sans biais au sens de la variance.

d) On dit que T est un estimateur **convergent** de $g(\theta)$ s'il converge en probabilité vers θ quand $n \rightarrow +\infty$.

Remarques.

- Comprendre l'utilité d'un estimateur sans biais: l'idée est simple: si on prélève un grand nombre d'échantillons de même taille, la moyenne des estimations obtenues donnera une valeur assez exacte du paramètre de la population.

- Une condition nécessaire et suffisante pour que T converge en probabilité vers θ est qu'il soit sans biais, ou asymptotiquement sans biais et tel que $\lim_{n \rightarrow +\infty} Var(T) = 0$.

- On peut comprendre facilement l'intérêt d'un estimateur convergent : il faut non seulement que la moyenne des estimations soit proche de θ , mais aussi que chaque estimation

soit le plus proche possible de θ , donc que la variabilité de l'estimateur soit faible, pour n suffisamment grand.

- Un estimateur biaisé peut être intéressant si son risque quadratique est inférieur à la variance d'un estimateur sans biais.

Exemples.

• Considérons un n -échantillon (X_1, X_2, \dots, X_n) de moyenne μ et de variance σ^2 . On a les résultats suivants:

(a) $E(\bar{X}_n) = \mu$ et $Var(\bar{X}_n) = \frac{\sigma^2}{n}$. L'estimateur \bar{X}_n est donc un estimateur sans biais et convergent de μ .

(b) $E(\hat{S}_n^2) = \sigma^2$. La variance empirique modifiée est un estimateur sans biais de la variance théorique.

Preuve

On peut écrire $E(\hat{S}_n^2) = \frac{n}{n-1} E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right) = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}_n)^2\right) = \frac{n}{n-1} ((\sigma^2 + \mu^2) - (\frac{\sigma^2}{n} + \mu^2)) = \sigma^2$.

(c) $E(S_n^2) = \frac{n-1}{n} \sigma^2$ (car $S_n^2 = \frac{n-1}{n} \hat{S}_n^2$). La variance empirique est un estimateur biaisé mais asymptotiquement sans biais de la variance théorique.

On peut dire que la variance empirique modifiée est un estimateur convergent de la variance théorique.

Comparer deux estimateurs sans biais de θ au sens du risque quadratique revient à comparer leur variance. Il est donc logique de se poser la question de l'existence d'une borne inférieure à l'ensemble des variances des estimateurs sans biais de θ . Pour cela, on doit définir la fonction de vraisemblance.

Définition 7.

On se place dans le cadre d'hypothèses suivant :

on appelle vraisemblance du paramètre θ , l'application $L : \Theta \rightarrow \mathbb{R}^+$ définie par:

- cas (a): si E est fini ou dénombrable, $L(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n P_\theta(x_i)$

- cas (b): si $E = \mathbb{R}$ ou \mathbb{R}^d et P_θ admet une densité f_θ , $L(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i)$.

On donne maintenant un résultat sur l'existence d'une borne inférieure à l'ensemble des variances des estimateurs sans biais de θ .

Théorème Supposons les hypothèses dites de Cramer-Rao vérifiées (drivabilit de $L(x_1, x_2, \dots, x_n, \theta)$ par rapport θ, \dots) .

Soit T un estimateur sans biais de θ , de variance finie.

$$\forall \theta \in \Theta, \text{Var}(T) \geq \frac{1}{I_n(\theta)}$$

où $I_n(\theta)$ l'**information de Fisher** de θ donnée par l'échantillon, est définie par :

$$I_n(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \text{Log}L(X, \theta) \right)^2 \right] = - E_\theta \left[\frac{\partial^2}{\partial \theta^2} \text{Log}L(X, \theta) \right]$$

L'information de Fisher mesure la quantité d'information sur θ contenue dans l'échantillon X .

Définition

La quantité $\frac{g'^2(\theta)}{I_n(\theta)}$ s'appelle **la borne de Fréchet-Cramer-Rao** (en abrégé FCR).

Remarques.

- Lorqu'on veut estimer le paramètre θ , c'est à dire $g(\theta) = \theta$, on a $\text{Var}_\theta(T) \geq \frac{1}{I_n(\theta)}$.
- On ne peut pas affirmer qu'il existe un estimateur sans biais dont la variance atteint la borne FCR mais, par contre, on peut affirmer qu'il n'existe pas d'estimateurs sans biais meilleurs, au sens du risque quadratique, que celui qui atteint la borne FCR.
- Si la valeur de la borne de FCR est grande, cela signifie que l'on ne peut pas estimer correctement $g(\theta)$.
- On peut démontrer facilement que $I_n(\theta) = nI_1(\theta)$.

Définition On dit qu'un estimateur sans biais de $g(\theta)$ est **efficace** si sa variance atteint la borne FCR.

Exemples.

• Considérons un n-échantillon (X_1, X_2, \dots, X_n) de loi Bernoulli de paramètre p . L'estimateur \bar{X}_n est un estimateur sans biais et efficace de p . En effet, des calculs simples donnent :

$$E_p(\bar{X}_n) = p, \text{Var}(\bar{X}_n) = \frac{p(1-p)}{n}, \quad I_n(p) = \frac{n}{p(1-p)}$$

et ainsi $\text{Var}(\bar{X}_n) = \frac{1}{I_n(p)}$.

Preuve

La fonction de vraisemblance est donnée par:

$$L(x_1, \dots, x_n; p) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} = \prod_{i=1}^n 1_{x_i \in \{0,1\}}$$

En dérivant deux-fois par rapport à p la log-vraisemblance $\text{Log}(L(x_1, \dots, x_n; p))$, on obtient:

$$\frac{\partial^2}{\partial p^2} \text{Log}(L(x_1, \dots, x_n; p)) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2}$$

pour $x_i \in \{0, 1\}$, $i = 1, \dots, n$.

L'information de Fisher est donnée par:

$$I_n(p) = -E_p\left(-\frac{\sum_{i=1}^n X_i}{p^2} - \frac{n - \sum_{i=1}^n X_i}{(1-p)^2}\right) = \frac{n}{p(1-p)}$$

• Considérons un n-échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, où σ^2 connue. L'estimateur \bar{X}_n est un estimateur sans biais et efficace de m . On peut facilement montrer que:

$$E(\bar{X}_n) = m, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{1}{I_n(m)}.$$

Preuve

En dérivant deux-fois par rapport à m la log-vraisemblance

$$\text{Log}L(x_1, \dots, x_n; m) = -\text{Log}(\sigma^n (2\pi)^{n/2}) - \frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}, \quad \forall x_i \in \mathbb{R}, i = 1, \dots, n, \text{ on obtient:}$$

$$\frac{\partial^2}{\partial m^2} \text{Log}(L(x_1, \dots, x_n; m)) = -\frac{n}{\sigma^2}$$

L'information de Fisher est donnée par:

$$I_n(m) = -E\left(-\frac{n}{\sigma^2}\right) = \frac{n}{\sigma^2}$$

• Considérons un n-échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, m connue. L'estimateur \hat{S}_n^2 est un estimateur sans biais, non efficace de σ^2 . On peut montrer que:

$$E(\hat{S}_n^2) = \sigma^2, \quad \text{Var}(\hat{S}_n^2) = \frac{2\sigma^4}{n-1} \neq \frac{1}{I_n(\sigma^2)} = \frac{2\sigma^4}{n}.$$

Preuve

En dérivant deux-fois par rapport à σ^2 la log-vraisemblance

$$\text{Log}L(x_1, \dots, x_n; \sigma^2) = -\text{Log}(\sigma^n (2\pi)^{n/2}) - \frac{\sum_{i=1}^n (x_i - m)^2}{2\sigma^2}, \quad \forall x_i \in \mathbb{R}, i = 1, \dots, n, \text{ on obtient:}$$

$$\frac{\partial^2}{\partial (\sigma^2)^2} \text{Log}(L(x_1, \dots, x_n; \sigma^2)) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - m)^2$$

L'information de Fisher est donnée par:

$$I_n(\sigma^2) = -E\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (X_i - m)^2\right) = \frac{n}{2\sigma^4}$$

Remarque Il n'y a pas de résultat sur les propriétés de \hat{S}_n comme estimateur de l'écart-type σ .

Exemple pratique

On souhaite estimer la proportion p des électeurs satisfaits du travail du premier ministre. Pour cela, on procède à un sondage d'opinion en prélevant deux échantillons indépendants entre eux et on demande, à chaque individu, s'il est satisfait du travail du premier ministre. $n = 200$ électeurs ont été choisis au hasard et avec remise, dans une circonscription rurale et $m = 300$ électeurs ont été choisis au hasard et avec remise, dans une circonscription urbaine. 150 ont répondu oui dans le premier échantillon et 200 dans le second. On appelle X_i la v.a. qui prend la valeur 1 si le i -ème individu du premier échantillon a répondu oui avec la probabilité p et 0 sinon avec la probabilité $1 - p$. On appelle Y_i la v.a. qui prend la valeur 1 si le i -ème individu du second échantillon a répondu oui avec la probabilité p et 0 sinon avec la probabilité $1 - p$.

On dispose donc d'un $(n + m)$ -échantillon $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ de loi de Bernoulli p . On cherche le meilleur estimateur de p à partir de ces données. D'après ce qui précède, $E(\bar{X}_n) = p$ (et resp. $E(\bar{Y}_m) = p$), on cherche donc un estimateur sans biais de p , fonction de \bar{X}_n et \bar{Y}_m . Considérons l'estimateur $\bar{U} = a\bar{X}_n + b\bar{Y}_m$, $a, b > 0$. Cherchons les constantes positives a et b telles que \bar{U} soit un estimateur sans biais de p et de variance minimale.

Comme $E(\bar{U}) = (a + b)p$, il suffit de choisir $a + b = 1$ pour que \bar{U} soit un estimateur sans biais de p . De plus, $Var(\bar{U}) = \left(\frac{a^2}{n} + \frac{b^2}{m}\right) (p(1 - p))$. Trouver a et b telles que \bar{U} soit un estimateur sans biais de p et de variance minimale revient à minimiser la fonction $f(a, b) = \left(\frac{a^2}{n} + \frac{b^2}{m}\right)$ sous la contrainte $a + b = 1$. C'est un problème mathématique classique et la solution est donnée par $a = \frac{n}{n+m}$ et $b = \frac{m}{n+m}$. Ainsi, le meilleur estimateur sans biais de p est $\bar{U} = \frac{n\bar{X}_n + m\bar{Y}_m}{n+m}$ de variance $\frac{p(1-p)}{n+m}$. D'après l'exemple précédent, l'information de Fisher de p donnée par l'échantillon $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ est $I_n(p) = \frac{n+m}{p(1-p)}$, qui est égale à $\frac{1}{Var(\bar{U})}$. L'estimateur \bar{U} est donc efficace et on ne peut pas en trouver un meilleur au sens du risque quadratique.

L'estimation de p est, donc $\frac{200\bar{x}_{200} + 300\bar{y}_{300}}{500}$ avec $\bar{x}_{200} = \frac{150}{200} = 0,75$ et $\bar{y}_{300} = \frac{200}{300} = 0,67$, c'est-à-dire 0,7.

B- Exhaustivité

Les résultats précédents donnent des directions pour choisir un "bon" estimateur, mais l'on doit vérifier si ce choix est correct. La notion **d'exhaustivité** va nous permettre de répondre partiellement à cette question.

Définition 8.

Une statistique T est dite **exhaustive** pour le paramètre θ si la loi conditionnelle du n -échantillon (X_1, \dots, X_n) sous P_θ sachant $T(X_1, \dots, X_n)$ est indépendante du paramètre θ .

En général, les calculs sur les lois conditionnelles sont assez compliqués et par conséquent, montrer qu'une statistique est exhaustive, peut être un problème délicat à résoudre. Ce problème peut être facilité grâce à un théorème d'utilisation très simple.

Théorème de factorisation

Une statistique T à valeurs dans \mathbb{R}^p est **exhaustive** pour le paramètre θ , dès qu'il existe une fonction h de Ω dans \mathbb{R}_+ et des fonctions q_θ mesurables de \mathbb{R}^p dans $]0, \infty[$ telles que

- dans le cas (a) (de la définition 7), on ait

$$P_\theta(\{x_1, x_2, \dots, x_n\}) = q_\theta(T(x_1, x_2, \dots, x_n))h(x_1, x_2, \dots, x_n),$$

- dans le cas (b) (de la définition 7), pour chaque θ , la probabilité P_θ admette la densité suivante sur \mathbb{R}^d :

$$f_\theta(x_1, x_2, \dots, x_n) = q_\theta(T(x_1, x_2, \dots, x_n))h(x_1, x_2, \dots, x_n).$$

Remarque. Il n'y a aucune raison pour que la statistique exhaustive soit unique.

Du point de vue statistique, la notion d'exhaustivité présente un intérêt grâce au résultat suivant, que nous admettrons également sans démonstration.

Proposition.

Soit T une statistique exhaustive de $g(\theta)$. Soit S un estimateur (resp. un estimateur sans biais) de $g(\theta)$. Il existe alors un estimateur S' fonction de T qui est meilleur que S (resp. et de plus, sans biais).

Ce résultat montre qu'un bon estimateur $g(\theta)$ est nécessairement une fonction de T . Cela justifie la signification du terme "exhaustif".

Exemples.

Reprenons les exemples précédents.

- Considérons un n-échantillon (X_1, X_2, \dots, X_n) de loi de Bernoulli de paramètre p .

On peut factoriser $P_p(\{x_1, x_2, \dots, x_n\}) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \prod_{i=1}^n 1_{x_i \in \{0,1\}}$, en posant

$$\begin{aligned} h(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n 1_{x_i \in \{0,1\}} \\ q_p(T) &= p^T (1-p)^{n-T} \\ T(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n x_i \end{aligned}$$

On obtient donc que $T = \sum_{i=1}^n X_i$ est une statistique exhaustive de p . L'estimateur \bar{X}_n sans biais et efficace de p , étudié précédemment, est bien fonction de cette statistique.

- Considérons un n-échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$.

On peut factoriser $f_{m,\sigma^2}(x_1, x_2, \dots, x_n) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2)$ en posant

$$h(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2}}$$

$$q_{m,\sigma^2}(u, v) = \frac{1}{\sigma^n} \exp\left(-\frac{n}{2\sigma^2}(v - 2mu + m^2)\right)$$

$$T(x_1, x_2, \dots, x_n) = (u(x_1, x_2, \dots, x_n), v(x_1, x_2, \dots, x_n)) \text{ avec}$$

$$u(x_1, x_2, \dots, x_n) = \bar{x}_n \text{ et } v(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i^2$$

On obtient ainsi, que $T = (\bar{X}_n, \frac{1}{n} \sum_{i=1}^n X_i^2)$ est une statistique exhaustive de (m, σ^2) . Les estimateurs \bar{X}_n et S_n^{*2} sans biais de m et σ^2 , étudiés précédemment sont bien fonctions de cette statistique.

C- Estimation par maximum de vraisemblance

La méthode du maximum de vraisemblance est l'une des plus usitées. La méthode est simple à mettre en place.

Définition 9.

Un estimateur T est appelé **estimateur du maximum de vraisemblance** de θ si

$$L(x_1, x_2, \dots, x_n; T(x_1, x_2, \dots, x_n)) = \sup_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta) \quad \forall x \in \Omega$$

(la fonction L est la "fonction de vraisemblance" définie dans la définition 7).

Remarques.

- Pour comprendre le principe de l'estimateur du maximum de vraisemblance, plaçons nous dans le cas discret, c'est-à-dire que $L(x_1, x_2, \dots, x_n; \theta) = P_\theta(x_1, x_2, \dots, x_n)$. Cette fonction représente la probabilité que l'on observe (x_1, x_2, \dots, x_n) quand la vraie valeur du paramètre est θ . Pour certaines valeurs de θ , cette probabilité sera petite, c'est-à-dire qu'il y a peu de chances d'observer (x_1, x_2, \dots, x_n) . Pour d'autres valeurs de θ , cette probabilité sera grande c'est-à-dire qu'il y a de fortes chances d'observer (x_1, x_2, \dots, x_n) . Par conséquent, la valeur "vraisemblable" pour θ est la valeur pour laquelle la probabilité d'observer (x_1, x_2, \dots, x_n) est la plus grande.

- Un tel estimateur n'existe pas toujours (la fonction $\theta \mapsto L(x_1, x_2, \dots, x_n; \theta)$ pouvant ne pas atteindre son supremum), et il peut aussi être multiple (pire: il peut ne pas exister pour certaines valeurs de x_1, x_2, \dots, x_n , et être multiple pour d'autres!). Cependant, dans la plupart des cas concrets, cet estimateur existe, est unique, et facile à calculer.

- Il n'y a pas vraiment de justification théorique à cette méthode, si ce n'est asymptotiquement: sous des hypothèses très générales, l'estimateur du maximum de vraisemblance T de θ vérifie:

(a) T converge en probabilité vers θ , quand n tend vers l'infini.

(b) $h(T)$ est l'estimateur du maximum de vraisemblance de $h(\theta)$, où h est une fonction quelconque. De plus, si h est dérivable, on a :

$$\sqrt{\frac{I_n(\theta)}{h'(\theta)^2}}(h(T) - h(\theta)) \text{ converge en loi vers } N(0, 1).$$

- Il n'y a aucune raison pour que cet estimateur soit sans biais ou efficace. Par contre, d'après la remarque (b) précédente, cet estimateur est asymptotiquement sans biais et

asymptotiquement efficace.

- S'il existe une statistique exhaustive dans le modèle, alors l'estimateur du maximum de vraisemblance est fonction de cette statistique.

Exemples. Reprenons les exemples précédents pour lesquels la fonction de vraisemblance $L(x; \theta)$ est deux fois dérivable en $\theta \in \mathbb{R}$. Alors, un estimateur du maximum de vraisemblance T qui maximise la fonction $\theta \rightarrow L(x; \theta) > 0$ maximise aussi la fonction $\theta \rightarrow \text{Log}(L(x; \theta))$. Il est donc à chercher parmi les estimateurs qui annulent la dérivée en θ de $\text{Log}(L(x; \theta))$, et de dérivée seconde négative en ce point.

- Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi Bernoulli $Be(p)$. L'unique estimateur du maximum de vraisemblance de p est donné par \bar{X}_n .

Preuve En reprenant les calculs du paragraphe 4.2.1A), la fonction de vraisemblance s'écrit

$$L(x_1, \dots, x_n; p) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n 1_{x_i \in \{0,1\}}$$

En dérivant une fois par rapport à p la log-vraisemblance $\text{Log}(L(x_1, \dots, x_n; p))$, on obtient:

$$\frac{\partial}{\partial p} \text{Log}(L(x_1, \dots, x_n; p)) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{(1-p)}$$

pour $x_i \in \{0, 1\}$, $i = 1, \dots, n$.

La valeur de p qui annule cette dérivée est $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$, qui peut être l'abscisse d'un extremum de la fonction. Vérifions que l'on obtient un maximum en calculant la dérivée seconde.

En dérivant deux fois par rapport à p la log-vraisemblance $\text{Log}(L(x_1, \dots, x_n; p))$, on obtient:

$$\frac{\partial^2}{\partial p^2} \text{Log}(L(x_1, \dots, x_n; p)) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2}$$

pour $x_i \in \{0, 1\}$, $i = 1, \dots, n$.

Calculons la valeur de cette dérivée seconde au point \bar{x}_n :

$$\frac{\partial^2}{\partial p^2} \text{Log}(L(x_1, \dots, x_n; p)) = -\frac{n^3}{\left(\sum_{i=1}^n x_i\right)\left(n - \sum_{i=1}^n x_i\right)}$$

Elle prend une valeur négative, \bar{x}_n est donc l'abscisse d'un maximum et, ainsi, l'unique estimateur du maximum de vraisemblance de p est donné par \bar{X}_n .

- Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, σ connu. L'unique estimateur \bar{X}_n est l'estimateur du maximum de vraisemblance de m . (Pour démontrer ce

résultat, on procède de la même manière que précédemment).

- Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, m connue. L'unique estimateur $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ est l'estimateur du maximum de vraisemblance de σ^2 .

- Donnons un exemple pour lequel la fonction de vraisemblance $L(x; \theta)$ n'est pas dérivable en $\theta \in \mathbb{R}$. Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi uniforme sur $[0, \theta]$, $\theta > 0$, c'est-à-dire de densité $\frac{1}{\theta} 1_{[0, \theta]}(x)$. La fonction de vraisemblance est donnée par:

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n 1_{[0, \theta]}(x_i)$$

Cette fonction de vraisemblance n'est donc pas dérivable en θ . C'est une fonction décroissante de θ . Elle prend donc sa valeur maximale pour θ le plus petit possible. Comme $\theta \geq x_i$, $i = 1, \dots, n$ c'est-à-dire $\theta \geq \sup_{i=1, \dots, n} x_i$, l'estimateur du maximum de vraisemblance est donné par $\hat{\Theta}_n = \sup_{i=1, \dots, n} X_i$.

Exemple pratique

L'autonomie d'une batterie d'un téléphone portable peut être représentée par une variable aléatoire X de densité $f_a(x) = ae^{-ax}$ si $x \geq 0$, 0 sinon, où a est un paramètre positif dont on cherche à estimer la valeur. On dit que X suit une loi exponentielle de paramètre a . On suppose, que pour $n = 100$ batteries, on a observé $\sum_{i=1}^n x_i = 3128$ heures.

N'ayant pas d'idée a priori sur l'estimateur à utiliser, on va chercher l'estimateur du maximum de vraisemblance de a et étudier ses propriétés.

La fonction de vraisemblance s'écrit, pour tout $x_i \geq 0$, $i = 1, \dots, n$

$$L(x_1, \dots, x_n; a) = a^n e^{-a \sum_{i=1}^n x_i}$$

En dérivant une fois par rapport à a la log-vraisemblance $\text{Log}(L(x_1, \dots, x_n; a))$, on obtient:

$$\frac{\partial}{\partial a} \text{Log}(L(x_1, \dots, x_n; a)) = \frac{n}{a} - \sum_{i=1}^n x_i$$

pour $x_i \geq 0$, $i = 1, \dots, n$.

La valeur de a qui annule cette dérivée est $\frac{1}{\bar{x}_n} = \frac{n}{\sum_{i=1}^n x_i}$, abscisse d'un extremum de la fonction. Vérifions que l'on obtient un maximum en calculant la dérivée seconde.

En dérivant deux fois par rapport à a la log-vraisemblance $\text{Log}(L(x_1, \dots, x_n; a))$, on obtient:

$$\frac{\partial^2}{\partial a^2} \text{Log}(L(x_1, \dots, x_n; a)) = -\frac{n}{a^2}$$

pour $x_i \geq 0$, $i = 1, \dots, n$.

Calculons la valeur de cette dérivée seconde au point $\frac{1}{\bar{x}_n}$:

$$\frac{\partial^2}{\partial a^2} \text{Log}(L(x_1, \dots, x_n; a)) = -\frac{\left(\sum_{i=1}^n x_i\right)^2}{n}$$

Elle prend une valeur négative, $\frac{1}{\bar{x}_n}$ est donc l'abscisse d'un maximum et, ainsi, l'unique estimateur du maximum de vraisemblance de a est donné par $\frac{1}{\bar{X}_n}$. De plus, l'information de Fisher de a vaut $I_n(a) = \frac{n}{a^2}$.

Étudions les propriétés de cet estimateur. Vérifions tout d'abord si c'est un estimateur sans biais de a . Pour cela, calculons $E\left(\frac{1}{\bar{X}_n}\right)$.

Notons $Y = \frac{1}{\sum_{i=1}^n X_i}$, cela revient à calculer $E\left(\frac{n}{Y}\right)$.

Il est facile de montrer que, pour un n -échantillon (X_1, \dots, X_n) de loi exponentielle de paramètre a , la v.a. $Y = \sum_{i=1}^n X_i$ suit une loi Gamma de paramètre (n, a) , et de densité

$$f(y) = \frac{a^n}{\Gamma(n)} e^{-ay} y^{n-1} \text{ si } y \geq 0, 0 \text{ sinon,}$$

$$\text{où } \Gamma(n) = \int_0^{+\infty} e^{-x} x^{n-1} dx.$$

On a donc $E\left(\frac{1}{Y}\right) = \int_0^{+\infty} \frac{a^n}{\Gamma(n)} e^{-ay} y^{n-2} dy$. Par une intégration par parties et le fait que $\int f(y) dy = 1$, on obtient que $E\left(\frac{1}{Y}\right) = \frac{a}{n-1}$ et $E\left(\frac{1}{\bar{X}_n}\right) = \frac{n}{n-1} a$.

L'estimateur du maximum de vraisemblance $\frac{1}{\bar{X}_n}$ de a est donc un estimateur biaisé de a . On peut en déduire un estimateur sans biais de a , noté $\hat{a}_n = \frac{n-1}{\sum_{i=1}^n X_i}$.

Vérifions que cet estimateur est convergent. Pour cela, il suffit de vérifier que $\text{Var}(\hat{a}_n)$ tend vers 0 quand $n \rightarrow +\infty$.

Par des calculs d'intégrales, on peut montrer que $\text{Var}\left(\left(\frac{1}{Y^2}\right)\right) = \frac{a^2}{(n-1)^2(n-2)}$ et ainsi, $\text{Var}(\hat{a}_n) = \frac{a^2}{(n-2)}$. On peut donc dire que \hat{a}_n est un estimateur convergent de a .

Il n'est pas efficace car $\text{Var}(\hat{a}_n) \neq \frac{1}{I_n(a)}$.

En conclusion, une bonne estimation de a est donnée par $\frac{n-1}{\sum_{i=1}^n x_i} = 0,316$.

3.2.2 Estimation par intervalle de confiance

Le risque quadratique $R_T(\theta) = E_\theta[(T - g(\theta))^2]$, qui mesure l'erreur commise en remplaçant $g(\theta)$ par T , dépend encore de la valeur inconnue θ . Pour pallier ce problème, on préfère souvent utiliser la notion d'estimation par intervalle de confiance, qui consiste à déterminer un intervalle recouvrant "vraisemblablement" la vraie valeur inconnue θ . Il faut, bien sûr, donner une signification au terme "vraisemblablement".

Définition 10.

On appelle **intervalle de confiance** pour le paramètre $g(\theta)$ de niveau de confiance $1 - \alpha$, $\alpha \in]0, 1[$, un intervalle aléatoire $I(X_1, X_2, \dots, X_n)$ dans lequel " $g(\theta)$ se trouve avec une

probabilité au moins égale à $1 - \alpha$, ce qui veut dire mathématiquement que

$$P_{\theta}(g(\theta) \in I(X_1, X_2, \dots, X_n)) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$$

Remarques

- Le nombre α est la probabilité que le paramètre $g(\theta)$ n'appartienne pas à l'intervalle I , c'est-à-dire la probabilité que l'on se trompe en affirmant $\theta \in I$. C'est donc une probabilité d'erreur, que l'on fixe proche de 0: typiquement 0,1 ou 0,05 ou 0,01.
- On s'est placé dans le cas $\theta \in \mathbb{R}$, ce qui nous permet de parler d'intervalle de confiance. Dans le cas $\theta \in \mathbb{R}^d, d > 1$, on parle de région de confiance.

Détaillons maintenant, sur un exemple, la démarche à suivre pour obtenir un intervalle de confiance :

- *Intervalle de confiance du paramètre p d'une loi de Bernoulli $Be(p)$* , pour un échantillon de grande taille, $n > 30$ et $p(1 - p) > 5$.

On considère un échantillon (X_1, X_2, \dots, X_n) de loi de Bernoulli $Be(p)$. La démarche à suivre pour obtenir un intervalle de confiance pour p est la suivante.

a)- Chercher un estimateur de p ayant de bonnes propriétés.

Dans le paragraphe 4.2.1, on a vu que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur sans biais et efficace de p .

b)- Donner la loi de l'estimateur et introduire une nouvelle v.a. dite **pivotale** T_n dépendant de l'estimateur et du paramètre p , de loi indépendante du paramètre p .

D'après le théorème de la limite centrale, la variable aléatoire $T_n = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ converge vers une v.a. de loi normale centrée, réduite $N(0, 1)$. Dans la pratique, on considère, que T_n suit approximativement une loi $N(0, 1)$ dès que $n > 30$ et $p(1 - p) > 5$.

c)- Construire l'intervalle de confiance à partir de T_n .

Soit $\alpha \in]0, 1[$ donné, u_{α} désigne le fractile d'ordre α de la loi $N(0, 1)$, c'est-à-dire que $P(T_n \leq u_{\alpha}) = \alpha$. On peut écrire:

$$P\left(u_{\alpha_1} \leq \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq u_{1-\alpha_2}\right) = 1 - \alpha$$

où $\alpha = \alpha_1 + \alpha_2$.

La loi normale étant une loi symétrique, unimodale, on choisit $u_{\alpha_1} = -u_{1-\alpha/2}$ et $u_{1-\alpha_2} = u_{1-\alpha/2}$ et on a, pour $\alpha \in]0, 1[$ donné:

$$P\left(-u_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq u_{1-\alpha/2}\right) = 1 - \alpha$$

et, ainsi, l'intervalle de confiance pour p , est donné par

$$P\left(\bar{X}_n - u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

En substituant, de part et d'autre de l'inégalité, le paramètre p par son estimation \bar{x}_n , ce qui peut être totalement justifié asymptotiquement de façon théorique, on obtient finalement une fourchette d'estimation asymptotique pour p , à partir de l'échantillon des observations (x_1, x_2, \dots, x_n) :

$$\bar{x}_n - t_{1-\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \leq p \leq \bar{x}_n + t_{1-\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}.$$

Exemple pratique

On veut évaluer la proportion p de foyers d'un département disposant d'un téléviseur et désireux de recevoir les émissions par câble. Ne voulant pas procéder à un recensement complet, on se propose d'estimer cette proportion à partir d'un échantillon de taille $n = 100$, prélevé au hasard et avec remise dans la population du département. Sur les 100 foyers interrogés, 64 foyers sont désireux de recevoir les émissions par câble. On appelle X_i la v.a. qui prend la valeur 1 si le i -ème foyer est désireux de recevoir les émissions par câble et 0 sinon. On veut déterminer un intervalle de confiance pour p de niveau de confiance 0,95.

On est dans les conditions d'application de l'exemple précédent, que l'on vérifie sur \bar{x}_n vu que p est inconnu: $n \geq 30$ et $\bar{x}_n(1-\bar{x}_n) > 5$ avec $\bar{x}_n = \frac{64}{100} = 0,64$.

L'intervalle de confiance pour p au niveau de confiance 0,95, est donné par

$$P\left(\bar{X}_n - u_{0,975}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \bar{X}_n + u_{0,975}\sqrt{\frac{p(1-p)}{n}}\right) = 0,95$$

avec $u_{0,975} = 1,96$, quantile lu dans la table de $N(0,1)$.

On obtient finalement une fourchette d'estimation asymptotique pour p , à partir de l'échantillon des observations $(x_1, x_2, \dots, x_{100})$, au niveau de confiance 0.95 :

$$0,546 \leq p \leq 0,734.$$

Interprétation La meilleure estimation possible de la proportion est $\bar{x}_n = 0,64$. De plus, on a une confiance de 95% dans le fait que la proportion soit comprise entre 0,546 et 0,734.

Remarque importante Les intervalles de confiance suscitent souvent des erreurs d'interprétation et des abus de langage. Reprenons l'exemple pratique précédent pour illustrer ce problème.

Il est correct de dire que " p a 95% de chances d'être compris entre $\bar{X}_n - u_{0,975}\sqrt{\frac{p(1-p)}{n}}$ et $\bar{X}_n + u_{0,975}\sqrt{\frac{p(1-p)}{n}}$ ", car les bornes sont aléatoires. Mais, il est incorrect de dire que " p a 95% de chances d'être comprise entre 0,546 et 0,734". Effectivement, dans cette écriture, il n'y a rien d'aléatoire. La proportion p est ou n'est pas dans l'intervalle $[0,546; 0,734]$. La probabilité que p soit comprise entre 0,546 et 0,734 est 0 ou 1, mais pas 0,95. En fait, si on recommence 100 fois le sondage précédent sur des échantillons différents, on aura 100

réalisations du couples $\left(\bar{X}_n - u_{0,975}\sqrt{\frac{p(1-p)}{n}}, \bar{X}_n + u_{0,975}\sqrt{\frac{p(1-p)}{n}}\right)$, et ainsi, des intervalles de confiance différents (dus aux fluctuations d'échantillonnage, voir paragraphe 2.3.). En moyenne, p sera dans 95 de ces intervalles et, c'est en ce sens que l'on doit comprendre qu'on a une confiance de 95% dans le fait que la proportion soit comprise entre 0,546 et 0,734.

Remarque Pour diminuer la largeur de l'intervalle d'estimation, on a deux possibilités: soit augmenter α , donc augmenter la probabilité de se tromper ou bien d'augmenter n , le nombre de personnes interrogées, les bornes de l'intervalle dépendant de $\frac{1}{\sqrt{n}}$. C'est pour cela qu'il est indispensable de donner la valeur de α et le nombre de personnes interrogées pour interpréter les résultats d'une fourchette d'estimation.

Intervalles de confiance classiques

Reprenons, de nouveau, les exemples précédents et développons la recherche d'intervalles de confiance.

- *Intervalle de confiance de l'espérance m d'une loi normale $N(m, \sigma^2)$.*

a- *Cas où σ est connu*

Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, σ connu.

Dans le paragraphe 4.2.1, on a vu que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur sans biais et efficace de m . De plus, d'après le paragraphe 3.1., la variable aléatoire $\sqrt{n} \frac{\bar{X}_n - m}{\sigma}$ suit une loi $N(0, 1)$.

Soit $\alpha \in]0, 1[$ donné, $u_{1-\alpha/2}$ désignant le fractile d'ordre $(1 - \alpha/2)$ de la loi $N(0, 1)$, on a :

$$P\left(-u_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - m}{\sigma} \leq u_{1-\alpha/2}\right) = 1 - \alpha$$

et, ainsi, l'intervalle de confiance pour m , est donné par :

$$P\left(\bar{X}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

On obtient finalement une fourchette d'estimation pour m , à partir de l'échantillon des observations (x_1, x_2, \dots, x_n) :

$$\bar{x}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

b- *Cas où σ est inconnu*

Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, σ inconnu.

Comme σ^2 est inconnu, on le remplace par son estimateur sans biais \hat{S}_n^2 défini dans le paragraphe 2.3.1. et d'après le paragraphe 3.3, on sait que la variable aléatoire $\sqrt{n} \frac{\bar{X}_n - m}{\hat{S}_n}$

suit une loi de Student $St(n - 1)$.

Soit $\alpha \in]0, 1[$ donné, t désignant le fractile d'ordre $(1 - \alpha/2)$ de la loi de Student $St(n - 1)$, on obtient, ainsi un intervalle de confiance pour l'espérance m :

$$P\left(\bar{X}_n - t_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}} \leq m \leq \bar{X}_n + t_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}}\right) = 1 - \alpha$$

On obtient finalement une fourchette d'estimation pour m , à partir de l'échantillon des observations (x_1, x_2, \dots, x_n) :

$$\bar{x}_n - t_{1-\alpha/2} \frac{\hat{s}_n}{\sqrt{n}} \leq m \leq \bar{x}_n + t_{1-\alpha/2} \frac{\hat{s}_n}{\sqrt{n}}$$

- *Intervalle de confiance de la variance σ^2 d'une loi normale $N(m, \sigma^2)$.*

a- Cas où m est connue

Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, m connue.

Dans le paragraphe 4.2.1C), on a vu que $\hat{\Sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ est l'estimateur du maximum de vraisemblance de σ^2 . De plus, d'après le paragraphe 3.2, la variable aléatoire $\frac{n}{\sigma^2} \hat{\Sigma}_n^2$ suit une loi du $\chi^2(n)$.

Soit $\alpha = \alpha_1 + \alpha_2 \in]0, 1[$ donné, u_1 et u_2 désignant respectivement, le fractile d'ordre (α_1) et le fractile d'ordre $(1 - \alpha_2)$ de la loi $\chi^2(n)$. Il y a une infinité de choix pour α_1 et α_2 . En général, on choisit $\alpha_1 = \alpha_2 = \alpha/2$ pour symétriser le risque.

on a :

$$P(u_1 \leq n \frac{\hat{\Sigma}_n^2}{\sigma^2} \leq u_2) = 1 - \alpha$$

et, on obtient, ainsi un intervalle de confiance pour la variance σ^2 :

$$P\left(\frac{n\hat{\Sigma}_n^2}{u_2} \leq \sigma^2 \leq \frac{n\hat{\Sigma}_n^2}{u_1}\right)$$

On obtient finalement une fourchette d'estimation pour σ^2 , à partir de l'échantillon des observations (x_1, x_2, \dots, x_n) :

$$\frac{n\hat{\sigma}_n^2}{u_2} \leq \sigma^2 \leq \frac{n\hat{\sigma}_n^2}{u_1}$$

b- Cas où m est inconnue

Considérons un n -échantillon (X_1, X_2, \dots, X_n) de loi normale $N(m, \sigma^2)$, m inconnue. Comme m est inconnue, on estime σ^2 par un estimateur sans biais \hat{S}_n^2 . D'après le paragraphe 3.2,

la v.a. $(n-1)\frac{\hat{S}_n^2}{\sigma^2}$ suit une loi du $\chi^2(n-1)$.

Soit $\alpha \in]0, 1[$ donné, u_1 et u_2 désignant respectivement, le fractile d'ordre (α_1) et le fractile d'ordre $(1 - \alpha_2)$ de la loi $\chi^2(n-1)$, on obtient

$$P(u_1 \leq (n-1)\frac{\hat{S}_n^2}{\sigma^2} \leq u_2) = 1 - \alpha$$

et, on obtient, ainsi un intervalle de confiance pour la variance σ^2 :

$$P\left(\frac{(n-1)\hat{S}_n^2}{u_2} \leq \sigma^2 \leq \frac{(n-1)\hat{S}_n^2}{u_1}\right)$$

On obtient finalement une fourchette d'estimation pour σ^2 , à partir de l'échantillon des observations (x_1, x_2, \dots, x_n) :

$$\frac{(n-1)\hat{S}_n^2}{u_2} \leq \sigma^2 \leq \frac{(n-1)\hat{S}_n^2}{u_1}$$

Exemple pratique

On mesure la compression d'un ciment en moulant des petits cylindres et en mesurant la pression X , mesurée en kg/cm^2 , à partir de laquelle ils se cassent. Pour 10 cylindres testés, on observe les pressions suivantes : 19,6 – 19,9 – 20,4 – 19,8 – 20,5 – 21 – 18,5 – 19,7 – 18,4 – 19,4.

On suppose que la variable aléatoire X suit une loi normale $N(m, \sigma^2)$. On cherche un intervalle de confiance au niveau de confiance 0,90 pour m , la valeur moyenne de la pression à partir de laquelle les cylindres se cassent, puis pour σ^2 . En appliquant les résultats précédents, on obtient un intervalle de confiance pour l'espérance m :

$$P\left(\bar{X}_n - t_{0,95}\frac{\hat{S}_n}{\sqrt{n}} \leq m \leq \bar{X}_n + t_{0,95}\frac{\hat{S}_n}{\sqrt{n}}\right) = 1 - \alpha$$

avec $t_{0,95} = 1,833$, quantile lu dans la table de $St(9)$.

On obtient une fourchette d'estimation pour m , à partir de l'échantillon des observations $(x_1, x_2, \dots, x_{10})$:

$$\bar{x}_{10} - 1,833\frac{\hat{s}_{10}}{\sqrt{10}} \leq m \leq \bar{x}_{10} + 1,833\frac{\hat{s}_{10}}{\sqrt{10}}$$

avec $\bar{x}_{10} = \frac{197,2}{10} = 19,72$ et $\hat{s}_{10} = \sqrt{109}\sqrt{\frac{3894,9}{10} - (19,72)^2} = 0,82$. Pour les calculs de la moyenne et la variance empirique.

Finalement, $19,245 \leq m \leq 20,195$.

L'intervalle de confiance pour la variance σ^2 est donnée par

$$P\left(\frac{(n-1)\hat{S}_n^2}{u_2} \leq \sigma^2 \leq \frac{(n-1)\hat{S}_n^2}{u_1}\right)$$

avec $u_1 = 3,33$ quantile d'ordre 0,05 et $u_2 = 16,92$ quantile d'ordre 0,95 lus dans la table $\chi^2(9)$.

On obtient finalement une fourchette d'estimation pour σ^2 , à partir de l'échantillon des observations $(x_1, x_2, \dots, x_{10})$:

$$0,36 \leq \sigma^2 \leq 1,82.$$