

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A. MIRA de Béjaia
Faculté des Sciences Exactes
Département de Mathématiques

POLYCOPIÉ DE COURS
STATISTIQUE INFÉRENTIELLE

N. SAADI

2019/2020

Chapitre 1

Échantillonnage

1.1 Introduction à la statistique inférentielle

1.1.1 Exemple $n^{\circ 1}$

Un industriel vient de recevoir un lot important de pièces. Afin d'avoir une idée de la proportion, notée θ , de pièces défectueuses, celui-ci réalise un tirage aléatoire avec remise de n pièces au sein du lot. L'échantillon est extrait du lot selon un schéma de Bernoulli. Autrement dit, si on note X_i la variable aléatoire à valeurs dans $\{0, 1\}$ associée au tirage numéro i , on suppose que les n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ sont indépendantes et de même loi ; à savoir $\mathbb{P}(X_i = 1) = \theta$ et $\mathbb{P}(X_i = 0) = 1 - \theta$ (l'événement $X_i = 1$) signifie donc que la pièce i est défectueuse). Le problème statistique est d'estimer la proportion inconnue θ à partir des observations, c'est-à-dire à partir des x_i (réalisations des variables aléatoires X_i).

1.1.2 Exemple $n^{\circ 2}$

Une entreprise E fabrique et commercialise des ampoules électriques de 100 Watts. Sur l'emballage de ces ampoules, l'entreprise E a porté la mention : "durée d'utilisation $T = 1000$ heures". Un organisme de défense du consommateur se propose de valider la mention portée par ce constructeur sur ces emballages. Il souhaite de plus pouvoir décider si la durée de vie (moyenne) d'une ampoule de 100 Watts fabriquée par l'entreprise E est supérieure ou égale à T . Pour ce faire, il teste un grand nombre n d'ampoules fabriquées par cette entreprise, et note x_i la durée de vie observée de l'ampoule

numero i . La durée de vie d'une ampoule électrique étant "par nature" une grandeur imprévisible au sens où la valeur de x_i ne peut, avant de l'avoir observée, être prédite avec certitude, il est commode de la modéliser par une variable aléatoire X . La loi de probabilité généralement retenue pour X est la loi exponentielle. Rappelons qu'une variable aléatoire réelle X suit une loi exponentielle de paramètre θ si sa densité est :

$$f_{\theta}(x) = \frac{1}{\theta} \exp \left\{ \frac{-x}{\theta} \right\}, x \in [0; \infty[.$$

On peut aisément vérifier que le paramètre θ de la loi exponentielle est égal à la durée de vie moyenne d'une ampoule; autrement dit, $\mathbb{E}(X) = \theta$. Le problème consiste : premièrement, à estimer le paramètre inconnu θ à partir des observations, c'est-à-dire à partir des x_i (réalisations des variables aléatoires X_i); et deuxièmement, à décider si $\theta \geq T$ ou si $\theta < T$.

Ces deux exemples ont en commun de baser l'inférence sur le paramètre θ (qu'il s'agisse d'estimer θ ou de tester une valeur particulière de θ)

1.1.3 Position du problème et objectifs

Afin de préciser l'objet et la place de la Statistique inférentielle (au sein de la Statistique en général), il convient de rappeler que le travail du Statisticien s'inscrit traditionnellement dans l'une des trois phases suivantes : production de données, exploration des données, et modélisation (aléatoire) des données. La production de données inclut en particulier l'organisation de la collecte des données (conception de plan de sondage par exemple). L'exploration des données repose essentiellement sur les méthodes de la Statistique descriptive et de l'analyse des données. La modélisation comporte classiquement plusieurs étapes :

- a. Au sein d'un ensemble de modèles probabilistes susceptibles d'avoir généré les observations, sélectionner l'un d'entre eux (cette étape étant réalisée à l'aide d'une procédure statistique de choix de modèle).
- b. Valider le modèle retenu.
- c. Estimer les paramètres du modèle retenu et éventuellement proposer des procédures de tests portant sur les paramètres du modèle. Cette présentation présuppose qu'un modèle paramétrique a été retenu, mais d'autres alternatives existent.

La Statistique inférentielle appelée encore **Statistique mathématique** consiste en un ensemble de concepts, de méthodes et d'outils élaborés en vue de réaliser chacune des étapes précédentes.

1.2 Échantillonnage

Définition 1.1. En statistique, les méthodes d'échantillonnage correspondent aux différentes manières de constituer un échantillon (partie) de la population étudiée, de manière à reproduire un échantillon aussi représentatif que possible de cette population.

1.2.1 Avantages de l'échantillonnage

- Impossibilité d'étudier toute la population lorsqu'elle est infinie.
- Le coût : Le choix d'un échantillon est de moindre coût qu'un recensement.
- Le temps : la rapidité nécessaire de certaines prises de décisions empêche le recours à un recensement.

1.2.2 Méthodes d'échantillonnage

L'échantillonnage peut se faire avec ou sans remise et une population peut être considérée comme finie ou infinie. Une population finie dans laquelle on procède à un échantillonnage avec remise peut être théoriquement considérée comme infinie. Dans la pratique, il en va de même pour des populations finies mais de grandes tailles. Pour chaque distribution d'échantillonnage, on peut calculer une moyenne, un écart-type, une variance...etc.

1. **Méthodes probabilistes (Aléatoires) :** L'échantillonnage probabiliste repose sur un choix d'unités dans la population fait au hasard. Une des caractéristiques de cette méthode est que chaque unité de la population a une probabilité mesurable d'être choisie. Voici les quatre types d'échantillonnage probabiliste que l'on peut effectuer

- a. **Échantillonnage aléatoire simple :** Chaque individu a la même probabilité d'être choisi.

Exemple 1.2.1. Un enseignant met le nom des élèves du collège dans un chapeau et, sans regarder, il en tire un certain nombre de noms des élèves qui constitueront l'échantillon. Cette méthode permet d'obtenir un échantillon représentatif de la population car elle donne la même probabilité à chaque individu de faire partie de l'échantillon.

- b. **Échantillonnage aléatoire stratifié :** La population est initialement subdivisée en sous-groupes homogènes (strates) définis selon un ou plusieurs critères

(appelés variables d'intérêt). Dans chaque strate, on prélève aléatoirement des individus pour obtenir des sous-échantillons aléatoires simples.

Exemple 1.2.2. Une enquête a été menée auprès de 100 étudiants. L'échantillon a été obtenu en choisissant aléatoirement 25 étudiants de première année, 25 étudiants de deuxième année, 25 étudiants de licence et 25 étudiants en Master I. Cette méthode permet d'obtenir un échantillon représentatif car tous les individus d'un groupe ont la même probabilité de faire partie du sous-échantillon et l'échantillon obtenu est représentatif de la population en ce qui concerne la variable d'intérêt (ici, le cycle universitaire).

- c. **Échantillonnage aléatoire par grappe** : On subdivise la population en sous-groupes appelés "grappes". Les grappes ont le même profil. On sélectionne par la suite un échantillon aléatoire de grappes. L'échantillon désiré est constitué de tous les individus de chaque grappe.

Exemple 1.2.3. Une compagnie aérienne décide de mener une enquête de satisfaction auprès de ses clients. Pour cela, elle choisit aléatoirement 5 vols de la journée et interroge tous les passagers de ces vols. Cette méthode permet d'obtenir un échantillon représentatif de la population si les grappes sont semblables entre elles et si dans une grappe, les individus sont hétérogènes.

- d. **Échantillonnage aléatoire systématique** : Cette méthode consiste à dresser la liste de tous les éléments de la population et de choisir au hasard chaque n^e start éléments pour constituer l'échantillon. L'échantillonnage systématique est donc une variante de la méthode aléatoire simple, dans laquelle on sélectionne un échantillon à intervalles prédéterminés.

Exemple 1.2.4. À partir de la liste alphabétique des noms de tous les élèves du collège, un Conseiller Pédagogique d'Education choisit au hasard un premier nom puis, à partir de ce nom, il choisit chaque 20^e nom de collégiens pour constituer l'échantillon.

2. **Méthodes non probabilistes (Raisonnées ou empiriques)** : Utilisent des connaissances préalables qu'on a sur la population. La méthode la plus utilisée est la méthode des quotas. Cette méthode se base sur la construction d'un échantillon de taille n dans lequel les proportions des individus sont égales à celles de la population. Une fois ces quotas déterminés, il faut les respecter dans le choix de l'échantillon.

1.3 Échantillon- Caractéristiques usuelles d'un échantillon

1.3.1 Population statistique

C'est un ensemble que l'on observe et qui sera soumis à une analyse statistique, chaque élément de cet ensemble est un individu ou **unité** statistique.

1.3.2 Échantillon

On suppose maintenant qu'on étudie un caractère statistique quantitatif X sur une population $\Omega = \{\omega_1, \dots, \omega_N\}$. On supposera vérifiée l'hypothèse d'échantillonnage aléatoire simple. Soit X_k le résultat aléatoire du k -ième tirage et on note (X_1, X_2, \dots, X_n) le résultat aléatoire de ces n tirages. Les variables aléatoires $(X_k)_{1 \leq k \leq n}$ sont indépendantes et suivent toutes la même loi de probabilité (celle de X : variable parente), d'espérance la moyenne de X et d'écart-type celui de la variable aléatoire X . Les valeurs observées x_1, x_2, \dots, x_k constituent n réalisations indépendantes de la variable aléatoire X ou encore, une réalisation unique du n -uplet (X_1, X_2, \dots, X_n) .

Définition 1.2. (X_1, X_2, \dots, X_n) est un n -uplet de variables aléatoires indépendantes de même loi (celle de X). Il est appelé n -échantillon aléatoire simple.

1.3.3 Distribution d'un échantillon-Vraisemblance d'un échantillon

Définition 1.3. Une statistique T est une variable aléatoire fonction mesurable de X_1, X_2, \dots, X_n .

$$T = f(X_1, \dots, X_n).$$

Remarque 1.3.1. Une statistique T peut être à valeurs dans \mathbb{R} ou \mathbb{R}^p , dans le cas de \mathbb{R}^p , on parlera de statistique vectorielle.

Définition 1.4. Vraisemblance d'un échantillon : Soit (X_1, X_2, \dots, X_n) un échantillon aléatoire simple issu d'une variable aléatoire X de densité de probabilité f_X . La vraisemblance de l'échantillon (la distribution du n -échantillon) (X_1, X_2, \dots, X_n) est :

$$L_X(x_1, x_2, \dots, x_n) = L_X(x) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i) = (f_X(x))^n.$$

Exemple 1.3.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une variable aléatoire $X \rightsquigarrow \mathcal{N}(0, 1)$.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), x \in \mathbb{R}.$$

La vraisemblance de l'échantillon est :

$$L_X(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right), x_i \in \mathbb{R}, \forall i = \overline{1, n}.$$

1.3.4 Fonction de répartition empirique d'un échantillon

Définition 1.5. Soit (X_1, X_2, \dots, X_n) un échantillon aléatoire simple issu d'une variable aléatoire X . Désignons par $F_n(x)$ la proportion des n variables X_1, X_2, \dots, X_n qui sont inférieures à x . $F_n(x)$ est donc une variable aléatoire pour tout x qui définit ainsi une fonction aléatoire appelée fonction de répartition empirique de l'échantillon, dont la réalisation sont des fonctions en escalier de sauts égaux à $\frac{1}{n}$. Si les x_i sont ordonnées par valeurs croissantes alors :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i \leq x)} = \begin{cases} 0, & \text{si } x < x_1; \\ \frac{1}{n}, & \text{si } x_1 \leq x < x_2; \\ \vdots & \\ \frac{i-1}{n}, & \text{si } x_{i-1} \leq x < x_i; \\ \vdots & \\ 1, & \text{si } x \geq x_n. \end{cases}$$

Propriétés asymptotiques de la fonction de répartition empirique

Théorème 1.3.1. Pour tout x , on a

$$F_n(x) \xrightarrow{P.s.} F(x).$$

Preuve. Pour x fixé, soit Y le nombre aléatoire de valeurs inférieures à x , qui est une somme de variables de Bernoulli du paramètre $F(x)$. D'après ce qui précède $F_n(x)$ ce n'est autre que $\frac{Y}{n}$ converge presque sûrement vers $F(x)$.

Théorème 1.3.2. (Glivenko-Cantelli)

La convergence de F_n vers F est presque sûrement uniforme, c'est-à-dire que :

$$D_n = \sup_x |F_n(x) - F(x)| \rightarrow 0.$$

Théorème 1.3.3. (Kolmogorov)

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n < y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y^2).$$

Ce théorème signifie que la distribution asymptotique de la variable aléatoire D_n est connue et ne dépend pas de la variable de départ X , et permet de calculer des limites pour les valeurs de D_n . La loi exacte de la variable D_n a été tabulée.

1.3.5 Échantillon ordonné et lois des valeurs extrêmes

Soit (X_1, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X . Les réalisations x_1, x_2, \dots, x_n peuvent être réordonnées en y_1, y_2, \dots, y_n où $y_1 < y_2 < \dots < y_n$. Les y_i constituent une permutation particulière des x_i . Les y_i sont des réalisations du n -uplet de variables aléatoires (Y_1, Y_2, \dots, Y_n) qui constitue l'échantillon ordonné de X . Soit F (resp. f) la fonction de répartition (resp. la densité) de la variable aléatoire X . Soit H_k (resp. h_k) les fonctions de répartition (resp. les densités) de Y_k .

A. Loi de $Y_1 = \inf_{1 \leq i \leq n} X_i$.

On a

$$\begin{aligned} H_1(y) &= \mathbb{P}(Y_1 < y) \\ &= 1 - \mathbb{P}(Y_1 \geq y) \\ &= 1 - \mathbb{P}\left(\inf_{1 \leq i \leq n} X_i \geq y\right) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(X_i \geq y) = 1 - \prod_{i=1}^n [1 - \mathbb{P}(X_i < y)] \\ &= 1 - \prod_{i=1}^n [1 - F(y)] \end{aligned}$$

donc

$$H_1(y) = 1 - [1 - F(y)]^n.$$

$$h_1(y) = n [1 - F(y)]^{n-1} f(y).$$

B. Loi de $Y_n = \sup_{1 \leq i \leq n} X_i$

$$\begin{aligned}
 H_n(y) &= \mathbb{P}(Y_n < y) \\
 &= \mathbb{P}\left(\sup_{1 \leq i \leq n} X_i < y\right) \\
 &= \prod_{i=1}^n \mathbb{P}(X_i < y) \\
 &= \prod_{i=1}^n [F(y)] \\
 &= [F(y)]^n
 \end{aligned}$$

et

$$h_n(y) = nF(y)^{n-1}f(y).$$

Remarque 1.3.2. Ces deux lois servent en particulier à détecter les valeurs aberrantes de l'échantillon : valeurs trop grandes ou trop petites.

1.3.6 Distributions d'échantillonnage de certains moments

Définition 1.6. Moment Centré d'une variable aléatoire.

Le k -ième moment non centré d'une variable aléatoire X est défini par :

$$\mu'_k = \mathbb{E}(X - E(X))^k.$$

Définition 1.7. Moment centré d'un échantillon

Soit (X_1, X_2, \dots, X_n) un n -échantillon aléatoire issu d'une variable aléatoire X et k un nombre entier ($k \in \mathbb{N}^*$). On appelle moment centré d'ordre k de l'échantillon et on note M'_k , la quantité :

$$M'_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k.$$

Définition 1.8. Moment non centré d'une variable aléatoire.

Le k -ième moment (théorique) non centré d'une variable aléatoire X est défini par :

$$\mu_k = \mathbb{E}(X^k).$$

Définition 1.9. Moment non centré d'un échantillon

Soit (X_1, X_2, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X et k un nombre entier ($k \in \mathbb{N}^*$). On appelle moment non centré d'ordre k de l'échantillon et on note M'_k , la quantité :

$$M_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Théorème 1.3.4. Soit (X_1, X_2, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X , alors

$$\mathbb{E}(M_k) = \mu_k.$$

En effet,

$$\mathbb{E}(M_k) = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i^k \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^k] = \mu_k.$$

Moyenne empirique d'un échantillon

Définition 1.10. la moyenne empirique \bar{X} ou le moment non centré d'ordre 1 de l'échantillon est :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Propriété 1.3.5. Soient μ et σ l'espérance et l'écart-type de la variable parente X , on a alors :

$$\mathbb{E}(\bar{X}) = \mu \text{ et } Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Preuve.

$$\begin{aligned} \mathbb{E}(\bar{X}) &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} n\mu = \mu. \\ Var(\bar{X}) &= \frac{1}{n^2} Var \left[\sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

Lois des grands nombres : Elles sont de deux types : lois faibles mettant en jeu la convergence en probabilité et lois fortes relatives à la convergence presque sûre. Nous considérons ici des suites de variables aléatoires X_1, \dots, X_n non nécessairement de même loi.

Théorème 1.3.6. (Loi faible des grands nombres.)

Soit (X_1, \dots, X_n) n-échantillon de variables aléatoires indépendantes d'espérance μ_1, \dots, μ_n finies et de variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ finies.

$$\text{Si } \frac{1}{n} \sum_{i=1}^n \mu_i \rightarrow \mu \text{ et si } \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \rightarrow 0, \text{ alors}$$

$$\bar{X} \xrightarrow{P} \mu.$$

Théorème 1.3.7. (Loi forte des grands nombres.)

Soit (X_1, \dots, X_n) n-échantillon de variables aléatoires indépendantes telles que $\frac{1}{n} \sum_{i=1}^n \mu_i \rightarrow \mu$ et $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}$ est convergente, alors

$$\bar{X} \xrightarrow{P.s} \mu.$$

Application : Cas des variables aléatoires indépendantes et identiquement distribuées

On voit aisément que $\bar{X} \xrightarrow{P.s} \mu$ car la condition $\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}$ est évidemment réalisée et s'écrit :

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} = \sigma^2 \sum_{i=1}^{\infty} \frac{1}{i^2}.$$

et l'on sait que la série $\sum_i \frac{1}{i^2}$ converge.

Théorème 1.3.8. Pour n assez grand et d'après le théorème central limite :

$$\frac{\bar{X} - \mathbb{E}(\bar{X})}{\sqrt{\text{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Ce résultat est d'une importance capitale en statistique.

Variance empirique d'un échantillon

Définition 1.11. Soit (X_1, \dots, X_n) un n-échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . On appelle variance empirique (moment centré d'ordre 2) de l'échantillon la quantité :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Propriété 1.3.9.

$$S^2 \xrightarrow{P.s.} \sigma^2.$$

Preuve. D'après les lois des grands nombres :

$$\frac{1}{n} \left(\sum_{i=1}^n X_i^2 \right) \xrightarrow{P.s.} \mathbb{E}(X^2)$$

et

$$\bar{X}^2 \longrightarrow [\mathbb{E}(X)]^2$$

donc

$$S^2 \xrightarrow{P.s.} \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \sigma^2.$$

Propriété 1.3.10.

$$\mathbb{E}(S^2) = \frac{n-1}{n} \sigma^2.$$

Preuve. Partons de $X_i - \mu = X_i - \bar{X} + \bar{X} - \mu$.

On a alors :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}).$$

Comme $\sum_{i=1}^n (X_i - \bar{X}) = 0$, on trouve :

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 + (\bar{X} - \mu)^2$$

d'où,

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2.$$

Calculons $\mathbb{E}(S^2)$

$$\begin{aligned} \mathbb{E}(S^2) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [X_i - \mu]^2 - \mathbb{E} [\bar{X} - \mu]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \text{Var}(X_i) - \text{Var}(\bar{X}) \\ &= \frac{1}{n} \sum_{i=1}^n \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} \\ &= \frac{n-1}{n} \sigma^2. \end{aligned}$$

Ce résultat montre que $\mathbb{E}(S^2) \neq \sigma^2$. On dit que S^2 est une statistique biaisée pour σ^2 .

Théorème 1.3.11. Soit (X_1, \dots, X_n) un n-échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . Si μ'_4 est le moment centré d'ordre 4 de X , alors :

$$\text{Var}(S^2) = \frac{n-1}{n^3} [(n-1)\mu'_4 - (n-3)\sigma^4].$$

$$\text{Si } n \rightarrow \infty \text{ alors } \text{Var}(S^2) \simeq \frac{\mu'_4 - \sigma^4}{n}.$$

Théorème limite pour S^2 .

$$\frac{S^2 - \frac{n-1}{n}\sigma^2}{\sqrt{\text{Var}(S^2)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Théorème 1.3.12. Soit (X_1, \dots, X_n) un n-échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . Lorsque la moyenne est connue, alors

$$\mathbb{E}[S^2] = \sigma^2.$$

$$\text{Var}(S^2) = \frac{\mu'_4 - \sigma^4}{n}.$$

Preuve. On a

$$\mathbb{E}(S^2) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n \text{Var}(X_i) = \sigma^2.$$

Par ailleurs, les variables $(X_i - \mu)^2$ étant indépendantes :

$$\begin{aligned} \text{Var}(S^2) &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[(X_i - \mu)^2] \\ &= \frac{1}{n} \left[\mathbb{E}(X - \mu)^4 - [\mathbb{E}(X - \mu)^2]^2 \right] \\ &= \frac{\mu'_4 - \sigma^4}{n}. \end{aligned}$$

Théorème 1.3.13. Soit (X_1, \dots, X_n) un n-échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . Lorsque la moyenne est connue et n est assez grand

$$S^2 \xrightarrow{\mathcal{L}} \mathcal{N}\left(\sigma^2, \frac{\mu'_4 - \sigma^4}{n}\right).$$

Variance empirique corrigée

Définition 1.12. Soit (X_1, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . On appelle variance empirique corrigée de l'échantillon la quantité :

$$S^{*2} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Propriété 1.3.14.

$$\mathbb{E}(S^{*2}) = \sigma^2.$$

1.3.7 Coefficient d'asymétrie et d'aplatissement d'un échantillon

Il existe plusieurs caractéristiques de forme fondées sur les moments. Nous présentons les caractéristiques les plus utilisées.

Définition 1.13. *Coefficient d'asymétrie.*

Le coefficient d'asymétrie d'une variable aléatoire X est le rapport noté γ_1 , entre le moment centré d'ordre 3 et le cube de l'écart-type, c'est-à-dire :

$$\gamma_1 = \frac{\mu'_3}{\sigma^3}.$$

Définition 1.14. *Coefficient d'aplatissement.*

Le coefficient d'aplatissement d'une variable aléatoire X est le rapport noté γ_2 , entre le moment centré d'ordre 4 et la puissance 4 de l'écart-type, rapport diminué de 3, c'est-à-dire :

$$\gamma_2 = \frac{\mu'_4}{\sigma^4} - 3.$$

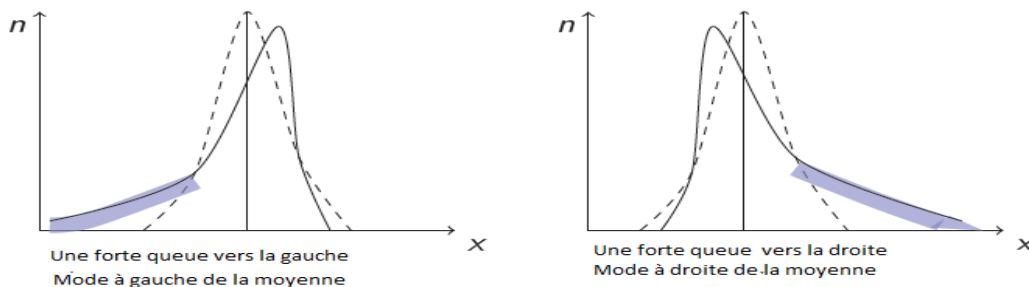
Définition 1.15. *Coefficient d'asymétrie d'un échantillon.*

Le coefficient d'asymétrie d'un échantillon est le rapport noté $\hat{\gamma}_1$, entre le moment empirique centré d'ordre 3 et le cube de l'écart-type empirique, c'est-à-dire :

$$\hat{\gamma}_1 = \frac{M'_3}{S^3}.$$

Interprétation :

- Si la répartition de l'échantillon ou de la distribution est symétrique autour de la moyenne, le coefficient d'asymétrie est nul.
- Dans le cas où il est positif, nous avons une asymétrie gauche (une forte queue vers la gauche), le graphique des fréquences est plus élevé à gauche.
- Lorsque la distribution possède une forte queue vers la droite, le coefficient est négatif.



Définition 1.16. Coefficient d'aplatissement empirique d'un échantillon.

Nous appelons le coefficient d'aplatissement d'un échantillon le rapport noté $\hat{\gamma}_2$, entre le moment empirique centré d'ordre 4 et la puissance 4 de l'écart-type empirique, rapport diminué de 3, c'est-à-dire :

$$\hat{\gamma}_2 = A = \frac{M'_4}{S^4} - 3.$$

Interprétation :

- lorsque le coefficient d'aplatissement est nul nous disons que la répartition des observations est de type gaussien ou normal, c'est-à-dire que la courbe des fréquences a la forme d'une cloche comme la densité d'une loi Normale.

- Lorsqu'il est positif, nous avons une répartition surgaussienne ou encore surnormale, c'est-à-dire moins aplatie qu'une densité normale.
- Lorsqu'il est négatif nous disons que la répartition est sous-gaussienne ou encore sous-normale, c'est-à-dire plus aplatie qu'une densité normale.

Théorème 1.3.15. Soit (X_1, \dots, X_n) un n -échantillon aléatoire simple issu d'une variable aléatoire X de moyenne μ et de variance σ^2 . Si μ'_3 et μ'_4 sont les moments centrés d'ordre 3 et 4 de X , alors

$$\begin{aligned}\mu'_3(\bar{X}) &= \frac{\mu'_3}{n^2} \text{ et } \mu'_4(\bar{X}) = \frac{\mu'_4 + 3\sigma^4(n-1)}{n^3}. \\ \gamma_1(\bar{X}) &= \frac{\gamma_1}{\sqrt{n}} \text{ et } \gamma_2(\bar{X}) = 3 + \frac{\gamma_2 - 3}{n}.\end{aligned}$$

Remarque 1.3.3. On voit que si $n \rightarrow \infty$, $\gamma_1(\bar{X}) \rightarrow 0$ et $\gamma_2(\bar{X}) \rightarrow 3$, ce qui traduit la normalité asymptotique de \bar{X} .

Corrélation entre \bar{X} et S^2

Cherchons $Cov(\bar{X}, S^2)$:

$$Cov(\bar{X}, S^2) = \mathbb{E}[(\bar{X} - \mathbb{E}(\bar{X})) (S^2 - \mathbb{E}(S^2))] = \mathbb{E}\left[(\bar{X} - \mu) \left(S^2 - \frac{n-1}{n}\sigma^2\right)\right].$$

Nous pouvons supposer que $\mu = 0$, car on sait que la covariance est insensible à un changement par translation sur un des termes :

$$\begin{aligned}Cov(\bar{X}, S^2) &= \mathbb{E}(\bar{X}S^2) \\ &= \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right)\right] \\ &= \frac{1}{n^2} \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \left(\frac{1}{n} \sum_{i=1}^n X_i^2\right)\right] - \mathbb{E}(\bar{X}^3) \\ &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n X_i X_j^2\right] - \mathbb{E}(\bar{X}^3) \\ &= \frac{1}{n^2} \mathbb{E}\left(\sum_{i=1}^n X_i^3\right) - \frac{1}{n^3} \mathbb{E}\left(\sum_{i=1}^n X_i^3\right)\end{aligned}$$

Car $\mathbb{E}(X_i X_j) = 0$ pour $i \neq j$ à cause de l'indépendance :

$$\mathbb{E}(\bar{X}S^2) = \frac{\mu_3}{n} - \frac{\mu_3}{n^2} = \frac{n-1}{n^2} \mu_3$$

Donc si $n \rightarrow \infty$, \bar{X} et S^2 sont asymptotiquement non corrélés et si la distribution est symétrique $\mu_3 = 0$, alors les deux variables \bar{X} et S^2 sont non corrélés quel que soit n .

1.3.8 Cas des échantillons gaussiens de petites tailles

On suppose que $X \rightarrow \mathcal{N}(\mu, \sigma)$.

Loi de \bar{X} .

\bar{X} combinaison linéaire de variables gaussiennes est aussi gaussiennes. Donc la loi de \bar{X} est :

$$\mathcal{N}\left(\mu, \frac{\sigma}{n}\right).$$

Il s'agit ici d'une loi exacte.

Espérance et variance de S^2 .

On suppose que $X \rightarrow \mathcal{N}(\mu, \sigma)$. D'après la décomposition de S^2 on peut écrire :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

En divisant par σ^2 :

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{nS^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2.$$

Soit $Z_i = \left(\frac{X_i - \mu}{\sigma} \right)$

$$Z_i \rightsquigarrow \mathcal{N}(0, 1) \implies Z_i^2 \rightsquigarrow \chi_1^2.$$

D'où,

$$\sum_{i=1}^n Z_i^2 \rightsquigarrow \chi_n^2$$

(Comme somme de n carrés de variables aléatoires indépendantes normales centrées réduites).

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1) \implies \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \rightsquigarrow \chi_1^2.$$

Donc

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

D'où, on en déduit :

$$S^2 \rightsquigarrow \frac{\sigma^2}{n} \chi_{n-1}^2.$$

$$\mathbb{E}(S^2) = \frac{\sigma^2}{n}(n-1), \text{Var}(S^2) = \frac{2\sigma^4}{n^2}(n-1).$$

Propriété 1.3.16.

$$\mathbb{E}(S^{*2}) = \sigma^2, \text{Var}(S^{*2}) = \frac{2\sigma^4}{n-1}.$$

Preuve.

$$\mathbb{E}(S^{*2}) = \mathbb{E}\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}\mathbb{E}(S^2) = \frac{n}{n-1}\left(\frac{n-1}{n}\sigma^2\right) = \sigma^2.$$

*Calculons la variance de S^{*2} .*

$$\begin{aligned} \text{Var}(S^{*2}) &= \text{Var}\left(\frac{n}{n-1}S^2\right) \\ &= \frac{n^2}{(n-1)^2}\text{Var}(S^2) \\ &= \frac{n^2}{(n-1)^2}\frac{2(n-1)}{n^2}\sigma^4 \\ &= \frac{2\sigma^4}{n-1}. \end{aligned}$$

Espérance et variance des principales caractéristiques d'un échantillon gaussien

Le tableau suivant récapitule les résultats concernant \bar{X} et S^2 .

Statistique	Espérance	Variance
\bar{X}	μ	$\frac{\sigma^2}{n}$
S^2	$\frac{n-1}{n}\sigma^2$	$\frac{2(n-1)}{n^2}\sigma^4$
S^{*2}	σ^2	$\frac{2\sigma^4}{n-1}$
$\hat{\gamma}_1$	$\simeq 0$	$\simeq \frac{6}{n}$
$\hat{\gamma}_2$	$\simeq 3$	$\simeq \frac{24}{n}$

TABLE 1.1 – Principales caractéristiques d'un échantillon gaussien.

Application :

Puisque

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1),$$

et

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2,$$

on aura :

$$T_{n-1} = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sqrt{n}}{\sqrt{\frac{nS^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{S} \sqrt{n-1},$$

où T_{n-1} est une variable aléatoire de Student à $n-1$ degré de liberté.

Remarque 1.3.4. Ce résultat est extrêmement utile car il dépend pas de σ et servira donc à chaque fois que σ est inconnu.

1.4 Modèle statistique paramétrique

Définition 1.17. Modèle statistique d'échantillonnage paramétrique : Soit X une variable aléatoire à valeurs dans un espace χ muni d'une tribu \mathcal{A} et de loi \mathbb{P}_θ où $\theta \in \Theta \in \mathbb{R}^p$. La donnée de n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ à valeurs également dans (χ, \mathcal{A}) , indépendantes et de même loi que celle de X constitue ce qu'on appelle un modèle d'échantillonnage paramétrique. Il est noté $(\chi, \mathcal{A}, \mathbb{P}_\theta, \theta \in \Theta)^n$.

1.4.1 Remarques et notations

- La famille de lois de probabilités \mathbb{P}_θ à laquelle appartient la loi de X est supposée connue, seul le paramètre θ est inconnu.
- Les n variables aléatoires X_1, X_2, \dots, X_n sont dites indépendantes et identiquement distribuées (en abrégé iid), et le **vecteur aléatoire** (X_1, X_2, \dots, X_n) constitue ce qu'on appelle un échantillon de variable aléatoires indépendantes et identiquement distribuées (v.a.iid) de \mathbb{P}_θ .
- (x_1, x_2, \dots, x_n) est l'échantillon observé et l'espace χ^n est appelé l'espaces des observations.

- Du point de vue du praticien, envisager un modèle paramétrique pour un ensemble d'observations x_1, x_2, \dots, x_n consiste à considérer les x_i comme des réalisations de variables aléatoires X_1, X_2, \dots, X_n indépendantes et de même loi \mathbb{P}_θ . On dit que le modèle est discret quand l'espace des observations χ est fini ou dénombrable. Dans ce cas, la tribu \mathcal{A} est l'ensemble des parties de χ : $\mathcal{A} = \mathcal{P}(\chi)$. C'est le cas quand l'élément aléatoire observé X a une loi de probabilité discrète. On dit que le modèle est continu quand $\chi \subset \mathbb{R}^p$. Dans ce cas, \mathcal{A} est la tribu des boréliens de χ (tribu engendrée par les ouverts de χ) : $\mathcal{A} = \mathcal{B}(\chi)$. C'est le cas quand l'élément aléatoire observé X a une loi de probabilité continue. On peut aussi envisager des modèles ni continus ni discrets, par exemple si l'observation a certains éléments continus et d'autres discrets.
- Le cas le plus fréquent est celui où l'élément aléatoire observé est constitué de variables aléatoires indépendantes et de même loi (i.i.d.) : $X = (X_1, X_2, \dots, X_n)$, où les X_i sont i.i.d. On dit que l'on a alors un modèle d'échantillon. Dans ce cas, on note $(\chi; \mathcal{A}; \mathbb{P}_\theta)$ le modèle statistique correspondant à un échantillon.

Exemple 1.4.1. Contrôle de qualité. Une chaîne de production produit un très grand nombre de pièces et on s'intéresse à la proportion inconnue de pièces défectueuses. Pour l'estimer, on prélève indépendamment n pièces dans la production et on les contrôle. L'observation est $= (x_1, \dots, x_n)$, où :

$$x_i = \begin{cases} 1 & \text{si la } i\text{ème pièce est défectueuse} \\ 0 & \text{sinon} \end{cases}.$$

Par conséquent, l'espace $\Omega = \{0, 1\}^n$ Il est fini, donc le modèle est discret et $\mathcal{A} = \mathcal{P}(\{0, 1\}^n)$. Les X_i sont indépendantes et de même loi de Bernoulli $\mathcal{B}(p)$, où p (inconnu) est la probabilité qu'une pièce soit défectueuse. Alors le modèle statistique peut s'écrire :

$$(\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), \{\mathcal{B}^{\otimes n}(p), p \in [0, 1]\})$$

ou

$$(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{\mathcal{B}(p), p \in [0, 1]\})^n.$$

Chapitre 2

Exhaustivité

2.1 Introduction

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon apporte une certaine information sur ce paramètre. Lorsque l'on résume cet échantillon par une statistique, il s'agit de ne pas perdre cette information ; une statistique qui conserve l'information sera qualifiée **d'exhaustivité**. Il convient de donner un sens précis à la notion d'information : une première approche consiste à remarquer qu'une variable aléatoire T ne peut nous renseigner sur la valeur d'un paramètre que dans la mesure où sa loi de probabilité dépend de ce paramètre ; si la variable T est une statistique relative à l'échantillon (X_1, \dots, X_n) et que la loi conditionnelle de (X_1, \dots, X_n) à T fixé ne dépend plus du paramètre θ , on peut dire alors, qu'une fois T connue, nous n'obtenons plus d'autre information de l'échantillon concernant θ et donc que T porte toute l'information disponible sur θ . Une deuxième approche consiste à définir mathématiquement une quantité d'information et chercher dans quelles circonstances cette quantité se conserve lorsque les données sont résumées par une statistique.

2.1.1 Statistique exhaustive

Soit un n -échantillon d'une variable aléatoire X . On notera $L(x_1, \dots, x_n; \theta)$ soit la densité de (X_1, \dots, X_n) si X est absolument continue, soit la probabilité conjointe

$\mathbb{P}((X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n))$ si X est discrète.

Soit T une statistique fonction de X_1, \dots, X_n de loi $g(t, \theta)$ (densité dans le cas continu, $\mathbb{P}(T = t)$ dans le cas discret).

Définition 2.1. Une statistique T est exhaustive pour θ si et seulement si la loi de probabilité conditionnelle de X sachant $[T = t]$ ne dépend pas de θ .

”Ceci veut dire qu’une fois T connu, aucune valeur de l’échantillon ni aucune autre statistique ne nous apportera de renseignements supplémentaires sur θ ”.

Exemple 2.1.1. • Exhaustivité dans un modèle de Bernoulli pour le contrôle statistique de la qualité.

Un industriel voudrait connaître la proportion p de pièces défectueuses qu’il fabrique dans une journée. Pour cela il prélève n pièces aléatoirement dans l’ensemble de sa production de la journée. Il suppose que la qualité de sa production n’a pas évolué au cours de la journée, autrement dit que cette proportion n’a pas varié. Il note alors le nombre k de pièces défectueuses observées dans cet échantillon et estime la qualité de sa production par $\frac{k}{n}$. Il néglige donc toute une partie de l’information apportée par l’échantillon de son contrôle, comme celle de savoir quelles pièces se sont révélées défectueuses. Est-ce handicapant pour la qualité de son estimation ?

En premier lieu, notons que le modèle statistique pour une telle expérience est le modèle associé au modèle de la loi de Bernoulli :

$$(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{\mathcal{B}(p), p \in [0, 1]\})^n$$

de v.a. générique X . De l’observation $x = (x_1, \dots, x_n)$ d’un échantillon $X = (X_1, \dots, X_n)$ où

$$x_i = \begin{cases} 1 & \text{si la } i\text{ème pièce est défectueuse} \\ 0 & \text{sinon} \end{cases}.$$

Les X_i sont des variables aléatoires indépendantes et de même loi $\mathcal{B}(p)$, où p est la probabilité qu’une pièce soit défectueuse. Dans ce modèle, l’industriel a donc retenu seulement l’information apportée par la statistique

$$T(X) = \sum_{i=1}^n X_i.$$

qui vaut k dans cet exemple. Or calculons la loi de X conditionnellement à $T(X) = k$.

On sait que $T(X) = \sum_{i=1}^n X_i$ est de loi binomiale $\mathcal{B}(n; p)$. Alors

$$\begin{aligned} \mathbb{P}(X = x | T(X) = k) &= \mathbb{P} \left\{ (X_1 = x_1) \cap \dots \cap (X_n = x_n) \mid \sum_{i=1}^n X_i = k \right\} \\ &= \frac{\mathbb{P} \{ (X_1 = x_1), \dots, (X_n = x_n), \sum_{i=1}^n X_i = k \}}{\mathbb{P}(\sum_{i=1}^n X_i = k)} \\ &= \begin{cases} 0 & \text{si } \sum_{i=1}^n x_i \neq k \\ \frac{\mathbb{P}\{X_1=x_1, \dots, X_n=x_n\}}{\mathbb{P}(\sum_{i=1}^n X_i = k)} & \text{si } \sum_{i=1}^n x_i = k \end{cases} \\ \mathbb{P}(X_i = x_i) &= \begin{cases} p & \text{si } x_i = 1 \\ 1 - p & \text{si } x_i = 0 \end{cases} = p^{x_i} (1 - p)^{1-x_i} \end{aligned}$$

et comme les X_i sont indépendantes, on a :

$$\begin{aligned} \frac{\mathbb{P} \{ X_1 = x_1, \dots, X_n = x_n \}}{\mathbb{P}(\sum_{i=1}^n X_i = k)} &= \frac{\prod_{i=1}^n \mathbb{P}(X_i = x_i)}{\mathbb{P}(\sum_{i=1}^n X_i = k)} = \frac{\prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i}}{\mathbb{P}(\sum_{i=1}^n X_i = k)} \\ &= \frac{p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}}{C_n^k p^k (1 - p)^{n-k}} = \frac{p^k (1 - p)^{n-k}}{C_n^k p^k (1 - p)^{n-k}} \\ \mathbb{P}(X = x | T = k) &= \begin{cases} 0 & \text{si } \sum_{i=1}^n x_i \neq k \\ \frac{p^k (1-p)^{n-k}}{C_n^k p^k (1-p)^{n-k}} = \frac{1}{C_n^k} & \text{si } \sum_{i=1}^n x_i = k \end{cases} \end{aligned}$$

On constate que cette loi ne dépend pas du paramètre p que l'on cherche à estimer. Ainsi, toute l'information sur le paramètre p contenue dans l'échantillon X est en fait contenue dans la statistique $T(X)$. On dit que cette statistique est exhaustive.

Théorème 2.1.1. (Théorème de factorisation de Fisher-Neyman). Pour qu'une statistique T soit exhaustive pour θ , il faut et il suffit qu'il existe deux fonctions mesurables g et h telles que :

$$L(x; \theta) = g(t, \theta)h(x).$$

Preuve. Effectuons la démonstration dans le cas d'un modèle discret. On a donc

$$L(x; \theta) = P(X = x; \theta).$$

(\Rightarrow) Si T est exhaustive, $\mathbb{P}(X = x|T(X) = k)$ ne dépend pas de θ . Par conséquent :

$$\begin{aligned} L(x; \theta) &= \mathbb{P}(X = x; \theta) = \mathbb{P}(X = x) \cap (T(X) = t(x)); \theta) \\ &= \mathbb{P}((X = x) \cap (T = t(x)); \theta) = \mathbb{P}(X = x|T = t(x)) \mathbb{P}(T = t(x); \theta) \\ &= h(x) \mathbb{P}(T = t(x); \theta) \end{aligned}$$

qui est bien de la forme $g(t(x), \theta)h(x)$.

(\Leftarrow) On suppose que

$$L(x; \theta) = \mathbb{P}(X = x; \theta) = g(t(x); \theta)h(x).$$

Il faut montrer qu'alors $\mathbb{P}(X = x|T = t)$ ne dépend pas de θ . On a :

$$\begin{aligned} \mathbb{P}(X = x|T = t_0; \theta) &= \frac{\mathbb{P}(X = x, T = t_0; \theta)}{\mathbb{P}(T = t_0; \theta)} \\ &= \begin{cases} 0 & \text{si } t(x) \neq t_0 \\ \frac{\mathbb{P}(X=x; \theta)}{\mathbb{P}(T=t_0; \theta)} & \text{si } t(x) = t_0. \end{cases} \end{aligned}$$

Or

$$\mathbb{P}(T = t_0; \theta) = \mathbb{P}(T(X) = t_0; \theta) = \sum_{y, t(y)=t_0} \mathbb{P}(X = y; \theta).$$

Donc, pour $t(x) = t_0$, on a :

$$\begin{aligned} \mathbb{P}(X = x|T = t_0; \theta) &= \frac{\mathbb{P}(X = x; \theta)}{\sum_{y, t(y)=t_0} \mathbb{P}(X = y; \theta)} = \frac{g(t(x); \theta)h(x)}{\sum_{y, t(y)=t_0} g(t(y); \theta)h(y)} \\ &= \frac{g(t_0; \theta)h(x)}{\sum_{y, t(y)=t_0} g(t_0; \theta)h(y)} = \frac{h(x)}{\sum_{y, t(y)=t_0} h(y)} \end{aligned}$$

qui ne dépend pas de θ . Donc T est exhaustive, d'où le théorème.

Exemple 2.1.2. Statistiques exhaustives dans le modèle de Poisson

Soit X_1, \dots, X_n issu de $X \rightsquigarrow \mathcal{P}(\theta)$, θ inconnu. La statistique $T(X) = \sum_{i=1}^n X_i$ est-elle exhaustive pour θ ?.

On a :

$$f(x, \theta) = e^{-\theta} \frac{\theta^x}{x!}, T = \sum_{i=1}^n X_i \rightsquigarrow \mathcal{P}(n\theta).$$

Notons P_θ la quantité $\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T = t)$.

$$\begin{aligned}
 P_\theta &= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T = t)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, X_n = t - \sum_{i=1}^{n-1} x_i)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\mathbb{P}_\theta(X_1 = x_1) \dots \mathbb{P}_\theta(X_n = x_n)}{\mathbb{P}_\theta(T = t)} \\
 &= \frac{\prod_{i=1}^n f(x_i, \theta)}{f(t, \theta)} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \cdot \frac{t!}{e^{-n\theta} (n\theta)^t} \\
 &= \frac{t!}{\prod_{i=1}^n x_i! n^t} = \frac{(\sum_{i=1}^n x_i)!}{\prod_{i=1}^n x_i!} \left(\frac{1}{n}\right)^{\sum_{i=1}^n x_i} \text{ indépendante de } \theta.
 \end{aligned}$$

Donc T est exhaustive pour θ .

Ou bien :

$$\begin{aligned}
 L(\underline{x}, \theta) &= \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \left(\frac{\theta^{x_i} e^{-\theta}}{x_i!} \right) = \frac{\theta^t e^{-n\theta}}{\prod_{i=1}^n x_i!} \\
 &= \frac{(n\theta)^t e^{-n\theta}}{t!} \cdot \frac{t!}{n^t (\prod_{i=1}^n x_i!)} \\
 &= g(t, \theta) \cdot h(\underline{x}).
 \end{aligned}$$

D'après le principe de factorisation, T est exhaustive pour θ .

Remarque 2.1.1. Le principe de factorisation nous donne donc un moyen de reconnaître si une statistique est exhaustive, mais permet difficilement de la construire ou même de savoir s'il en existe une.

Le théorème de Darrois répond à ces deux préoccupations :

Définition 2.2. (Famille exponentielle). Soit X une variable aléatoire réelle, dont la loi de probabilité dépend d'un paramètre θ . On dit que la loi de X appartient à la famille exponentielle si et seulement si $\mathbb{P}(X = x; \theta)$ (cas discret) ou $f_X(x; \theta)$ (cas continu) est de la forme :

$$\exp \{a(x)\alpha(\theta) + b(x) + \beta(\theta)\}.$$

Ce théorème est un util très puissant dans la recherche des statistiques exhaustives.

Exemple 2.1.3. La plupart des lois usuelles appartiennent à la famille exponentielle :

• **La loi Bernoulli $\mathcal{B}(p)$:**

$$\begin{aligned} P(X = x; p) &= \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases} = p^x(1 - p)^{1-x} = \exp [x \ln p + (1 - x) \ln(1 - p)] \\ &= \exp [x(\ln p - \ln(1 - p)) + \ln(1 - p)] \\ &= \exp \left[x \ln \frac{p}{1 - p} + \ln(1 - p) \right] \end{aligned}$$

qui est de la forme souhaitée avec

$$a(x) = x, \alpha(p) = \ln \frac{p}{1 - p}, b(x) = 0 \text{ et } \beta(p) = \ln(1 - p).$$

• **Loi exponentielle $\exp(\lambda)$.**

$$f_X(x; \lambda) = \lambda e^{-\lambda x} = \exp[-\lambda x + \ln \lambda]$$

qui est de la forme souhaitée avec

$$a(x) = x, \alpha(\lambda) = -\lambda, b(x) = 0, \text{ et } \beta(p) = \ln \lambda.$$

Le lien entre famille exponentielle et exhaustivité est donné par le théorème suivant :

Théorème 2.1.2. (Théorème de Darmois).

Soit une variable aléatoire X dont le domaine de définition ne dépend pas de θ . Une condition nécessaire et suffisante pour que l'échantillon (X_1, \dots, X_n) admette une statistique exhaustive est que la forme de la densité soit :

$$f(x, \theta) = \exp[a(x)\alpha(\theta) + b(x) + \beta(\theta)] \quad (\text{famille exponentielle}).$$

Si la densité est de cette forme et si de plus l'application $x_i \rightarrow \sum_{i=1}^n a(x_i)$ est bijective et continûment différentiable pour tout i , alors $T = \sum_{i=1}^n a(x_i)$ est une statistique exhaustive particulière.

Exemples

- Loi de Bernoulli de paramètre p inconnu : $T = \sum_{i=1}^n X_i$ est exhaustive pour p .
- Loi normale $\mathcal{N}(\mu, \sigma)$
 - si σ connu, $T = \sum_{i=1}^n X_i$ est exhaustive pour μ
 - si μ est connu, $T = \sum_{i=1}^n (X_i - \mu)^2$ est exhaustive pour σ^2
- Loi exponentielle de densité $\frac{1}{\lambda}e^{-\frac{x}{\lambda}}$, λ inconnu : $T = \sum_{i=1}^n X_i$ est exhaustive pour λ .

Remarques

- On peut remarquer que toute fonction injective d'une statistique exhaustive est encore exhaustive.
- Une statistique exhaustive T , qui est fonction de toute statistique exhaustive est dite exhaustive minimale.

On a relié la notion d'exhaustivité à celle d'information sans définir précisément l'information. Il y a en fait un lien entre l'exhaustivité et l'information de Fisher, comme on le verra plus tard.

2.2 Cas multidimensionnel

Théorème 2.2.1. (Théorème de factorisation de Fisher-Neyman). Dans un modèle d'échantillonnage $(\chi; \mathcal{A}; \mathbb{P}_\theta; \theta \in \Theta \subset \mathbb{R}^p)^n$, pour qu'une statistique T soit exhaustive pour θ , il faut et il suf-

fit qu'il existe deux fonctions mesurables g et h telles que :

$$\forall x \in \chi, \forall \theta \in \Theta, L(x; \theta) = g(t(x), \theta)h(x).$$

Exemple 2.2.1. Échantillon de loi normale $(\mathcal{N}(\mu, \sigma^2))$.

La vraisemblance de l'échantillon est donnée par :

$$L(x; \mu; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right)\right]$$

qui est de la forme $g((\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2); \mu, \sigma^2)$, donc le couple $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ ou (\bar{X}, S^2) est une statistique exhaustive pour le paramètre $\theta = (\mu, \sigma^2)$.

Théorème 2.2.2. (Théorème de Darmois). Dans un modèle d'échantillonnage

$(\chi; \mathcal{A}; \mathbb{P}_\theta; \theta \in \Theta \subset \mathbb{R}^p)^n$, où le support de la loi des observations ne dépend pas de θ , il existe une **statistique exhaustive** si et seulement si cette loi appartient à la famille exponentielle. Alors

$$T(x) = \left(\sum_{i=1}^n a_1(x_i), \dots, \sum_{i=1}^n a_d(x_i) \right)$$

est une **statistique exhaustive**.

Exemple 2.2.2. Loi normale $\mathcal{N}(\mu; \sigma^2)$, $\theta = (\mu, \sigma^2) : a_1(x) = x$ et $a_2(x) = x^2$, donc on retrouve $T(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ ou (\bar{X}, S^2) est une statistique exhaustive pour le couple (μ, σ^2) .

Chapitre 3

Qualité d'un estimateur et méthodes d'estimation

3.1 Introduction

L'estimation consiste à donner des valeurs approchées aux paramètres d'une population (μ, σ , etc) à l'aide d'un échantillon de n observations issues de cette population. Les lois des grands nombres justifient l'usage de \bar{X} et S^2 comme estimation de μ et de σ^2 respectivement : on sait que $\bar{X} \xrightarrow{P.s} \mu$ et $S^2 \xrightarrow{P.s} \sigma^2$. De même la fréquence empirique F d'un événement est une estimation de sa probabilité. Les variables aléatoires \bar{x}, S^2, F sont appelées alors des **estimateurs** de μ, σ^2 et p respectivement.

Cependant le même paramètre peut être estimé à l'aide d'estimateurs différents : pour une distribution. Afin de choisir entre plusieurs estimateurs possibles d'un même paramètre il faut définir les qualités exigées d'un estimateur.

3.2 Estimateur ponctuel

On souhaite estimer un paramètre θ d'une population. Un estimateur de θ est une statistique T (donc une fonction de (X_1, \dots, X_n)) dont la réalisation est envisagée comme une **"bonne valeur"** du paramètre θ . On parle d'estimation de θ notée $\hat{\theta}$ associée à cet estimateur la valeur observée lors de l'expérience, c'est-à-dire la valeur prise par la fonc-

tion au point observé (x_1, \dots, x_n) :

$$\hat{\theta} = f(x_1, \dots, x_n).$$

3.3 Qualité d'un estimateur

Définition 3.1. (Estimateur sans biais). On appelle biais de T pour θ la valeur

$$\text{Biais}(T) = \mathbb{E}(T) - \theta.$$

- Un estimateur T est dit **sans biais** si $\mathbb{E}(T) = \theta$.

"Cela signifie que la valeur moyenne de T est égale à θ ".

Remarque 3.3.1. De façon générale, on peut écrire

$$T - \theta = (T - \mathbb{E}(T)) + (\mathbb{E}(T) - \theta),$$

ainsi

- la grandeur $T - \mathbb{E}(T)$ représente les fluctuations de T autour de sa moyenne
- et $\mathbb{E}(T) - \theta$ représente l'erreur systématique (biais).

Définition 3.2. (Erreur quadratique moyenne). La qualité d'un estimateur se mesure également par l'erreur quadratique moyenne ou (**risque quadratique**) définie par

$$R(T, \theta) = \mathbb{E}[(T - \theta)^2].$$

Théorème 3.3.1. Soit T un estimateur du paramètre θ à étudier. On a :

$$R(T, \theta) = \mathbb{E}[(T - \theta)^2] = \text{Var}(T) + (\text{Biais}(T))^2 = \text{Var}(T) + [\mathbb{E}(T) - \theta]^2.$$

Preuve.

$$\begin{aligned} \mathbb{E}[(T - \theta)^2] &= \mathbb{E}[(T - \theta + \mathbb{E}(T) - \mathbb{E}(T))^2] \\ &= \mathbb{E}[(T - \mathbb{E}(T))^2] + \mathbb{E}[(\mathbb{E}(T) - \theta)^2] + 2\mathbb{E}[(T - \mathbb{E}(T))(\mathbb{E}(T) - \theta)] \\ &= \text{Var}(T) + (\mathbb{E}(T) - \theta)^2 \quad \text{car } \mathbb{E}[(T - \mathbb{E}(T))(\mathbb{E}(T) - \theta)] = 0. \end{aligned}$$

Le critère d'erreur quadratique moyenne n'est pas parfait mais il est préféré à d'autres critères qui semblent plus naturels comme l'erreur absolue moyenne $\mathbb{E}(T - \theta)$ car il s'exprime en fonction de notions simples comme le biais et la variance et est relativement facile à manipuler analytiquement.

Théorème 3.3.2. (Meilleur estimateur). Soient T_1 et T_2 deux estimateur de θ . On dit que T_1 est meilleur que T_2 si :

$$R(T_1, \theta) < R(T_2, \theta), \forall \theta.$$

Théorème 3.3.3.

Si T est sans biais alors $R(T, \theta) = \text{Var}(T)$.

3.3.1 Performances asymptotiques d'un estimateur

Définition 3.3. (Estimateur asymptotiquement sans biais). Un estimateur T est dit asymptotiquement sans biais si $\lim_{n \rightarrow \infty} \mathbb{E}(T) = \theta$.

"Cela signifie que la valeur moyenne de T doit être proche de θ quand n est assez grand".

Remarque 3.3.2. Un estimateur sans biais est asymptotiquement sans biais.

Définition 3.4. (Estimateur convergent). Un estimateur T est dit **convergent** si

$$\lim_{n \rightarrow \infty} \text{Var}(T) = 0.$$

Définition 3.5. (Estimateur faiblement consistant). On dit qu'un estimateur T est **faiblement consistant** si T converge en probabilité vers θ lorsque n tend vers l'infini, i.e. pour tout $\epsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}[(T - \theta) \geq \epsilon] = 0.$$

" On l'interprète comme le fait que la probabilité de s'éloigner de la valeur à estimer de plus de ϵ tend vers 0 quand la taille de l'échantillon augmente. Cela signifie que T s'écarte de θ avec une faible probabilité quand n est assez grand".

Définition 3.6. (Estimateur fortement consistant). On dit qu'un estimateur T est **fortement consistant** si et seulement si T converge presque sûrement vers θ lorsque n tend vers l'infini, i.e.

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} T = \theta \right] = 1.$$

"Cela signifie que T est proche de θ quand n est assez grand".

La consistance est la propriété la plus importante d'un estimateur. Ainsi, pour de grands échantillons, l'approximation de θ par T est correcte. Par exemple, la loi forte des grands nombres affirme que, \bar{X}_n est un estimateur consistant de la moyenne μ .

Théorème 3.3.4. Si T est convergent et de variance tendant vers 0 lorsque n tend vers l'infini alors T est consistant.

Preuve. On a, pour tous réels θ et $\alpha > 0$,

$$|T - \theta| > \alpha \Rightarrow |\mathbb{E}(T) - T| > \alpha - |\theta - \mathbb{E}(T)|.$$

Si $\lim \mathbb{E}(T) = \theta$, alors à partir d'un certain rang N , on a $|\theta - \mathbb{E}(T)| \leq \frac{1}{2}$. Ainsi

$$\begin{aligned} \mathbb{P}[|T - \theta|] &= \mathbb{P}[|\mathbb{E}(T) - T| > \alpha - |\theta - \mathbb{E}(T)|] \\ &= \mathbb{P}\left(|\theta - \mathbb{E}(T)| \leq \frac{1}{2}\right) \\ &= \frac{4}{\alpha^2} \text{Var}(T), \end{aligned}$$

borne supérieure qui tend vers 0 lorsque n tend vers l'infini.

Proposition 3.3.5. Soit T un estimateur convergent du paramètre θ , et ϕ une fonction de \mathbb{R} dans \mathbb{R} , continue au point θ . Alors $(\phi(T))$ est un estimateur convergent de $\phi(\theta)$.

Exemple 3.3.1. Considérons par exemple comme modèle la loi uniforme sur $[0, \theta]$, où le paramètre θ est inconnu. La moyenne empirique \bar{X}_n est un estimateur convergent de l'espérance de la loi, qui vaut $\theta/2$. Donc $T = 2\bar{X}_n$ est un estimateur convergent de θ .

Exemples

1. \bar{X} est un estimateur sans biais de la moyenne μ . Son estimation \bar{x} est la moyenne observée dans une réalisation de l'échantillon.
2. S_n^2 est un estimateur consistant de σ^2 (mais biaisé).
3. $S_n^* = \frac{n}{n-1} S_n^2$ est un estimateur sans biais et consistant de σ^2 .
4. Si p est la fréquence d'un caractère, F constitue un estimateur sans biais et consistant de p . Son estimation est notée f .

Entre deux estimateurs sans biais, le "meilleur" sera celui dont la variance est minimale (on parle d'efficacité).

3.4 Réduction de la variance et statistique complète

Le théorème suivant permet, à partir d'un estimateur sans biais, de construire un autre estimateur sans biais de variance inférieure, pour peu qu'il existe une statistique **exhaustive**.

Théorème 3.4.1. (Théorème de Rao-Blackwell). S'il existe une statistique **exhaustive** T et un estimateur sans biais T_1 de θ , alors $Z = \mathbb{E}[T_1|T]$ est un estimateur sans biais de θ , de variance inférieure à celle de T_1 .

Définition 3.7. (Statistique complète). Une statistique T est **complète** ou totale si et seulement si pour toute fonction mesurable φ , on a :

$$\mathbb{E}(\varphi(T)) = 0, \forall \theta \in \Theta \Rightarrow \varphi = 0$$

presque partout sur le support de la loi de T ; c'est-à-dire partout sauf sur un ensemble de mesure nulle.

Exemple 3.4.1. (Contrôle de qualité).

$X = (X_1, \dots, X_n)$, où les X_i sont i.i.d. de loi de Bernoulli $\mathcal{B}(p)$. On sait que

$T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ est une statistique exhaustive pour p . Est-elle complète ?

On sait que $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ est de loi binomiale $\mathcal{B}(n; p)$, donc :

$$\mathbb{E}(\varphi(T)) = \sum_{k=0}^n \varphi(k) \mathbb{P}(T = k) = \sum_{k=0}^n \varphi(k) C_n^k p^k (1-p)^{n-k}.$$

Il faut montrer que

$$\sum_{k=0}^n \varphi(k) C_n^k p^k (1-p)^{n-k} = 0, \forall p \in [0, 1] \Rightarrow \forall k \in \{0, 1, \dots, n\}, \varphi(k) = 0.$$

En effet, comme le support de T est fini, doit être nulle partout sur le support.

Or,

$$\sum_{k=0}^n \varphi(k) C_n^k p^k (1-p)^{n-k} = (1-p)^n \sum_{k=0}^n \varphi(k) C_n^k \left[\frac{p}{1-p} \right]^k.$$

Soit $\theta = \frac{p}{1-p}$. On a

$$\sum_{k=0}^n \varphi(k) C_n^k p^k (1-p)^{n-k} = 0, \forall p \in [0, 1] \Rightarrow \sum_{k=0}^n \varphi(k) C_n^k \theta^k = 0, \forall \theta \in \mathbb{R}^+.$$

C'est un polynôme de degré n en θ qui est identiquement nul, donc tous ses coefficients sont nuls. Par conséquent,

$$\forall k \in \{0, 1, \dots, n\}, \varphi(k) C_n^k = 0,$$

et donc

$$\forall k \in \{0, 1, \dots, n\}, \varphi(k) = 0,$$

ce qui prouve que $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ est une statistique complète.

Les notions d'exhaustivité et de complétude permettent de trouver un ESBVM (estimateur sans biais de variance minimale) de θ à partir d'un estimateur sans biais.

3.5 Estimateur sans biais et de variance minimale

Théorème 3.5.1. (Théorème de Lehmann-Scheffé). Si T_1 est un estimateur sans biais de θ et T est une statistique **exhaustive et complète**, alors $Z = \mathbb{E}[T_1|T]$ est l'unique estimateur sans biais de θ , de variance minimale parmi tous les estimateurs sans biais de θ .

Corollaire 3.5.2. Pour trouver un estimateur optimal, il suffit de trouver un estimateur sans biais fonction d'une statistique **exhaustive et complète**.

Exemple 3.5.1. (Contrôle de qualité).

$T_1 = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur sans biais de p , fonction de la statistique exhaustive et complète $\sum_{i=1}^n X_i$, donc c'est l'ESBVM de p .

Propriété 3.5.3. Le théorème de **Lehmann-Scheffé** reste valable si on remplace θ par $\varphi(\theta)$, où φ est une fonction mesurable quelconque. Autrement dit, l'ESBVM de $\varphi(\theta)$ est un estimateur sans biais de $\varphi(\theta)$ fonction d'une statistique exhaustive et complète.

Théorème 3.5.4. Dans un modèle statistique où la loi des observations appartient à la famille exponentielle, si $\alpha(\theta)$ est bijective, alors la statistique **exhaustive** $\sum_{i=1}^n a(x_i)$ est complète.

3.6 Information de Fisher et efficacité

On a dit qu'une statistique exhaustive contenait autant d'information sur θ que l'observation x toute entière, mais on n'a pas défini ce qu'était l'information sur un paramètre. Il y a en fait plusieurs façons de la définir.

"Intuitivement, l'information mesure la capacité de l'observation à estimer avec précision le paramètre θ ".

Définition 3.8. On appelle quantité **d'information de Fisher** $I_n(\theta)$ apportée par un échantillon sur le paramètre θ , la quantité suivante positive ou nulle (si elle existe) :

$$I_n(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln L(X_1, \dots, X_n; \theta) \right)^2 \right].$$

Théorème 3.6.1. Si le domaine de définition de X ne dépend pas de θ alors :

$$I_n(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln L(X_1, \dots, X_n; \theta) \right].$$

Propriété 3.6.2. Si le domaine de définition de X ne dépend pas de θ alors :

$$I_n(\theta) = nI_1(\theta) = -n\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln f(X; \theta) \right].$$

Cette propriété traduit l'idée naturelle que, dans un échantillon, chaque observation porte la même quantité d'information sur θ , et que la quantité d'information est additive.

Exemple 3.6.1. Soit $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Calculer l'information de Fisher apportée par un n -échantillon issu de X sur le paramètre μ .

$$\begin{aligned}
I_n(\mu) &= nI_1(\mu) = -n\mathbb{E}\left[\frac{\partial^2}{\partial\mu^2}\ln(f_X(x;\mu))\right] \\
\ln(f_X(x,\mu)) &= \ln\left[\frac{e^{-\frac{1}{2\sigma^2}(x-\mu)^2}}{\sigma\sqrt{2\pi}}\right] = -\left(\frac{1}{2\sigma^2}(x-\mu)^2\right) - \ln(\sigma\sqrt{2\pi}) \\
\frac{\partial}{\partial\mu}\ln(f_X(x,\mu)) &= \frac{1}{\sigma^2}(x-\mu) \\
\frac{\partial^2}{\partial\mu^2}\ln(f_X(x,\mu)) &= -\frac{1}{\sigma^2} \Rightarrow I_n(\mu) = \frac{n}{\sigma^2}.
\end{aligned}$$

L'intérêt principal de la quantité d'information est qu'elle fournit une borne inférieure pour la variance de n'importe quel estimateur sans biais de θ .

3.6.1 Borne de Freshet-Damois-Cramer-Rao (FDCR)

Théorème 3.6.3. (Inégalité de FDCR).

1. Si le domaine de définition de X ne dépend pas de θ alors pour tout estimateur sans biais :

$$Var(T) \geq \frac{1}{I_n(\theta)}.$$

2. Si T est un estimateur sans biais de $h(\theta)$, alors

$$Var(T) \geq \frac{1}{I_n(h(\theta))}.$$

Estimateur efficace

Définition 3.9. Un estimateur T est dit **efficace** si sa variance est égale à la borne de FDCR.

Propriété 3.6.4.

1. Un estimateur sans biais efficace est convergent.
2. Un estimateur efficace T est un estimateur sans biais de variance minimale.

Exemple 3.6.2. Soit (X_1, \dots, X_n) un échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

On a les propriétés suivantes :

- \bar{X} est un estimateur sans biais de la moyenne μ .
- $\mathbb{E}(\bar{X}) = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$, $\lim_{n \rightarrow \infty} Var(\bar{X}) = 0$ (\bar{X} est un estimateur convergent) .
- $I_n(\mu) = \frac{n}{\sigma^2}$, la borne de FDCR est $\frac{1}{I_n(\mu)} = \frac{\sigma^2}{n} = Var(\bar{X})$.

Donc \bar{X} est efficace pour μ et c'est un estimateur sans biais de variance minimale de μ .

3.7 Cas multidimensionnel

Dans cette section, nous allons approfondir cette notion d'information de Fisher, en commençant par la définir pour un paramètre θ de dimension d quelconque.

3.7.1 Score et matrice d'information

On se place dans un modèle paramétrique $(\chi, \mathcal{A}, \mathbb{P}_\theta, \theta \in \mathbb{R}^d)$. Le paramètre θ s'écrit donc $\theta = (\theta_1, \dots, \theta_d)^t$.

Quand on estime un paramètre de dimension d , les notions usuelles liées à l'estimation s'écrivent sous forme vectorielle. Par exemple :

- Le vecteur aléatoire $T = (T_1, \dots, T_n)^t$ est un estimateur sans biais de θ si $\mathbb{E}(T) = \theta$:

$$\mathbb{E}(T) = \begin{pmatrix} \mathbb{E}(T_1) \\ \mathbb{E}(T_2) \\ \vdots \\ \mathbb{E}(T_d) \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix} \text{ où } \forall j \in \{1, \dots, d\}, \mathbb{E}(T_j) = \theta_j.$$

- L'erreur quadratique moyenne de l'estimateur T est

$$\mathbb{E} [\|T - \theta\|^2] = \sum_{j=1}^d \mathbb{E} ((T_j - \theta_j)^2).$$

- Pour définir les notions qui vont suivre, on a besoin de faire les hypothèses suivantes :

1. Le support de \mathbb{P}_θ ne dépend pas de θ (ce qui, par exemple, exclut la loi uniforme sur $[0; \theta]$) et $\forall \theta, \forall x, L(x : \theta) > 0$.
2. $\ln L(x : \theta)$ est dérivable 2 fois par rapport à chaque composante θ_j de θ .

Sous ces hypothèses, on peut définir les quantités suivantes.

Définition 3.10. Le **score** est le gradient de la log-vraisemblance :

$$Z(X; \theta) = \nabla \ln L(x; \theta) = \begin{pmatrix} Z_1(X; \theta) \\ Z_2(X; \theta) \\ \vdots \\ Z_d(X; \theta) \end{pmatrix}, \quad \text{où } \forall j \in \{1, 2, \dots, d\}, Z_j(X; \theta) = \frac{\partial}{\partial \theta_j} \ln L(X; \theta)$$

Définition 3.11. La matrice d'information de Fisher $I(\theta)$ est la matrice de covariance du score, de terme général

$$I_{j,k}(\theta) = \text{Cov} [Z_j(X; \theta), Z_k(X; \theta)].$$

Propriété 3.7.1. Le score est centré :

$$\mathbb{E}(Z(X; \theta)) = 0.$$

On en déduit que :

$$\begin{aligned} I_{j,k}(\theta) &= \text{Cov} [Z_j(X; \theta), Z_k(X; \theta)] \\ &= \mathbb{E} [Z_j(X; \theta) Z_k(X; \theta)] \\ &= \mathbb{E} \left[\frac{\partial}{\partial \theta_j} \ln L(X; \theta) \frac{\partial}{\partial \theta_k} \ln L(X; \theta) \right] \\ &= -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln L(X; \theta) \right] \end{aligned}$$

Exemple 3.7.1. Soit (X_1, X_2, \dots, X_n) un n-échantillon issu de $\mathcal{N}(\mu, \sigma^2)$, avec μ, σ^2 inconnus.

$$\begin{aligned} \theta &= \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}. \\ L(x; \mu; \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \\ \ln L(x; \mu; \sigma^2) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \theta_1} \ln L(x; \theta) &= \frac{\partial}{\partial \mu} \ln L(x; \mu; \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\
\frac{\partial^2}{\partial \mu^2} \ln L(x; \mu; \sigma^2) &= -\frac{n}{\sigma^2} \\
\frac{\partial}{\partial \theta_2} \ln L(x; \theta) &= \frac{\partial}{\partial \sigma^2} \ln L(x; \mu; \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \\
\frac{\partial^2}{\partial (\sigma^2)^2} \ln L(x; \mu; \sigma^2) &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \\
\frac{\partial^2}{\partial \mu \partial \sigma^2} \ln L(x; \mu; \sigma^2) &= -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu)
\end{aligned}$$

Pour calculer l'information de Fisher, on remplace la réalisation x par la variable aléatoire X , et on prend l'espérance

$$\begin{aligned}
I_{ij}(\theta) &= -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln L(X; \theta) \right] \\
I_{11}(\theta) &= -\mathbb{E} \left[\frac{\partial^2}{\partial^2 \mu^2} \ln L(X; \mu; \sigma^2) \right] = \frac{n}{\sigma^2} \\
I_{22}(\theta) &= -\mathbb{E} \left[\frac{\partial^2}{\partial^2 (\sigma^2)^2} \ln L(x; \theta) \right] = -\mathbb{E} \left[\frac{\partial^2}{\partial^2 (\sigma^2)^2} \ln L(X; \mu, \sigma^2) \right] \\
&= -\left[\frac{n}{2\sigma^4} - \frac{n}{\sigma^6} (\mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2) \right] \\
&= -\left[\frac{n}{2\sigma^4} - \frac{n}{\sigma^6} (\mathbb{E}(X^2) - \mu^2) \right] = -\left[\frac{n}{2\sigma^4} - \frac{n}{\sigma^6} (\sigma^2 + \mu^2 - \mu^2) \right] = -\frac{n}{2\sigma^4} + \frac{n}{\sigma^4} = \frac{n}{2\sigma^4} \\
I_{12}(\theta) = I_{21}(\theta) &= -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ln L(X; \theta) \right] = -\mathbb{E} \left[\frac{\partial^2}{\partial \mu \partial \sigma^2} \ln L(X; \mu, \sigma^2) \right] = -\left[-\frac{n}{\sigma^4} (\mathbb{E}(X) - \mu) \right] = 0.
\end{aligned}$$

La matrice d'information de **Fisher** est donnée par :

$$I_n(\theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

Théorème 3.7.2. Inégalité de Fréchet-Darmois-Cramer-Rao (FDCR). On considère $(\chi, \mathcal{A}, \mathbb{P}_\theta, \theta \in \mathbb{R}^d)$ un modèle paramétrique vérifiant les hypothèses de cette section et tel que la matrice d'information $I_n(\theta)$ soit inversible. Soit T une statistique à valeurs dans \mathbb{R}^q , Λ la matrice de covariance de T et Δ la matrice de terme général

$$\Delta_{ij} = \frac{\partial}{\partial \theta_j} \mathbb{E}(T), \forall i = \overline{1, q}, j = \overline{1, d}.$$

Alors $\forall \theta \in \mathbb{R}^d$, la matrice $\Lambda_T - \Delta I_n^{-1}(\theta) \Delta^t$ est semi-définie positive. Quand $d = q = 1$, alors : $\Lambda_T = \text{Var}(T)$, $\Delta = \frac{\partial}{\partial \theta} \mathbb{E}(T)$. Alors on obtient :

$$\text{Var}(T) - \frac{\left[\frac{\partial}{\partial \theta} \mathbb{E}(T) \right]^2}{I_n(\theta)} \geq 0.$$

Propriété 3.7.3. Quand $\theta \in \mathbb{R}^d$, l'inégalité FDCR appliquée aux termes diagonaux de Λ_T permet d'obtenir une borne inférieure pour la variance de chaque composante de T : $\forall i \in \{1, \dots, q\}$, on a

$$\text{Var}(T_i) = \sum_{k=1}^q \sum_{j=1}^d I_{jk}^{-1}(\theta) \frac{\partial \mathbb{E}(T_i)}{\partial \theta_j} \frac{\partial \mathbb{E}(T_i)}{\partial \theta_k}.$$

Propriété 3.7.4. En particulier, si T est un estimateur sans biais de θ , on a pour tout i , $\mathbb{E}(T_i) = \theta_i$. Donc

$$\frac{\partial \mathbb{E}(T_i)}{\partial \theta_j} = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{sinon} \end{cases}, \text{ d'où } \text{Var}(T_i) \geq I_{ii}^{-1}(\theta),$$

qui est la borne de Cramer-Rao.

Définition 3.12. Un estimateur sans biais T est efficace si et seulement si $\Lambda = I^{-1}(\theta)$. Alors, pour tout i ,

$$\text{Var}(T_i) = I_{ii}^{-1}(\theta).$$

Le théorème suivant donne une condition d'existence d'un estimateur efficace dans un modèle d'échantillonnage paramétrique, liée à la famille exponentielle.

Théorème 3.7.5. Dans un modèle statistique $(\chi, \mathcal{A}, \mathbb{P}_\theta, \theta \in \mathbb{R}^d)^n$, la borne de Cramer-Rao ne peut être atteinte que si \mathbb{P}_θ appartient à la famille exponentielle. La vraisemblance s'écrit :

$$L(X, \theta) = \exp \left[\sum_{i=1}^n \sum_{j=1}^d a_j(x_i) \alpha_j(\theta) + \sum_{i=1}^n b(x_i) + n\beta(\theta) \right].$$

3.8 Méthodes de calcul d'un estimateur

3.8.1 Méthode des Moments

Soit (X_1, X_2, \dots, X_n) un n -échantillon issu d'une variable aléatoire X de densité $f(x, \theta_1, \theta_2, \dots, \theta_k)$ où : $\theta_1, \theta_2, \dots, \theta_k$ sont des paramètres inconnus. La méthode des moments consiste à estimer les paramètres $\theta_1, \theta_2, \dots, \theta_k$, en égalisant les moments empiriques calculés à partir de l'échantillon avec les moments théoriques de même ordre. Soit $\mu_r = \mathbb{E}(X^r)$, $r = 1, 2, \dots, k$ moments d'ordre r de la population (théorique) et on note $M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$ moment empirique d'ordre r de l'échantillon. La solution du système $M_r = \mu_r$, $r = 1, \dots, k$ nous donne les estimateurs de $\theta_1, \theta_2, \dots, \theta_k$.

$$\begin{cases} M_1 = \mu_1, \\ M_1 = \mu_2, \\ \vdots \\ M_k = \mu_k \end{cases} \quad k \text{ équations à } k \text{ inconnus};$$

Propriété 3.8.1. Dans la plupart des cas, les estimateurs obtenus par la méthode des moments sont consistants, convergents, asymptotiquement normaux mais en général ne sont pas efficaces.

Exemple 3.8.1. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, où μ, σ^2 sont inconnus. Estimer μ et σ^2 par la méthode des moments.

$$\begin{aligned} \mu_1 &= \mathbb{E}(X) = \mu \\ \mu_2 &= \mathbb{E}(X^2) = \sigma^2 + \mu^2 \\ M_1 &= \frac{1}{n} \sum_{i=1}^n X_i \\ M_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2. \end{aligned}$$

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2. \end{cases},$$

l'estimateur de μ est $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$.

$$\begin{aligned}\sigma^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.\end{aligned}$$

3.8.2 Méthode du maximum de vraisemblance

Définition 3.13. La méthode consistant à estimer θ par la valeur qui maximise L (vraisemblance) s'appelle méthode du maximum de vraisemblance.

$$\hat{\theta} = \left\{ \theta / L(\hat{\theta}) = \sup_{\theta} L(x; \theta) \right\}.$$

La vraisemblance étant positive et le logarithme népérien une fonction croissante, il est équivalent et souvent plus simple de maximiser le logarithme népérien de la vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver). Cette méthode consiste à résoudre :

$$\begin{cases} \frac{\partial L(x_1, \dots, x_n; \theta)}{\partial \theta} = 0 & \text{ou} & \frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} = 0, & \text{permet de trouver la valeur } \hat{\theta} \\ \frac{\partial^2 L(x_1, \dots, x_n; \theta)}{\partial \theta^2} \leq 0 & \text{ou} & \frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} \leq 0, & \text{pour assurer l'existence du } \max_{\theta} L(x; \theta) \end{cases}$$

Exemple 1 : Avec une loi discrète

On souhaite estimer le paramètre λ d'une loi de Poisson à partir d'un n -échantillon. On a $f(x; \lambda) = P_\lambda(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$. La fonction de vraisemblance s'écrit

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda n} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}.$$

Il est plus simple d'utiliser le logarithme, la vraisemblance étant positive :

$$\ln L(x_1, \dots, x_n; \lambda) = \ln e^{-\lambda n} + \ln \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} = -\lambda n + \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} = -\lambda n + \ln \lambda \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!).$$

La dérivée première

$$\frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} = -n + \frac{\sum_{i=1}^n x_i}{\lambda}$$

s'annule pour $\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$. La dérivée seconde

$$\frac{\partial^2 \ln L(x_1, \dots, x_n; \theta)}{\partial \theta^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2}$$

est toujours négative ou nulle. Ainsi l'estimation donnée par $\Lambda = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ conduit à un estimateur du maximum de vraisemblance égal à $\hat{\lambda} = \bar{x}$. Il est normal de retrouver la moyenne empirique qui est le meilleur estimateur possible pour le paramètre λ (qui représente aussi l'espérance d'une loi de Poisson).

Exemple 2 : Avec une loi continue

On souhaite estimer les paramètres μ et σ d'une loi normale à partir d'un n -échantillon.

La loi normale $\mathcal{N}(\mu, \sigma)$ a pour fonction densité

$$f(x; \mu, \sigma) = f_{(\mu, \sigma)}(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Ecrivons la fonction de vraisemblance pour une réalisation d'un échantillon de n variables indépendantes :

$$f(x_1, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n f(x_i; \mu, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right).$$

Or $\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$, où \bar{x} représente la moyenne de l'échantillon. Ainsi la fonction de vraisemblance peut être écrite sous la forme

$$f(x_1, \dots, x_n; \mu, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

$$\frac{\partial}{\partial \mu} \ln L = \frac{\partial}{\partial \mu} \left(\ln \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} - \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) \right) = 0 - \frac{-2n(\bar{x} - \mu)}{2\sigma^2}$$

On obtient donc l'estimateur par le maximum de vraisemblance de l'espérance :

$$\hat{\mu} = \bar{x} = \sum_{i=1}^n x_i / n.$$

Pour le second paramètre, on calcule

$$\frac{\partial}{\partial \sigma} \ln L = \frac{\partial}{\partial \sigma} \left(\frac{n}{2} \ln \left(\frac{1}{2\pi\sigma^2} \right) - \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{\sigma^3}$$

Donc

$$\hat{\sigma}^2 = \sum_{i=1}^n (x_i - \hat{\mu})^2 / n$$

que l'on peut traduire par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

On vérifie que c'est bien des maxima locaux :

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -n/\sigma^2 \leq 0$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = n/\sigma^2 - \frac{3}{\sigma^4} (\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2).$$

Au point $\hat{\sigma}$,

$$\frac{\partial^2 \ln L}{\partial \sigma^2}(\hat{\sigma}) = n/\hat{\sigma}^2 - \frac{3}{\hat{\sigma}^4} (n\hat{\sigma}^2 + n(\bar{x} - \mu)^2) \leq 0.$$

La méthode fournit un estimateur non biaisé de la moyenne ($E(\hat{\mu}) = \mu$) mais par contre, l'estimateur de la variance est biaisé ($E(\hat{\sigma}^2) = \frac{n}{n-1}\sigma^2$). Néanmoins l'estimateur est asymptotiquement sans biais.

3.8.3 Estimation par intervalle de confiance

Au lieu de se donner une fonction (estimateur) qui donne une estimation ponctuelle d'un paramètre, on cherche un intervalle dans lequel se trouve le paramètre étudié avec une probabilité contrôlée (et généralement grande).

Définition 3.14. Un intervalle de confiance pour le paramètre θ , de niveau de confiance $1 - \alpha \in]0, 1[$, est l'intervalle qui a la probabilité $1 - \alpha$ de contenir la vraie valeur du paramètre θ .

"Un intervalle de confiance indique la précision d'une estimation car pour un risque α donné, l'intervalle est d'autant plus grand que la précision est faible".

Principe de construction

Le principe de la méthode d'estimation par intervalle de confiance est le suivant : Soit T un estimateur de θ (meilleur estimateur possible) dont on connaît la loi en fonction de θ . On détermine par la suite un intervalle de probabilité de niveau $1 - \alpha$ pour T i.e :

$$\mathbb{P}(t_1(\theta) < T < t_2(\theta)) = 1 - \alpha.$$

Il faut ensuite inverser cet intervalle pour T , dont les bornes dépendent de θ pour obtenir un intervalle pour θ ,

$$\begin{aligned} \text{i.e. } \mathbb{P}_\theta(a(T) < \theta < b(T)) &= 1 - \alpha \\ \text{ou bien } \mathbb{P}(t_2^{-1}(\theta) < T < t_1^{-1}(\theta)) &= 1 - \alpha \end{aligned}$$

Intervalle de l'espérance μ d'une loi $\mathcal{N}(0, 1)$.

a. Cas où σ^2 est connue

$$\bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1).$$

$$\begin{aligned} \mathbb{P}\left(-u_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\ \mathbb{P}\left(\mu - \frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}\right) &= 1 - \alpha \end{aligned}$$

d'où :

$$IC_{(1-\alpha)}(\mu) = \left[\bar{x} - \frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}} \right].$$

où : $u_{\frac{\alpha}{2}}$ est le fractile d'ordre $(1 - \frac{\alpha}{2})$ de $\mathcal{N}(0, 1)$, i.e :

$$\mathbb{P} \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\frac{\alpha}{2}} \right) = 1 - \frac{\alpha}{2}.$$

D'où :

$$u_{\frac{\alpha}{2}} = \phi^{-1} \left(1 - \frac{\alpha}{2} \right).$$

Exemple 3.8.2. Pour $1 - \alpha = 0.95$

$$\mathbb{P} \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{\frac{\alpha}{2}} \right) = 1 - \frac{\alpha}{2} = 0.975.$$

Sur la table de $\mathcal{N}(0, 1)$ on aura $u_{\frac{\alpha}{2}} = 1.96$

$$IC_{(1-\alpha)}(\mu) = \left[\bar{x} - \frac{\sigma}{\sqrt{n}} 1.96, \bar{x} + \frac{\sigma}{\sqrt{n}} 1.96 \right].$$

b. Cas où σ^2 est inconnue :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n (\bar{X} - \mu)^2.$$

En divisant par σ^2 :

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{nS^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2.$$

Soit $Z_i = \left(\frac{X_i - \mu}{\sigma} \right)$

$$Z_i \rightsquigarrow \mathcal{N}(0, 1) \implies Z_i^2 \rightsquigarrow \chi_1^2.$$

D'où,

$$\sum_{i=1}^n Z_i^2 \rightsquigarrow \chi_n^2$$

(Comme somme de n carrés de variables aléatoires indépendantes normales centrées réduites).

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1) \implies \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \rightsquigarrow \chi_1^2.$$

Donc

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

D'où, on en déduit :

$$\frac{(n-1)S^{*2}}{\sigma^2} \rightsquigarrow \frac{\sigma^2}{n} \chi_{n-1}^2.$$

Alors la variable aléatoire :

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^{*2}}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{S^*} \sqrt{n} \rightsquigarrow T_{n-1},$$

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

$$\begin{aligned} \mathbb{P}\left(-t_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S^*} \sqrt{n} \leq t_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\ \mathbb{P}(|T_{n-1}| \leq t_{\alpha}) &= 1 - \alpha \\ \mathbb{P}(|T_{n-1}| > t_{\alpha}) &= \alpha \text{ (} t_{\alpha} \text{ est tiré de la table de Student).} \end{aligned}$$

On trouve l'intervalle de confiance pour μ :

$$\left[\bar{x} - t_{\frac{\alpha}{2}} \frac{S^*}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \frac{S^*}{\sqrt{n}} \right]$$

si $\alpha = 0.05, n = 10$ on trouve $t = 2.262$.

$$IC(\mu) = \left[\bar{x} - 2.262 \frac{S^*}{\sqrt{n}}, \bar{x} + 2.262 \frac{S^*}{\sqrt{n}} \right].$$

Intervalle de confiance pour σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$.

a. μ connu :

On a $T = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ est le meilleur estimateur de σ^2 .

$$\frac{nT}{\sigma^2} \rightsquigarrow \chi_n^2.$$

$$\mathbb{P}\left(k_1 \leq \frac{nT}{\sigma^2} \leq k_2\right) = 1 - \alpha.$$

L'intervalle de confiance pour σ^2 est :

$$\frac{nt}{k_2} \leq \sigma^2 \leq \frac{nt}{k_1}.$$

Avec $k_1 = \phi_{\chi_n^2}^{-1}(\frac{\alpha}{2})$ et $k_2 = \phi_{\chi_n^2}^{-1}(1 - \frac{\alpha}{2})$.

b. μ inconnu :

On a

$$\frac{(n-1)S^{*2}}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

$$\mathbb{P}\left(k_1 \leq \frac{(n-1)S^{*2}}{\sigma^2} \leq k_2\right) = 1 - \alpha.$$

D'où :

$$IC(\sigma^2) = \left[\frac{(n-1)S^{*2}}{k_2}; \frac{(n-1)S^{*2}}{k_1} \right].$$

Exemple 3.8.3. Si $1 - \alpha = 0.95$, on cherche :

$$k_1 \text{ tel que } \mathbb{P}\left(\frac{(n-1)S^{*2}}{\sigma^2} < k_1\right) = \frac{\alpha}{2} = 0.025.$$

$$k_2 \text{ tel que } \mathbb{P}\left(\frac{(n-1)S^{*2}}{\sigma^2} < k_2\right) = 1 - \frac{\alpha}{2} = 0.975.$$

Estimation d'une proportion par intervalle de confiance

On considère une population telle que pour le caractère observé la proportion p d'une certaine catégorie est inconnue. On souhaite estimer cette proportion p de cette population à partir d'un échantillon de taille n dont la fréquence de la catégorie étudiée est f . Soit F la variable aléatoire qui à chaque échantillon de taille n associe la fréquence du nombre d'éléments qui appartiennent à la catégorie choisie. On sait que F suit approximativement la loi $\mathcal{N}(p; \sigma)$ avec $\sigma = \sqrt{\frac{pq}{n}}$, pour n suffisamment grand ($n > 30$). On dispose de

$$\sigma' = \sqrt{\frac{f(1-f)}{n}}$$

l'écart type associé à la fréquence f de l'échantillon de taille n . On se sert de l'estimation ponctuelle de σ puisque p est inconnue :

$$\sigma' = \sigma \sqrt{\frac{n}{n-1}} = \sqrt{\frac{f(1-f)}{n}} \sqrt{\frac{n}{n-1}}.$$

Donc la variable aléatoire Z définie par :

$$Z = \frac{F - f}{\sigma'}$$

suit approximativement une loi normale centrée réduite $\mathcal{N}(0, 1)$. On cherche un intervalle de confiance de la proportion p , c'est-à-dire un intervalle tel que la probabilité que la proportion p n'appartienne pas à cet intervalle soit égale à α où $\alpha \in [0, 1]$. On appelle cet intervalle de confiance avec le risque α ou avec le coefficient de confiance $c = 1 - \alpha$.

Le risque que l'on prend à dire que p appartient à cet intervalle est donc de α ou encore la probabilité que p n'appartienne pas à cet intervalle est le risque α . Déterminons cet intervalle de confiance : On rappelle que l'on a défini $z_{\frac{\alpha}{2}}$ comme étant la valeur telle que

$$\mathbb{P}(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}.$$

où Z suit $\mathcal{N}(0, 1)$. A l'aide des propriétés de la loi normale centrée réduite, on a $\mathbb{P}(Z < -z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ et

$$\mathbb{P}(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

$$\begin{aligned} \mathbb{P}(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) &\Leftrightarrow \mathbb{P}\left(-z_{\frac{\alpha}{2}} < \frac{F - p}{\sigma} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\ &\Leftrightarrow \mathbb{P}(F - z_{\frac{\alpha}{2}}\sigma < p < F + z_{\frac{\alpha}{2}}\sigma) = 1 - \alpha \\ &\Leftrightarrow \mathbb{P}\left(F - z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n-1}} < p < F + z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n-1}}\right) = 1 - \alpha. \end{aligned}$$

L'intervalle de confiance de la proportion p avec un niveau de confiance de $1 - \alpha$ est :

$$IC(p) = \left[f - z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n-1}}, f + z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n-1}} \right].$$

Remarque 3.8.1. Lorsque n est grand, la différence entre n et $n-1$ devient négligeable, aussi la formule devient

$$IC(p) = \left[f - z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}, f + z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}} \right].$$

C'est la formule la plus couramment utilisée.