# Table des matières

**1** **2**

1.1 Rappels des lois fondamentales de l'échantillonnage . . . . . 2

1.1.1 Phénomène et échantillons aléatoires . . . . . 2

1.1.2 Moyenne, variance, moments empiriques . . . . . 3

1.2 Théorie de l'estimation paramétrique ponctuelle : . . . . . 5

1.2.1 La classe exponentielle de lois . . . . . 6

1.3 Qualités des estimateurs . . . . . 7

1.3.1 Biais d'un estimateur . . . . . 7

# Chapitre 1

Ce rappel ne constitue pas un cours élémentaire de Lois fondamentales de l'échantillonnage mais cherche à regrouper modestement certaines notions essentielles qui nous seront indispensables pour la suite.

## 1.1 Rappels des lois fondamentales de l'échantillonnage

### 1.1.1 Phénomène et échantillons aléatoires

Le terme d'échantillon est souvent associé à un sous ensemble de cardinal n tiré d'une population selon certaines règles. On s'intéresse aux échantillons de variables que l'on relie aux échantillons d'individus par la considération élémentaire suivante :
Sur chaque individu tiré, on mesure une certaine grandeur et on note $x_1, x_2...x_n$ les valeurs observées. $(x_1, x_2, ....x_n)$ est un échantillon de valeurs.

**Définition 1.1.1.** *On appelle échantillon aléatoire de taille n (en bref n-échantillon) une suite de n variables aléatoires indépendantes et de même loi (ou v.a.i.i.d). Cette loi est appelée la loi mère de l'échantillon.*

**Définition 1.1.2.** *Soit $X_1, X_2....X_n$un n-échantillon, on appelle* ***statistique*** *toute v.a $T_n = h(X_1, X_2, ..., X_n)$ fonction de $X_1, X_2, ...X_n$.*

*On peut concrétiser la loi d'une statistique (donc d'une caractéristique, telle la moyenne de l'échantillon) en imaginant une simulation en très grand nombre d'échantillons de taille n, en calculant pour chacun d'eux la valeur prise par la statistique et en étudiant la distribution de ces valeurs. De façon imagée on peut dire qu'il s'agit de la distribution d'échantillonnage de la statistique sur "l'univers" de tous les échantillons possibles. Notons qu'une statistique peut être une fonction à valeurs dans$\mathbb{R}, \mathbb{R}^2$ ou $\mathbb{R}^p$. En particulier les moments empiriques ci-après sont*

*à valeurs dans $\mathbb{R}$. Les définitions qui suivent se rapportent toutes à un échantillon aléatoire noté $X_1, X_2, ....X_n$*

### 1.1.2 Moyenne, variance, moments empiriques

**Définition 1.1.3.** *On appelle* ***moyenne de l'échantillon*** *ou* ***moyenne empirique*** *la statistique, notée $\overline{X}$, définie par :*

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

**Définition 1.1.4.** *On appelle* ***variance empirique*** *la statistique notée $S^{*2}$, définie par :*

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

**Proposition 1.1.1.** *Soit $\mu$ et $\sigma^2$ , respectivement la moyenne et la variance de la loi mère on a :*

$$E(\overline{X}) = \mu \quad V(\overline{X}) = \frac{\sigma^2}{n}$$

**Preuve .**

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Puis, en raison de l'indépendance des $X_i$

$$V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

■

**Proposition 1.1.2.** *La moyenne de la loi de la variance empirique est :*

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

**Preuve .**

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \frac{1}{n}\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)+(\overline{X}-\mu)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-2(\overline{X}-\mu)^2+(\overline{X}-\mu)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\overline{X}-\mu)^2\end{aligned}$$

D'où

$$E(S^{*2})=\frac{1}{n}\sum_{i=1}^{n}V(X_i)-V(\overline{X})=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

■

**Définition 1.1.5.** *On appelle* ***variance de l'échantillon*** *la statistique*

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

*Dorénavant on étudiera $S^2$ plutôt que $S^{*2}$ à laquelle on pourra éventuellement se référer en conservant le terme de variance empirique.*

**Proposition 1.1.3.** *Toute combinaison linéaire de v.a.gaussiennes indépendantes est une variable aléatoire gaussienne*

**Preuve .** Il suffit de démontrer cela avec deux v.a. l'extension à plusieurs v.a se faisant de proche en proche. De plus, on a vu que si X est gaussienne alors Y=aX est gaussienne. Il suffit donc de démontrer la proposition pour $Y_1+Y_2$ où $Y_1$ et$Y_2$ sont indépendantes.
Soient

$$Y_1\to\mathcal{N}(\mu,\sigma_1^2)\;\; Y_2\to\mathcal{N}(\mu,\sigma_2^2)$$

$$\begin{aligned}\Psi_{Y_1+Y_2}(t) &= \Psi_{Y_1}(t)\Psi_{Y_2}(t)\\ &= e^{t\mu_1+\frac{\sigma_1^2}{2}t^2}e^{t\mu_2+\frac{\sigma_2^2}{2}t^2}\\ &= e^{t(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}\end{aligned}$$

qui est la fonction génératrice des moments de la loi $\mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

**Loi mère gaussienne :**
Si la loi mère est $\mathcal{N}(\mu, \sigma^2)$ alors $\overline{X}$ est gaussienne, en tant que combinaison linéaire de gaussiennes indépendantes. Par conséquent :

$$\overline{X} \to \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

■

**Proposition 1.1.4.** *Si la loi mère est gaussienne, $\overline{X}$ et $S^2$ sont des v.a indépendantes.*

**Définition 1.1.6.** *On appelle* ***moment empirique*** *d'ordre r ,noté $M_r$ la statistique*

$$M_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

**Définition 1.1.7.** *On appelle* ***moment centré empirique*** *d'ordre r, noté $M'_r$, la statistique*

$$M'_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$$

## 1.2 Théorie de l'estimation paramétrique ponctuelle :

En estimation paramétrique la loi de X est réputée appartenir à une famille de lois dont la forme dependant d'un où plusieurs paramètres inconnus réels.on note $\theta$ ce paramètre ou vecteur de paramètres et $F(x, \theta)$, $f(x, \theta)$ ou $p(x, \theta)$ respectivement la fonction de répartition, la densité et la probabilité en $x$.

L'ensemble des valeurs possibles pour $\theta$ , appelée espace paramétrique sera noté $\Theta$, lequel est inclus dans $\mathbb{R}^k$ où k est la dimension du paramètre $\theta$ par exemple la famille des lois de Gauss est décrite par la famille des densités de la forme $1/\sigma\sqrt{2\pi}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$, pour tout $x \in \mathbb{R}$ où intervient un paramètre $(\mu, \sigma^2)$ de dimension 2, l'espace paramétrique étant la partie de $\mathbb{R}^2$ :
$\{(\mu, \sigma^2)\ \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}, \sigma^2 > 0\}$.

Dans ce cadre paramétrique le problème est celui de l'estimation du paramètre $\theta$ grâce à laquelle on obtiendra une estimation complète de la loi de $X$ par voie de conséquence, de toute caractéristique de cette loi. Distinguons bien ici la notion de paramètre d'une loi de celle de caractéristique (moyenne, variance, médiane, ...) de la loi : le paramètre identifie chaque loi (chaque membre) dans la famille considérée mais n'est pas nécessairement une caractéristique usuelle de cette loi. Par contre toute caractéristique usuelle dépend du membre de la famille et donc du paramètre $\theta$.

### 1.2.1 La classe exponentielle de lois

Cette classe regroupe des familles paramétriques de lois qui, de par leur forme particulière, partagent beaucoup de propriétés dans la théorie de l'estimation ou la thorie de l'estimation du fait que leurs densités peuvent s'écrire sous une même expression canonique.

**Définition 1.2.1.** *Soit une famille paramétrique de lois admettant des fonctions de densité (cas continu) ou des fonctions de probabilité (cas discret)* $\{f(x,\theta), \theta \in \Theta \subset \mathbb{R}^k\}$. *On dit qu'elle appartient à la classe exponentielle de lois si* $f(x,\theta)$ *peut s'écrire :*

$$f(x,\theta) = a(\theta)b(x) + \exp\{c_1(\theta)d_1(x) + c_2(\theta)d_2(x) + ..... + c_k(\theta)d_k(x)\}$$

*pour tout* $x \in \mathbb{R}$.
*En particulier, si* $\theta$ *est de dimension 1, on a*

$$a(\theta)b(x) + \exp\{c(\theta)d(x)\}$$

**Exemple 1.2.1.** *La famille des lois*$\{\mathcal{U}[0,1], \theta > 0\}$ *, on a*

$$f(x,\theta) = \frac{1}{\theta} I_{[0,\theta]}(x) \quad \forall x \in \mathbb{R}$$

**Remarque 1.2.1.** *Si n connu,* $\{\mathcal{B}(n,p), p \in ]0,1]\}$ *appartient à la famille exponentielle , ce qui ne serait pas le cas si n était inconnu.*

**Exemple 1.2.2.** *Loi* $\mathcal{B}(n,p)$ *avec n connu*

$$\begin{aligned} f(x,n,p) &= C_n^x p^x (1-p)^{n-x} \text{ pour } x = 0,1,2,....,n \\ &= (1-p)^n C_n^x \exp\{x \ln \frac{p}{1-p}\} \end{aligned}$$

*d'où d(x) = x. Le cas de la loi de Bernoulli* $\mathcal{B}(p)$ *est identique avec n = 1.*

**Exemple 1.2.3.** *Loi* $\mathcal{P}(\lambda)$

$$\begin{aligned} f(x,\lambda) &= e^{-\lambda}\frac{\lambda^x}{x!} \text{ pour } x \in \mathbb{N} \\ &= e^{-\lambda}\frac{1}{x!}\exp\{\ x.\ln\lambda\} \end{aligned}$$

*d'où, également,* $d(x) = x$, *et* $\mathcal{P}(\lambda)$ *appartient à la famille exponentielle.*

**Exemple 1.2.4.** *Loi* $\mathcal{N}(\mu,\sigma^2)$

$$\begin{aligned} f(x,\mu,\sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\{\frac{-1}{2}.\frac{(x-\mu)^2}{\sigma^2}\} \\ &= \frac{1}{\sigma\sqrt{2\pi}}\exp\{\frac{\mu^2}{2\sigma^2}\}\exp\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\} \end{aligned}$$

*d'où,* $d_1(x) = x^2$ *et* $d_2(x) = x$

## 1.3 Qualités des estimateurs

Un des objectifs essentiels de la théorie de l'estimation, est d'opérer des choix parmi les différents estimateurs auxquels on peut penser. Pour cela il est nécessaire de se donner des critères de qualité pertinents. De façon générique nous noterons $T_n$ l'estimateur de $\theta$ à étudier. Etant donné que la valeur de $\theta$ est inconnue, nous souhaitons que le comportement de $T_n$ soit satisfaisant quel que soit $\theta \in \Theta$, c'est-à-dire quelle que soit la loi mère effective dans la famille paramétrique donnée, et les critères de qualité seront à étudier comme des fonctions de $\theta$.

### 1.3.1 Biais d'un estimateur

**Définition 1.3.1.** *Soit une v.a. $X$ de loi de densité (ou fonction de probabilité) $f(x, \theta)$ où $\theta \in \Theta \subset \mathbb{R}$. Soit $X_1, X_2..., X_n$ un n-échantillon issu de cette loi et $T_n$ un estimateur de $\theta$. On appelle biais de $T_n$ pour $\theta$ la quantité :*

$$b_\theta(T_n) = E_\theta(T_n) - \theta$$

*Si $b_\theta(T_n) = 0$ quel que soit $\theta \in \Theta$, on dit que $T_n$ est sans biais pour $\theta$.*
*Si $E_\theta(T_n) \neq \theta$ on dit que $T_n$ est baisé.*