# Chapitre 2 : Estimation

*N. Biheng*

## 1 Cadre général

> « L'objectif de la statistique mathématique est, principalement, d'aider à établir un jugement sur un échantillon. »
> (Philippe Tassi in [5])

Notre point de départ sera une variable aléatoire $X$ dont l'ensemble des valeurs est noté $\mathcal{H}$. Dans ce chapitre, nous supposerons que la variable aléatoire suit une loi $\mathbb{P}_\theta$ où $\theta \in \Theta$. $\Theta$ sera appelé l'espace des paramètres.

**Exemple 1**
Nous savons que le poids des pommes d'un verger peut être modélisé par une loi normale. Si nous connaissons sa variance et ignorons $m$, nous aurons :
$\Theta = ]0; +\infty[ \subset \mathbb{R}$.
Notons que $m$ est strictement positif car il s'agit du poids moyen des pommes du verger.
Nous pourrions également ignorer les paramètres $m$ et $\sigma^2$.
Dans ce cas, $\Theta = ]0; +\infty[ \times ]0; +\infty[ \subset \mathbb{R}^2$.

**Définition 1**
Un *modèle statistique* sera la donnée de l'ensemble des valeurs que peut prendre la variable aléatoire $X$ noté $\mathcal{H}$, d'un ensemble $\Theta \subset \mathbb{R}^d$ appelé l'ensemble des paramètres et d'une famille $\mathcal{P} := \{\mathbb{P}_\theta | \theta \in \Theta\}$ de lois indexée par $\Theta$.

**Définition 2**
Un échantillon de taille $n$ est la donnée de $n$ variables aléatoires indépendantes et identiquement distribuées (i.i.d.) qui suivent la même loi que la variable aléatoire $X$.

**Définition 3**
Un *estimateur* est une fonction (mesurable) de $\hat{\theta} : \mathcal{H}^n \longrightarrow \mathbb{R}^d$.
Plus généralement, une *statistique* est une fonction (mesurable) définie sur $\mathcal{H}^n$.

Nous voudrons estimer $g(\theta)$ pour une certaine fonction $g$.

Par exemple, nous pourrons estimer $E(X) = \dfrac{1}{\lambda}$ pour la loi exponentielle.

Fondamentalement, un estimateur sera conçu comme une fonction des observations à valeurs dans $\mathbb{R}^d$.

**Définition 4**

Le biais d'un estimateur $\hat{\theta}$ de $g(\theta)$ est défini par : $b(\hat{\theta}) := E(\hat{\theta}) - g(\theta)$.

Lorsque $g(\theta) = \theta$, le biais est simplement $b(\hat{\theta}) = E(\hat{\theta}) - \theta$.

**Définition 5**

L'estimateur $\hat{\theta}$ dit sans biais lorsque $b(\hat{\theta}) = 0$.

**Exemple 2**

La moyenne empirique $\bar{X} := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ (parfois notée $\bar{X}_n$) est un estimateur sans biais de l'espérance de $X$.

Soit $(X_1, \ldots, X_n)$ un échantillon de taille $n$. Les variables aléatoires $X_i$ suivent la même loi que la variable aléatoire $X$ donc $E(X_i) = E(X)$.

En effet, par linéarité de l'espérance,

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \times nE(X) = E(X)$$

car $E(X_i) = E(X)$ pour tout $i$.

Ainsi, pour la loi normale $\bar{X}$ est un estimateur sans biais du paramètre $m$ et pour la loi géométrique $\bar{X}$ est un estimateur sans biais de $\dfrac{1}{p}$.

**Exemple 3**

Par contre, la variance empirique $S'^2 := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$ est biaisée en tant qu'estimateur de la variance.

En effet, $E(S'^2) = \dfrac{n-1}{n}\sigma^2$ où $\sigma^2$ désigne la variance de la variable aléatoire $X$ [1].

Lorsque, nous voudrons considérer un estimateur sans biais de la variance, nous aurons recours à la variance empirique modifiée (corrigée) $S^2 := \dfrac{n}{n-1}S'^2$ qui est clairement sans biais car

$E(S^2) = E\left(\dfrac{n}{n-1}S'^2\right) = \dfrac{n}{n-1} \times \dfrac{n-1}{n}\sigma^2 = \sigma^2$.

1. Nous renvoyons le lecteur vers [3] ou [5] pour la démonstration de ce résultat.

En dehors des contextes où nous étudierons la convergence d'un estimateur, nous noterons simplement $\hat{\theta}$ comme précédemment mais lorsque nous aborderons la convergence, il sera pertinent d'indiquer l'indice $n$ comme dans la définition suivante.

**Définition 6**
L'estimateur $\hat{\theta}_n$ est dit *asymptotiquement sans biais* si $\lim\limits_{n \longrightarrow +\infty} b(\hat{\theta}_n) = 0$.

Nous voyons aisément que $S_n'^2$ est asymptotiquement sans biais.
L'espérance d'un estimateur sans biais de $g(\theta)$ est $g(\theta)$ donc sa variance est l'écart en moyenne quadratique (ou la distance $L^2$) entre l'estimateur et la valeur estimée d'où la définition suivante :

**Définition 7**
Soient $\hat{\theta}^1$ et $\hat{\theta}^2$ deux estimateurs sans biais de $g(\theta)$.
L'estimateur $\hat{\theta}^1$ est *plus efficace* que l'estimateur $\hat{\theta}^2$ si $V(\hat{\theta}^1) < V(\hat{\theta}^2)$ pour tout $\theta \in \Theta$.

Nous définirons l'efficacité d'un estimateur en fin de chapitre.

**Définition 8**

1. Un estimateur $\hat{\theta}_n$ est dit *convergent* si $\hat{\theta}_n$ *converge en probabilité* vers $\theta$.
2. Un estimateur $\hat{\theta}_n$ est dit *fortement convergent* si $\hat{\theta}_n$ *converge presque sûrement* vers $\theta$.

La convergence presque sûre est beaucoup plus forte que la convergence en probabilité. La convergence presque sûre, dans ce contexte, signifie que $\hat{\theta}(X(\omega))$ converge vers $\theta$ (sauf éventuellement sur un sous-ensemble de mesure nulle). En dehors de cet éventuel sous-ensemble, il y a convergence de $\hat{\theta}(X(\omega))$ vers $\theta$ pour tout $\omega$.
Par contre, la convergence en probabilité signifie simplement, pour tout $\varepsilon > 0$, la probabilité de ne pas appartenir à l'intervalle $[\theta - \varepsilon; \theta + \varepsilon]$ peut être rendue aussi petite que nous le souhaitons pour $n$ assez grand.

**Théorème 1**
Soit $X$ une variable aléatoire de carré intégrable i.e. telle que $E(X^2) < +\infty$.

1. $\bar{X}$ est un estimateur fortement convergent de l'espérance.
2. $S^2$ et $S'^2$ sont des estimateurs fortement convergents de la variance.

Ce théorème est une conséquence directe de la loi forte des grands nombres.

# 2 Estimation paramétrique

## 2.1 Méthode des moments

Elle peut être exprimée à l'aide d'équations mais nous nous contenterons d'une présentation plus intuitive. Nous exprimerons le paramètre que nous voudrons estimer en fonction des moments $m_k := E(X^k)$ ou des moments centrés $m'_k := E((X - E(X))^k)$ notamment $m'_2 = V(X)$. Ensuite, nous remplacerons ces moments par les moments empiriques correspondants.

**Exemple 4**
Pour la loi géométrique, $E(X) = \dfrac{1}{p}$ donc $p = \dfrac{1}{E(X)}$.
A l'ordre 1, la méthode des moments nous donne l'estimateur $\hat{p} := \dfrac{1}{\bar{X}}$.

**Exemple 5**
Pour la loi Gamma[2] de paramètres $\alpha$ et $\beta$, il faudra résoudre un système :

$$\begin{cases} E(X) = \dfrac{\alpha}{\beta} \\ V(X) = \dfrac{\alpha}{\beta^2} \end{cases} \tag{1}$$

donc

$$\begin{cases} E(X) = \dfrac{\alpha}{\beta} \\ \dfrac{E(X)}{V(X)} = \dfrac{\frac{\alpha}{\beta}}{\frac{\alpha}{\beta^2}} = \beta \end{cases} \tag{2}$$

d'où

$$\begin{cases} \alpha = \dfrac{E(X)^2}{V(X)} \\ \beta = \dfrac{E(X)}{V(X)} \end{cases} \tag{3}$$

Nous en déduisons les estimateurs donnés par la méthode des moments : $\hat{\alpha} = \dfrac{\bar{X}^2}{S^2}$ et $\beta = \dfrac{\bar{X}}{S^2}$

2. cf. Complément sur la loi Gamma.

## 2.2 Méthode du maximum de vraisemblance

« C'est une vérité très certaine que si nous ne pouvons pas déterminer ce qui est vrai, nous devons suivre ce qui est le plus probable. »
(Descartes)

La loi dépend du paramètre $\theta$ donc la densité associée aussi, nous la noterons $f(x,\theta)$.
Nous supposerons $f(x,\theta) > 0$ pour tout $x$ et tout $\theta$.
Nous définirons sur $\mathcal{H}^n \times \Theta$ la *fonction de vraisemblance* par :

$$L(x_1,\ldots,x_n,\theta) := \prod_{i=1}^{n} f(x_i,\theta)$$

Pour obtenir le paramètre le plus probable à partir des observations $(x_1,\ldots,x_n)$, nous maximiserons sur $\Theta$ la fonction $L(x_1,\ldots,x_n,.)$ i.e. nous résoudrons le problème de maximisation suivant :

$$\left\{\begin{array}{l} \max L(x_1,\ldots,x_n,\theta) \\ \theta \in \Theta \end{array}\right. \tag{4}$$

Supposons que la densité soit de classe $\mathcal{C}^2$ par rapport à $\theta$ (i.e. deux fois dérivable par rapport à $\theta$ et dont la dérivée seconde par rapport à $\theta$ est continue).
Le problème est un problème d'optimisation différentiable. Nous pourrons ainsi considérer le système suivant :

$$\left\{\begin{array}{l} \dfrac{\partial L}{\partial \theta} = 0 \\ \dfrac{\partial^2 L}{\partial \theta^2} < 0 \end{array}\right. \tag{5}$$

La première condition est une condition nécessaire et la seconde est une condition suffisante. En pratique, la première condition nous permettra d'identifier des valeurs de $\theta$ susceptibles d'être des solutions.

Pour simplifier les calculs, nous remplacerons souvent la vraisemblance par la log-vraisemblance car la fonction logarithme est de classe $\mathcal{C}^2$ et strictement croissante sur $]0;+\infty[$.
Le système précédent est équivalent au système :

$$\left\{\begin{array}{l} \dfrac{\partial \log L}{\partial \theta} = 0 \\ \dfrac{\partial^2 \log L}{\partial \theta^2} < 0 \end{array}\right. \tag{6}$$

L'équation $\dfrac{\partial \log L}{\partial \theta} = 0$ s'appelle l'*équation de vraisemblance.*
Rappelons quelques propriétés du logarithme qui nous seront utiles. Soient $a$ et $b$ des réels strictement positifs et $x$ un réel, on a :

1. $\log(ab) = \log a + \log b$
2. $\log\left(\dfrac{a}{b}\right) = \log a - \log b$
3. $\log a^x = x \log a$

**Exemple 6**
Déterminons l'estimateur du maximum de vraisemblance pour la loi exponentielle de paramètre $\lambda$.
Par définition, $\mathcal{H} = [0; +\infty[$ et $\Theta = ]0; +\infty[$.
On sait que : $f(x, \lambda) = \lambda e^{-\lambda x}$.
Remarquons que$f(x, \lambda) > 0$ pour tout $x$ et tout $\lambda$.
Soit $n$ un entier naturel supérieur ou égal à 1.
Par définition, $L(x, \lambda) = \prod_{k=1}^{n} \lambda e^{-\lambda x_k} = \lambda^n e^{-\lambda \sum_{k=1}^{n} x_k}$.
Passons au logarithme :

$$\log L(x_1, \ldots, x_n, \lambda) = n \log \lambda - \lambda \sum_{k=1}^{n} x_k$$

Ecrivons les conditions du premier et second ordre :

$$\begin{cases} \dfrac{\partial \log L}{\partial \lambda} = 0 \\ \dfrac{\partial^2 \log L}{\partial \lambda^2} < 0 \end{cases} \tag{7}$$

On a : $\dfrac{\partial L}{\partial \lambda}(x, \lambda) = \dfrac{n}{\lambda} - \sum_{k=1}^{n} x_k$.

La condition nécessaire est que la solution $\hat{\lambda}(x)$ doit vérifier :

$$\frac{n}{\lambda} - \sum_{k=1}^{n} x_k = 0$$

D'où : $\hat{\lambda}(x) = \dfrac{n}{\sum_{k=1}^{n} x_k}$.

Pour conclure qu'il s'agit bien d'un maximum, vérifions la condition suffisante : $\frac{\partial^2 \log L}{\partial \lambda^2}(x_1, \ldots, x_n, \lambda) = -\frac{n}{\lambda^2} < 0$ pour tout $\lambda$ donc a fortiori pour $\hat{\lambda}(x)$.
La condition suffisante est satisfaite donc $\hat{\lambda}(x)$ est bien le maximum de vraisemblance.

Nous en déduisons l'estimateur du maximum de vraisemblance :
$\hat{\lambda}_n = \frac{n}{\sum_{k=1}^n X_k} = \frac{1}{\bar{X}_n}$. Notons qu'il est *fortement convergent* d'après la loi forte des grands nombres en notant que la fonction inverse est continue.

Etudions un autre exemple qui concernera une loi discrète.
Naturellement, la fonction de densité sera remplacée par $P(X = x)$.

**Exemple 7**
Déterminons l'estimateur du maximum de vraisemblance pour la loi de Poisson de paramètre $\lambda$. Ici, on a : $\mathcal{H} = \mathbb{N}$ et $\Theta = ]0; +\infty[$.
On sait que : $P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} > 0$ pour $x \in \mathbb{N}$ et $\lambda \in ]0; +\infty[$.
Par définition, $L(x_1, \ldots, x_n, \lambda) = \prod_{i=1}^n e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$.
Passons au logarithme :
$\log L(x_1, \ldots, x_n, \lambda) = -n\lambda + \sum_{i=1}^n x_i \log(\lambda) - \log \prod_{i=1}^n x_i!$.
Dérivons cette fonction par rapport à $\lambda$, on a :

$$\frac{\partial \log L}{\partial \lambda}(x_1, \ldots, x_n, \lambda) = -n + \frac{\sum_{i=1}^n x_i}{\lambda}$$

La condition nécessaire est que $\hat{\lambda}(x)$ doit être solution de :

$$-n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0$$

ce qui nous donne : $\hat{\lambda}(x) = \frac{1}{n}\sum_{k=1}^n x_k$.
Pour conclure, nous devons vérifier que la **condition suffisante** est satisfaite :

$$\frac{\partial^2 \log L}{\partial \lambda^2}(x_1, \ldots, x_n, \lambda) = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0$$

Ceci est vrai pour tout $\lambda$ donc a fortiori pour $\hat{\lambda}(x)$.
La condition suffisante est satisfaite donc $\hat{\lambda}(x)$ est bien le maximum de vraisemblance.
$\hat{\lambda}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k = \bar{X}_n$ est l'EMV et nous savons qu'il est *fortement convergent* d'après la loi forte des grands nombres.

**Exemple 8**
Déterminons l'estimateur du maximum de vraisemblance du couple $(m, \sigma^2)$ pour la loi normale $\mathcal{N}(m, \sigma^2)$.
La densité associée à cette loi est :

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

donc

$$L(x_1, \ldots, x_n, m, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\frac{-\sum_{i=1}^{n}(x_i-m)^2}{2\sigma^2}\right)$$

Passons au logarithme

$$\log L(x_1, \ldots, x_n, m, \sigma^2) = -n\log\sigma - n\log\sqrt{2\pi} - \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^2}$$

ou encore

$$\log L(x_1, \ldots, x_n, m, \sigma^2) = -\frac{n}{2}\log\sigma^2 - n\log\sqrt{2\pi} - \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^2}$$

car $\log(a^x) = x\log a$ donc $\log\sigma = \log((\sigma^2)^{\frac{1}{2}}) = \dfrac{1}{2}\log\sigma$.
Comme nous l'avons déjà vu (cf. correction des exercices), en dérivant par rapport à $m$, on obtient :

$$\hat{m} := \bar{X}$$

Dérivons par rapport à $\sigma^2$ (d'où la modification précédente de la log-vraisemblance).

$$\frac{\partial \log L}{\partial\sigma^2}(x_1, \ldots, x_n, m, \sigma^2) = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^4}$$

Donc $\dfrac{\partial \log L}{\partial \sigma^2}(x_1, \dots, x_n, m, \sigma^2) = 0$ donne

$$\frac{n}{2\sigma^2} = \sum_{i=1}^{n} \frac{(x_i - m)^2}{2\sigma^4}$$

d'où

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

Il faut naturellement remplacer $m$ par la solution trouvée en dérivant par rapport à $m$.
Vérifions la condition du second ordre.

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \sum_{i=1}^{n} \frac{(x_i - m)^2}{\sigma^6}$$

En $(\hat{m}, \hat{\sigma^2})$, on veut savoir si la condition suivante est vraie :

$$\frac{n}{2\hat{\sigma}^4} - \sum_{i=1}^{n} \frac{(x_i - \hat{m})^2}{\hat{\sigma}^6} < 0$$

qui se ramène à

$$\frac{\hat{\sigma}^6}{2\hat{\sigma}^4} < \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m})^2$$

Or $\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \hat{m})^2$ donc la condition se ramène à

$$\frac{1}{2}\hat{\sigma}^2 < \hat{\sigma}^2$$

qui est vraie car $\hat{\sigma}^2 > 0$.
En conclusion, $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S'^2$ est l'estimateur du maximum de vraisemblance de la variance.
Nous avons vu précédemment l'égalité $E(S'^2) = \dfrac{n-1}{n}\sigma^2$.
L'estimateur $\hat{\sigma}^2 = S'^2$ est biaisé tandis que l'estimateur $\hat{m} = \bar{X}$ est sans biais. *Lorsque nous estimerons une distribution gaussienne par la méthode du maximum de vraisemblance, l'espérance sera correctement estimée tandis que la variance sera systématiquement sous-estimée.* Naturellement, à mesure que la valeur de $n$ augmentera, le biais deviendra moins important.

Proposons un autre exemple classique.

**Exemple 9**
Déterminer l'estimateur du maximum de vraisemblance pour la loi géométrique de paramètre $p$.
La fonction de vraisemblance est :

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$$

i.e.

$$L(x_1, \ldots, x_n, p) = p^n (1-p)^{\left(\sum_{i=1}^{n} x_i\right) - n}$$

En passant au logarithme (ce qui est licite car les quantités considérées sont strictement positives), nous obtenons :

$$\log L(x_1, \ldots, x_n, p) = n \log(p) + \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \log(1-p)$$

puis en dérivant :

$$\frac{\partial \log L}{\partial p}(x_1, \ldots, x_n, p) = \frac{n}{p} - \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \frac{1}{1-p}$$

Ecrivons l'équation de log-vraisemblance.

$$\frac{\partial \log L}{\partial p}(x_1, \ldots, x_n, p) = 0 \Leftrightarrow \frac{n}{p} - \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \frac{1}{1-p} = 0$$

En multipliant par $p(1-p)$ puis, en simplifiant, nous obtenons :

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Vérifions que la condition suffisante est satisfaite.

$$\frac{\partial^2 \log L}{\partial p^2}(x_1, \ldots, x_n, p) = -\frac{n}{p^2} - \left[(\sum_{i=1}^{n} x_i) - n\right] \frac{1}{(1-p)^2}$$

En remarquant que[3], pour tout $i \in \{1, \ldots, n\}$, $x_i \geq 1$, nous déduisons : $\sum_{i=1}^n x_i - n \geq 0$ et $\dfrac{\partial^2 \log L}{\partial p^2}(x_1, \ldots, x_n, p) < 0$.
Par conséquent, $\hat{p} = \dfrac{n}{\sum_{i=1}^n X_i} = \dfrac{1}{\bar{X}}$ est l'estimateur du maximum de vraisemblance.
Remarquons qu'il s'agit de l'estimateur obtenu à l'aide de la méthode des moments d'ordre 1.

Nous conclurons par un théorème qui assure la convergence de l'estimateur du maximum de vraisemblance vers le paramètre considéré.

**Théorème 2**
Sous les conditions suivantes :

1. $\Theta$ ouvert de $\mathbb{R}$
2. $\mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ si $\theta \neq \theta'$
3. $f(x, \theta) > 0$ pour tout réel $x \in \mathcal{H}$ et tout réel $\theta \in \theta$
4. $\dfrac{\partial \log L}{\partial \theta}$ existe pour tout réel $x \in \mathcal{H}$ et tout réel $\theta \in \theta$

Alors il existe une suite de solutions $\hat{\theta}_n$ de l'équation de vraisemblance qui converge fortement vers la vraie valeur du paramètre $\theta_0$.

Ce théorème est démontré dans [5]. Nous y renvoyons le lecteur.

# 3 Information de Fisher

Nous souhaitons déterminer la quantité d'information contenue dans une observation ou dans un échantillon. En d'autres termes, nous souhaitons savoir en quoi la connaissance d'une ou de plusieurs observations nous renseignent sur la valeur d'un paramètre d'une loi. Sous les hypothèses suivantes

1. $\forall x \in \mathcal{H}$ et $\forall \theta \in \Theta \subset \mathbb{R}$, $f(x, \theta) > 0$,
2. $f$ est dérivable par rapport à $\theta$.

3. La loi géométrique ne charge que les valeurs entières strictement positives i.e. $\mathcal{H} = \mathbb{N} \setminus \{0\}$.

nous pourrons définir l'*information de Fisher*[4] :

$$I(\theta) := E\left(\frac{f'(X,\theta)}{f(X,\theta)}\right)$$

ou encore la définition équivalente

$$I(\theta) := E\left[\left(\frac{\partial \log f}{\partial \theta}(X,\theta)\right)^2\right]$$

**Définition 9**
Le *score* est défini par :

$$S(x,\theta) := \frac{\partial \log f}{\partial \theta}(x,\theta)$$

Pour utiliser cette formule, les calculs reposeront en général sur la linéarité de l'espérance, la connaissance des moments des lois et sur les identités remarquables suivantes :

- $(a+b)^2 = a^2 + 2ab + b^2$,
- $(a-b)^2 = a^2 - 2ab + b^2$,

**Exemple 10**
Pour la loi exponentielle, comme nous l'avons vu, $f(x,\lambda) = \lambda e^{-\lambda x} > 0$.
Donc

$$\log f(x,\lambda) = \log \lambda - \lambda x$$

et

$$\frac{\partial \log f}{\partial \lambda}(x,\lambda) = \frac{1}{\lambda} - x$$

Par conséquent,

$$I(\lambda) = E\left[\left(\frac{\partial \log f}{\partial \lambda}(X,\lambda)\right)^2\right] = E\left[\left(\frac{1}{\lambda} - X\right)^2\right]$$

D'où

$$I(\lambda) = \frac{1}{\lambda^2} - \frac{2}{\lambda}E(X) + E(X^2)$$

et

$$I(\lambda) = \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} = \boxed{\frac{1}{\lambda^2}}$$

4. Ce concept est dû comme le maximum de vraisemblance au biologiste et statisticien Ronald Aylmer Fisher.

**Exemple 11**
De même, nous calculerons $I(p)$ pour la loi binomiale de paramètres $n$ et $p$.
On sait que $f(x,p) = \binom{n}{x} p^x(1-p)^{n-x} > 0$. Donc

$$\log f(x,p) = \log \binom{n}{x} + x\log p + (n-x)\log(1-p)$$

et

$$\frac{\partial \log f}{\partial p}(x,p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$$

Par conséquent,

$$I(p) = E\left[\left(\frac{\partial \log f}{\partial p}(X,p)\right)^2\right] = \left[E\left(\frac{X-np}{p(1-p)}\right)^2\right]$$

donc

$$I(p) = \frac{1}{p^2(1-p)^2}\left[E(X^2) - 2npE(X) + n^2p^2\right]$$

Or $E(X) = np$ et $E(x^2) = n(n-1)p^2 + np$ d'après le chapitre 1.
Donc

$$I(p) = \frac{1}{p^2(1-p)^2}(n(n-1)p^2 + np - 2n^2p^2 + n^2p^2)$$

et

$$I(p) = \frac{np(1-p)}{p^2(1-p)^2} = \boxed{\frac{n}{p(1-p)}}$$

Nous privilégierons une autre méthode de calcul.

**Théorème 3**
Sous réserve que $f$ soit deux fois dérivable en $\theta$, $I(\theta) = -E\left(\frac{\partial^2 \log f}{\partial \theta^2}(X,\theta)\right)$.

**Exemple 12**
Revenons à la loi binomiale de paramètres $n$ et $p$.
On a : $\frac{\partial \log f}{\partial p}(x,p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$.
Par la dérivée d'un quotient [5] :

$$\frac{\partial^2 \log f}{\partial p^2}(x,p) = -\frac{-np(1-p) - (x-np)(1-2p)}{p^2(1-p)^2}$$

5. $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$

Après simplification, nous obtenons :

$$\frac{\partial^2 \log f}{\partial p^2}(x,p) = \frac{(2p-1)x - 2np^2}{p^2(1-p)^2}$$

Par conséquent

$$I(p) = -E\left(\frac{\partial^2 \log f}{\partial p^2}(x,p)\right) = -\frac{1}{p^2(1-p)^2}\left[(2p-1)E(X) - np^2\right]$$

Donc

$$I(p) = -\frac{1}{p^2(1-p)^2}(2p-1)np - np^2)$$

et

$$I(p) = \frac{np(1-p)}{p^2(1-p)^2} = \boxed{\frac{n}{p(1-p)}}$$

**Exemple 13**
Pour la loi normale de paramètre $m$ et $\sigma^2$ avec $\sigma^2$ fixé et connu.
On a : $f(x,m) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$.
Donc $\log f(x,m) = -\dfrac{1}{2}\log(2\pi\sigma^2) - \dfrac{(x-m)^2}{2\sigma^2}$
et $\dfrac{\partial \log f}{\partial m}(x,m) = -\dfrac{-2x+2m}{2\sigma^2} = \dfrac{-m+x}{\sigma^2}$.
D'où

$$\frac{\partial^2 \log f}{\partial m^2}(x,m) = \frac{-1}{\sigma^2}$$

En conclusion,

$$I(m) = -E\left(\frac{\partial^2 \log f}{\partial m^2}(X,m)\right) = \frac{1}{\sigma^2}$$

**Exemple 14**
Pour la loi exponentielle de paramètre $\lambda$, nous pouvons la calculer de nouveau.
On a : $f(x,\lambda) = \lambda e^{-\lambda x}$.
Donc $\log f(x,\lambda) = \log\lambda - \lambda x$.
D'où $\dfrac{\partial \log f}{\partial \lambda}(x,\lambda) = \dfrac{1}{\lambda} - x$.
Par conséquent,

$$\frac{\partial^2 \log f}{\partial \lambda^2}(x,\lambda) = -\frac{1}{\lambda^2}$$

En conclusion,

$$I(\lambda) = -E\left(\frac{\partial^2 \log f}{\partial \lambda^2}(X, \lambda)\right) = \frac{1}{\lambda^2}$$

Voici les informations usuelles :

- loi de Poisson $I(\lambda) = \dfrac{1}{\lambda}$
- loi binomiale $I(p) = \dfrac{n}{p(1-p)}$
- loi normale $I(m) = \dfrac{1}{\sigma^2}$
- loi normale $I(\sigma^2) = \dfrac{1}{2\sigma^4}$
- loi exponentielle $I(\lambda) = \dfrac{1}{\lambda^2}$.

Lorsque des variables aléatoires $X_1$,...,$X_n$ sont **indépendantes**, l'information de Fisher associée au n-uplet $(X_1, \ldots, X_n)$ est :

$$\boxed{I_n(\theta) := I_{(X_1,\ldots,X_n)}(\theta) = nI(\theta)}$$

où $I(\theta) := I_{X_1}(\theta)$.

**Exemple 15**
Par exemple, pour la loi normale,

$$I_n(m) = \frac{n}{\sigma^2}$$

# 4 Inégalité FDCR

Nous étudierons dans cette partie l'inégalité de Fréchet-Darmois-Cramer-Rao.
Comme nous l'avons vu, pour éviter la sous-estimation ou la sur-estimation des paramètres, nous privilégierons des estimateurs sans biais.
De plus, nous souhaitons déterminer parmi eux ceux dont la variance sera la plus faible car la variance correspondra à la distance (dans l'espace fonctionnel des variables aléatoires de carré intégrable) entre l'estimateur et $\theta$ (vu

comme une fonction constante de carré intégrable).
L'inégalité FDCR fournira une borne inférieure à la variance d'un estimateur sans biais.
Nous notons, à titre culturel, les hypothèses de Cramer-Rao. Elles seront toujours vérifiées dans les exercices que nous traiterons.
**Hypothèses de Cramer-Rao** [6]

- $H_1$ : $\Theta$ est un ouvert de $\mathbb{R}$ et $f(x,\theta) > 0$ pour tout $x$ et tout $\theta$.
- $H_2$ : $\dfrac{\partial f}{\partial \theta}(x,\theta)$ et $\dfrac{\partial^2 f}{\partial \theta^2}(x,\theta)$ existent et sont finies presque sûrement et pour tout $\theta$.
- $H_3$ : Pour tout $A$ (mesurable), $\int_A f(x,\theta)dx$ est dérivable au moins deux fois sous le signe $\int$ et
$$\frac{\partial}{\partial \theta}\int_A f(x,\theta)dx = \int_A \frac{\partial f}{\partial \theta}(x,\theta)dx$$
- $H_4$ : pour tout $\theta$, $0 < I(\theta) < +\infty$.

Les hypothèses sont d'ordre technique et assurent notamment la licéité de la dérivation sous le signe $\int$.

**Théorème 4**
Sous les hypothèses de Cramer-Rao. Soit $T$ un estimateur sans biais de $g(\theta)$ de variance finie.
Sous l'hypothèse supplémentaire :
$H5$ : Pour tout $A$ (mesurable), $\int_A T(x)f(x,\theta)dx$ est dérivable par rapport à $\theta$ et $\int_A |T(x)|\dfrac{\partial f}{\partial \theta}(x,\theta)dx$.
On a :

- $g(\theta)$ est dérivable.
- pour tout $\theta$,
$$\boxed{V(T) \geq \frac{g'^2(\theta)}{I_n(\theta)}}$$

6. Nous renvoyons le lecteur intéressé à [5].

La borne $\dfrac{g'^2(\theta)}{I_n(\theta)}$ est appelée *borne de Fréchet* ou borne de *Cramer-Rao*.
Lorsque nous étudierons des estimateurs sans biais du paramètre $\theta$, $g(\theta) = \theta$ et donc la borne de Fréchet sera $\dfrac{1}{I_n(\theta)}$.

**Définition 10**
Un estimateur sans biais $T$ de $g(\theta)$ est dit *efficace* lorsque la borne de Fréchet est atteinte i.e.

$$V(T) = \frac{g'^2(\theta)}{I_n(\theta)}$$

**Exemple 16**
Avant d'étudier la moyenne empirique, rappelons deux propriétés importantes sur la variance.
soient $X$ et $Y$ deux variables aléatoires indépendantes et $a$ un réel.
On a :

- $V(aX) = a^2V(X)$
- $V(X+Y) = V(X) + V(Y)$

La preuve de ces propriétés est laissée en exercice. La première propriété procède de la définition de la variance et la seconde procède du fait que $E(XY) = E(X)E(Y)$ pour des variables aléatoires $X$ et $Y$ indépendantes.
Nous savons que la moyenne empirique $\bar{X}$ est un estimateur sans biais de l'espérance pour toutes les lois (qui admettent une espérance).
Pour la loi normale, $V(\bar{X}) = V\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} V(X_i)$ car les variables aléatoires $X_i$ sont indépendantes.
$V(\bar{X}) = \dfrac{n}{n^2}V(X_1) = \dfrac{\sigma^2}{n}$ car les variables aléatoires $X_i$ sont de même loi.
D'après ce qui précède, $I_n(m) = \dfrac{n}{\sigma^2}$.
Donc

$$V(\bar{X}) = \frac{1}{I_n(m)}$$

La moyenne empirique est donc un estimateur efficace du paramètre $m$.

Nous verrons dans les compléments et en exercices comment traiter les estimateurs biaisés et nous étudierons d'autres méthodes d'estimation.

# Références

[1] Dacunha-Castelle, D. : *Chemins de l'aléatoire*. Flammarion (2000)

[2] Foata, D., Franchi, J. Fuchs, A. : *Calcul de probabilités*. Dunod (2012)

[3] Monfort, A. : *Cours de statistique mathématique*. Economica (1997)

[4] Revuz, D. : *Probabilités*. Hermann (1997)

[5] Tassi, P. : *Méthodes statistiques*. Economica (2004)

[6] Toulouse, P.S. : *Thèmes de probabilités et de statistiques*. Dunod (1999)

[7] Williams, D. : *Probability with Martingales*. Cambrige Mathematical Textbooks (1991)