L3 Economie

# Statistiques inférentielles

## Chapitre 3 : Estimation ponctuelle

Cécile Durot
cecile.durot@gmail.com

Université Paris Nanterre

Introduction

Méthode de substitution

Estimation non paramétrique : Estimation empirique

Estimation paramétrique : EMV

## Cadre d'étude

Dans toute la suite, nous supposerons que les données collectées sont des nombres réels $x_1, \ldots, x_n$, et que le procédé de collecte est suffisament proche du tirage aléatoire simple pour que l'on puisse supposer que :

Les données $x_1, \ldots, x_n$ sont les réalisations de v.a.r. $X_1, \ldots, X_n$ indépendantes et de même loi (i.i.d.).

On notera respectivement $P_X$ et $F_X$ la loi et la fonction de répartition commune des v.a.r. $X_1, \ldots, X_n$ tandis que $P_n$ et $F_n$ désigneront respectivement la loi empirique et la fonction de répartition empirique. On notera également $X$ une variable aléatoire abstraite de même loi que $X_1, \ldots, X_n$ et $\Omega$ l'univers sur lequel sont définies $X_1, \ldots, X_n$.

## Problématique

On s'intéresse à une caractéristique de la loi $P_X$. La caractéristique est inconnue puisque la loi est inconnue. Pour simplifier, on considère principalement ici le cas d'une caractéristique à valeurs réelles ; on l'appellera paramètre d'intérêt.

**Exemples :**

- Le paramètre d'intérêt peut être par exemple $P(X \leqslant 0)$, $P(X \leqslant 2)$, ou plus généralement $F_X(t)$ pour un $t \in \mathbb{R}$ donné.
- Si le modèle statistique suppose que $X_1, \ldots, X_n$ possèdent une espérance, le paramètre d'intérêt peut être $E(X)$. S'il suppose que ces variables possèdent une variance, le paramètre d'intérêt peut être $var(X)$.
- Si le modèle statistique suppose que $X_1, \ldots, X_n$ possèdent une densité $f_X$, le paramètre d'intérêt peut être $f_X(t)$ pour un $t \in \mathbb{R}$ donné.

# Estimateur

Dans la suite, on note $\theta_0$ la vraie valeur du paramètre d'intérêt et $\Theta$ l'ensemble des valeurs possibles pour $\theta_0$. Par exemple, $\Theta = ]0, +\infty[$ si $\theta_0 = var(X)$ tandis que $\Theta = \mathbb{R}$ si $\theta_0 = E(X)$.

**Définition :** Un estimateur est une variable aléatoire souvent notée $\widehat{\theta}_n$, à valeurs dans $\Theta$, telle que pour une certaine fontion $T_n$, on a

$$\widehat{\theta}_n(\omega) = T_n(X_1(\omega), \ldots, X_n(\omega)) \text{ pour tout } \omega \in \Omega.$$

**Remarques :**

- Donner un estimateur $\widehat{\theta}_n$ est donc équivalent à donner la fonction $T_n$.
- Un estimateur ne dépend du résultat $\omega$ de l'expérience aléatoire qu'au travers des réalisations de $X_1, \ldots, X_n$, c'est-à-dire des données.
- Il existe toujours plusieurs (une infinité d') estimateurs de $\theta_0$.

# Qualités d'un estimateur

**Définitions :** Soit $\widehat{\theta}_n$ un estimateur de $\theta_0$.

- Le biais de $\widehat{\theta}_n$ est le nombre $E(\widehat{\theta}_n) - \theta_0$.
- On dit que $\widehat{\theta}_n$ est sans biais si $E(\widehat{\theta}_n) = \theta_0$.
- On dit que $\widehat{\theta}_n$ est asymptotiquement sans biais si $\lim_{n\to\infty} E(\widehat{\theta}_n) = \theta_0$.
- On dit que $\widehat{\theta}_n$ est consistant si $\widehat{\theta}_n \xrightarrow[n\to\infty]{P} \theta_0$.
- Le risque quadratique de $\widehat{\theta}_n$ est $E[(\widehat{\theta}_n - \theta_0)^2]$.

**Qualités requises :**
On choisira toujours un estimateur sans biais ou asymptotiquement sans biais, consistant, et de risque quadratique le plus petit possible. Ainsi, entre deux estimateurs sans biais, on retiendra celui dont la variance est la plus petite.

# Estimation

**Définition :** La réalisation d'un estimateur, $\widehat{\theta}_n(\omega) = T_n(x_1, \ldots, x_n)$, est appelée estimation.

**Remarque :**
L'estimation est la valeur que l'on va pouvoir calculer à partir des données. C'est elle qui fournit la valeur approchée de $\theta_0$.

Introduction

Méthode de substitution

Estimation non paramétrique : Estimation empirique

Estimation paramétrique : EMV

## Méthode de substitution

Il découle du théorème de Slutsky que si l'on dispose d'un estimateur consistant $\widehat{\theta}_n$ de $\theta_0$, alors pour toute fonction continue $g$, $g(\widehat{\theta}_n)$ est un estimateur consistant de $g(\theta_0)$.

Par exemple, il découle de la LGN que $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur consistant de $E(X)$, donc $(\overline{X}_n)^2$ est un estimateur consistant de $(E(X))^2$.

Introduction

Méthode de substitution

Estimation non paramétrique : Estimation empirique

Estimation paramétrique : EMV

## Estimation empirique

La méthode d'estimation empirique repose sur le fait que la loi empirique $P_n$ fournit une bonne approximation de la vraie loi $P_X$ lorsque $n$ est grand.

**Définition :** L'estimateur empirique d'une caractéristique de $P_X$ est la même caractéristique de $P_n$.

**Exemples :**

- Si $\theta_0 = E(X)$ alors l'estimateur empirique est l'espérance d'une variable aléatoire de loi $P_n$, c'est-à-dire $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- Si le paramètre d'intérêt est $var(X)$ alors l'estimateur empirique est $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
- L'estimateur empirique de $F_X(t)$ est $F_n(t)$ pour tout $t \in \mathbb{R}$.
- Pour toute fonction $g$ telle que cela a un sens, l'estimateur empirique de $E(g(X))$ est $\frac{1}{n}\sum_{i=1}^n g(X_i)$.
- Il n'existe pas d'estimateur empirique de la densité de $X$.

# Propriétés des estimateurs empiriques

On peut utiliser la méthode d'estimation empirique dans un modèle non paramétrique mais aussi dans un modèle paramétrique. Cette méthode fournit le plus souvent des estimateurs consistants et asymptotiquement sans biais (voire sans biais).

**Exemples :**

- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais et consistant (LGN) de $E(X)$.
- $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ est un estimateur asymptotiquement sans biais et consistant (exercice) de $var(X)$. Un estimateur sans biais et consistant est l'estimateur empirique corrigé
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$
- $F_n(t)$ et $\frac{1}{n}\sum_{i=1}^n g(X_i)$ sont des estimateurs sans biais et consistants de $F_X(t)$ et $E(g(X))$, respectivement.

Introduction

Méthode de substitution

Estimation non paramétrique : Estimation empirique

Estimation paramétrique : EMV

## Cadre

Dans la suite, on suppose qu'on dispose d'une famille de lois indexée par $\theta$, $\{P_\theta, \theta \in \Theta\}$ avec $\Theta \subset \mathbb{R}^p$ donné, à laquelle appartient la loi de $X$ : il existe donc $\theta_0$ tel que $P_X = P_{\theta_0}$. On suppose dans la suite $\theta_0$ identifiable au sens où $P_\theta \neq P_{\theta_0}$ dès lors que $\theta \neq \theta_0$. Connaître $P_X$ revient alors à connaître $\theta_0$. On étudiera simultanément les deux cadres suivants :

1. La loi de $X$ est discrète et on note $f_\theta(x) = P_\theta(\{x\})$ la probabilité de $\{X = x\}$ si $\theta_0 = \theta$, pour tout $x \in \Omega_X$.
2. Chaque $P_\theta$ possède une densité notée $f_\theta$.

## Vraisemblance

**Définitions :**

- La vraisemblance est la fonction (aléatoire) $L_n : \Theta \to [0, +\infty[$ telle que pour tout $\theta \in \Theta$,

$$L_n(\theta) = f_\theta(X_1) \times \cdots \times f_\theta(X_n).$$

- Si $f_\theta(x) > 0$ pour tout $x \in \Omega_X$ et tout $\theta \in \Theta$, alors la log-vraisemblance est la fonction (aléatoire) $\ell_n : \Theta \to \mathbb{R}$ telle que pour tout $\theta \in \Theta$,

$$\ell_n(\theta) = \log(L_n(\theta)) = \log(f_\theta(X_1)) + \cdots + \log(f_\theta(X_n)).$$

# Exemple 1

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0, 1[$. Avec les notations de la diapositive 14, $\theta_0 = p_0$, $\Theta = ]0, 1[$ et pour tout $p \in \Theta$, $P_p$ est la loi $\mathcal{B}(p)$. Pour tout $p \in ]0, 1[$, on a

$$f_p(x) = p^x(1-p)^{1-x} \text{ pour tout } x \in \{0, 1\},$$

donc

$$\log(f_p(X_i)) = X_i \log(p) + (1 - X_i)\log(1-p).$$

On en déduit que pour tout $p \in ]0, 1[$, on a

$$\ell_n(p) = \sum_{i=1}^{n} \log(f_p(X_i)) = S_n \log p + (n - S_n)\log(1-p)$$

où $S_n = X_1 + \cdots + X_n$.

# Exemple 1

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0,1[$. Avec les notations de la diapositive 14, $\theta_0 = p_0$, $\Theta = ]0,1[$ et pour tout $p \in \Theta$, $P_p$ est la loi $\mathcal{B}(p)$. Pour tout $p \in ]0,1[$, on a

$$f_p(x) = p^x(1-p)^{1-x} \text{ pour tout } x \in \{0,1\},$$

donc

$$\log(f_p(X_i)) = X_i \log(p) + (1 - X_i)\log(1-p).$$

On en déduit que pour tout $p \in ]0,1[$, on a

$$\ell_n(p) = \sum_{i=1}^{n} \log(f_p(X_i)) = S_n \log p + (n - S_n)\log(1-p)$$

où $S_n = X_1 + \cdots + X_n$.

# Exemple 1

Supposons que $X_1, \dots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0,1[$. Avec les notations de la diapositive 14, $\theta_0 = p_0$, $\Theta = ]0,1[$ et pour tout $p \in \Theta$, $P_p$ est la loi $\mathcal{B}(p)$. Pour tout $p \in ]0,1[$, on a

$$f_p(x) = p^x(1-p)^{1-x} \text{ pour tout } x \in \{0,1\},$$

donc

$$\log(f_p(X_i)) = X_i \log(p) + (1-X_i)\log(1-p).$$

On en déduit que pour tout $p \in ]0,1[$, on a

$$\ell_n(p) = \sum_{i=1}^{n} \log(f_p(X_i)) = S_n \log p + (n-S_n)\log(1-p)$$

où $S_n = X_1 + \cdots + X_n$.

# Exemple 2

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$.

Pour tout $\theta > 0$ on a

$$f_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}}.$$

De plus,

$$P(X_i \leqslant 0) = \int_{-\infty}^{0} f_{\theta_0}(x)dx = 0$$

donc $X_i > 0$ p.s. On a donc $f_\theta(X_i) = \theta \exp(-\theta X_i)$ p.s. et

$$\log(f_\theta(X_i)) = \log\theta - \theta X_i \ p.s.$$

Donc,

$$\ell_n(\theta) = \sum_{i=1}^{n} \log(f_\theta(X_i)) = n\log\theta - \theta S_n \text{ p.s. pour tout } \theta > 0$$

où $S_n = X_1 + \cdots + X_n$.

# Exemple 2

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$.

Pour tout $\theta > 0$ on a

$$f_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}}.$$

De plus,

$$P(X_i \leqslant 0) = \int_{-\infty}^{0} f_{\theta_0}(x)dx = 0$$

donc $X_i > 0$ p.s. On a donc $f_\theta(X_i) = \theta \exp(-\theta X_i)$ p.s. et

$$\log(f_\theta(X_i)) = \log \theta - \theta X_i \ p.s.$$

Donc,

$$\ell_n(\theta) = \sum_{i=1}^{n} \log(f_\theta(X_i)) = n \log \theta - \theta S_n \text{ p.s. pour tout } \theta > 0$$

où $S_n = X_1 + \cdots + X_n$.

# Exemple 2

Supposons que $X_1, \dots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$.

Pour tout $\theta > 0$ on a

$$f_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}}.$$

De plus,

$$P(X_i \leqslant 0) = \int_{-\infty}^{0} f_{\theta_0}(x) dx = 0$$

donc $X_i > 0$ p.s. On a donc $f_\theta(X_i) = \theta \exp(-\theta X_i)$ p.s. et

$$\log(f_\theta(X_i)) = \log \theta - \theta X_i \ p.s.$$

Donc,

$$\ell_n(\theta) = \sum_{i=1}^{n} \log(f_\theta(X_i)) = n \log \theta - \theta S_n \text{ p.s. pour tout } \theta > 0$$

où $S_n = X_1 + \cdots + X_n$.

## Exemple 2

Supposons que $X_1, \dots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$.

Pour tout $\theta > 0$ on a

$$f_\theta(x) = \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}}.$$

De plus,

$$P(X_i \leqslant 0) = \int_{-\infty}^{0} f_{\theta_0}(x)dx = 0$$

donc $X_i > 0$ p.s. On a donc $f_\theta(X_i) = \theta \exp(-\theta X_i)$ p.s. et

$$\log(f_\theta(X_i)) = \log \theta - \theta X_i \ p.s.$$

Donc,

$$\ell_n(\theta) = \sum_{i=1}^{n} \log(f_\theta(X_i)) = n \log \theta - \theta S_n \text{ p.s. pour tout } \theta > 0$$

où $S_n = X_1 + \dots + X_n$.

# Estimateur du maximum de vraisemblance (EMV)

**Définition :** On appelle estimateur du maximum de vraisemblance toute solution $\widehat{\theta}_n$, si elle existe, de l'équation

$$L_n(\widehat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta).$$

Si $f_\theta(x) > 0$ pour tout $x \in \Omega_X$ et tout $\theta \in \Theta$, alors de façon équivalente, $\widehat{\theta}_n$ est solution, si elle existe, de l'équation

$$\ell_n(\widehat{\theta}_n) = \sup_{\theta \in \Theta} \ell_n(\theta).$$

Cela vient du fait que la fonction log est strictement croissante et continue. L'intérêt de cette écriture est qu'elle fait intervenir la log-vraisemblance, qui étant définie comme une somme de fonctions, est plus facile à dériver que la vraisemblance $L_n$.

# Exemple 1

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0,1[$. Rappelons qu'alors,

$$\ell_n(p) = S_n \log p + (n - S_n) \log(1 - p)$$

pour tout $p \in ]0,1[$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(p) = \frac{S_n}{p} - \frac{n - S_n}{1 - p} = \frac{S_n(1-p) - p(n - S_n)}{p(1-p)} = \frac{S_n - np}{p(1-p)}$$

On en déduit le tableau de signes :

| $p$ | $0$ | | $S_n/n$ | | $1$ |
|---|---|---|---|---|---|
| $\ell_n'$ | | $+$ | $0$ | $-$ | |
| $\ell_n$ | | $\nearrow$ | | $\searrow$ | |

Donc l'EMV de $p_0$ est $S_n/n$.

# Exemple 1

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0,1[$.
Rappelons qu'alors,

$$\ell_n(p) = S_n \log p + (n - S_n) \log(1-p)$$

pour tout $p \in ]0,1[$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(p) = \frac{S_n}{p} - \frac{n - S_n}{1-p} = \frac{S_n(1-p) - p(n-S_n)}{p(1-p)} = \frac{S_n - np}{p(1-p)}$$

On en déduit le tableau de signes :

| $p$ | $0$ | | $S_n/n$ | | $1$ |
|---|---|---|---|---|---|
| $\ell_n'$ | | $+$ | $0$ | $-$ | |
| $\ell_n$ | | $\nearrow$ | | $\searrow$ | |

Donc l'EMV de $p_0$ est $S_n/n$.

# Exemple 1

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(p_0)$ avec $p_0 \in ]0,1[$. Rappelons qu'alors,

$$\ell_n(p) = S_n \log p + (n - S_n) \log(1-p)$$

pour tout $p \in ]0,1[$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(p) = \frac{S_n}{p} - \frac{n - S_n}{1-p} = \frac{S_n(1-p) - p(n-S_n)}{p(1-p)} = \frac{S_n - np}{p(1-p)}$$

On en déduit le tableau de signes :

| $p$ | $0$ | | $S_n/n$ | | $1$ |
|---|---|---|---|---|---|
| $\ell_n'$ | | $+$ | $0$ | $-$ | |
| $\ell_n$ | | $\nearrow$ | | $\searrow$ | |

Donc l'EMV de $p_0$ est $S_n/n$.

# Exemple 2

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$. Rappelons qu'alors,

$$\ell_n(\theta) = n \log \theta - \theta S_n \text{ p.s.}$$

pour tout $\theta > 0$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(\theta) = \frac{n}{\theta} - S_n$$

On en déduit le tableau de signes :

| $\theta$ | $0$ | | $n/S_n$ | | $+\infty$ |
|---|---|---|---|---|---|
| $\ell_n'$ | | $+$ | $0$ | $-$ | |
| $\ell_n$ | | $\nearrow$ | | $\searrow$ | |

Donc l'EMV de $\theta_0$ est $n/S_n$ c'est-à-dire $1/\bar{X}_n$, où $\bar{X}_n = \sum_{i=1}^n X_i/n$.

# Exemple 2

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$. Rappelons qu'alors,

$$\ell_n(\theta) = n \log \theta - \theta S_n \text{ p.s.}$$

pour tout $\theta > 0$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(\theta) = \frac{n}{\theta} - S_n$$

On en déduit le tableau de signes :

| $\theta$ | $0$ | $n/S_n$ | $+\infty$ |
|---|---|---|---|
| $\ell_n'$ | $+$ | $0$ | $-$ |
| $\ell_n$ | $\nearrow$ | | $\searrow$ |

Donc l'EMV de $\theta_0$ est $n/S_n$ c'est-à-dire $1/\bar{X}_n$, où $\bar{X}_n = \sum_{i=1}^n X_i/n$.

# Exemple 2

Supposons que $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{E}(\theta_0)$ avec $\theta_0 > 0$. Rappelons qu'alors,

$$\ell_n(\theta) = n \log \theta - \theta S_n \text{ p.s.}$$

pour tout $\theta > 0$ où $S_n = X_1 + \cdots + X_n$. Pour déterminer le lieu de maximum de la fonction $\ell_n$, on la dérive :

$$\ell_n'(\theta) = \frac{n}{\theta} - S_n$$

On en déduit le tableau de signes :

| $\theta$ | $0$ | | $n/S_n$ | | $+\infty$ |
|---|---|---|---|---|---|
| $\ell_n'$ | | $+$ | $0$ | $-$ | |
| $\ell_n$ | | $\nearrow$ | | $\searrow$ | |

Donc l'EMV de $\theta_0$ est $n/S_n$ c'est-à-dire $1/\bar{X}_n$, où $\bar{X}_n = \sum_{i=1}^n X_i/n$.

# Remarques

- L'EMV n'existe pas toujours. S'il existe, il n'est pas nécessairement unique, et pas nécessairement explicite.
- Dans le cas d'un paramètre réel $\Theta \subset \mathbb{R}$, si $f_\theta(x) > 0$ pour tout $x \in \Omega_X$ et tout $\theta \in \Theta$, et si $\theta \mapsto f_\theta(x)$ est dérivable pour tout $x$, alors pour calculer l'EMV :
  - ▷ on calcule $L_n$ puis $\ell_n$
  - ▷ on calcule la dérivée de $\ell_n$
  - ▷ on dresse le tableau de variations de $\ell_n$
  - ▷ on en déduit la valeur de l'EMV (le point, s'il existe), où cette fonction est maximale.

## Familles exponentielles

**Définition :** Dans le cas d'un paramètre réel avec $\Theta$ un intervalle ouvert de $\mathbb{R}$, on dit que la famille de lois $\{P_\theta, \theta \in \Theta\}$ est exponentielle si $f_\theta$ peut s'écrire sous la forme

$$f_\theta(x) = h(x)C(\theta)\exp(G(\theta)T(x))$$

pour tout $\theta \in \Theta$ et tout $x \in \Omega_X$, pour certaines fonctions $h$, $C$, $G$, $T$ avec $G$ indéfiniment dérivable telle que $G'(\theta) > 0$ pour tout $\theta$.

**Exemples :** Les familles de lois suivantes sont exponentielles :

- $\{\mathcal{B}(\theta), \theta \in ]0,1[\}$
- $\{\mathcal{E}(\theta), \theta > 0\}$
- $\{\mathcal{P}(\theta), \theta > 0\}$
- $\{\mathcal{N}(m, \sigma^2), m \in \mathbb{R}\}$, avec $\sigma > 0$ donné

# Exemple 1

$$
\begin{aligned}
f_p(x) &= p^x(1-p)^{1-x} \\
&= (1-p)\left(\frac{p}{1-p}\right)^x \\
&= (1-p)\exp\left(x\log\left(\frac{p}{1-p}\right)\right)
\end{aligned}
$$

Donc

$$f_p(x) = h(x)C(p)\exp(G(p)T(x))$$

si on pose (par exemple) $h(x) = 1$, $C(p) = 1-p$, $T(x) = x$, $G(p) = \log(p/(1-p))$.

# Exemple 2

$$
\begin{aligned}
f_\theta(x) &= \theta \exp(-\theta x) \mathbb{1}_{\{x>0\}} \\
&= h(x) C(\theta) \exp(G(\theta) T(x))
\end{aligned}
$$

si on pose (par exemple) $h(x) = \mathbb{1}_{\{x>0\}}$, $C(\theta) = \theta$, $T(x) = -x$, $G(\theta) = \theta$.

**Définitions :** Si $\{P_\theta, \theta \in \Theta\}$ est une famille exponentielle de lois, alors

- Le score est $\dot{\ell}_{\theta_0}(X) = \frac{\partial}{\partial \theta_0} \log f_{\theta_0}(X)$
- L'information de Fisher est $I(\theta_0) = E[(\dot{\ell}_{\theta_0}(X))^2]$.

# Exemple 1

On rappelle (voir la diapositive 16) que

$$\log(f_p(X)) = X\log(p) + (1-X)\log(1-p).$$

donc le score est

$$\begin{aligned}\frac{\partial}{\partial p_0}\log f_{p_0}(X) &= \frac{X}{p_0} - \frac{1-X}{1-p_0}\\ &= \frac{X-p_0}{p_0(1-p_0)}\end{aligned}$$

et l'information de Fisher est

$$\begin{aligned}I(\theta_0) &= \frac{E[(X-p_0)^2]}{(p_0(1-p_0))^2}\\ &= \frac{var(X)}{(p_0(1-p_0))^2}\\ &= \frac{1}{p_0(1-p_0)}\end{aligned}$$

**Théorème :** Si $\{P_\theta, \theta \in \Theta\}$ est une famille exponentielle de lois, alors

- $E(\dot{\ell}_{\theta_0}(X)) = 0$ et $I(\theta_0) = var(\dot{\ell}_{\theta_0}(X))$.
- $I(\theta_0) = -E[\frac{\partial^2}{\partial^2\theta_0} \log f_{\theta_0}(X)]$.
- Si l'EMV $\widehat{\theta}_n$ de $\theta_0$ existe et $I(\theta_0) > 0$, alors
$$\sqrt{nI(\theta_0)}(\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)\,.$$
- Si l'EMV $\widehat{\theta}_n$ de $\theta_0$ existe, $I(\theta_0) > 0$, et $\theta \mapsto I(\theta)$ est une fonction continue, alors $\sqrt{nI(\widehat{\theta}_n)}(\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)\,.$

Par exemple, si $X_1, \ldots, X_n$ sont i.i.d. $\mathcal{B}(\theta_0)$ pour un $\theta_0 \in ]0,1[$, alors
$$\sqrt{n}\frac{\overline{X}_n - \theta_0}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)\,.$$

**Remarque :** Sous les hypothèses précédentes, l'EMV est consistant.

**Théorème (Borne de Cramer-Rao) :** Si $\{P_\theta, \theta \in \Theta\}$ est une famille exponentielle de lois, alors tout estimateur sans biais $\widehat{\theta}_n$ de $\theta_0$ vérifie

$$var(\widehat{\theta}_n) \geqslant \frac{1}{nI(\theta_0)}.$$

**Définition :** Dans le cadre des familles exponentielles, un estimateur sans biais $\widehat{\theta}_n$ est dit efficace si

$$var(\widehat{\theta}_n) = \frac{1}{nI(\theta_0)}.$$

**Exemple 1 :** L'EMV est efficace dans le modèle de Bernoulli, puisque l'EMV est $\bar{X}_n$, de variance $p_0(1-p_0)/n = 1/(nI(p_0))$.