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

L'Université Djillali Liabes Sidi Bel Abbès, Algérie
Faculté des Sciences Exactes
Département de Mathématiques

# Polycopié

**Sur**

# Théorie de l'estimation statistique
# Cours et Exercices

**Présentée par**

Zoulikha Kaid

Année Universitaire 2014-2015

# Table des matières

## Préface

L'objectif de la statistique mathématique est la détermination d'une caractéristique de la population à partir d'un échantillon. Elle est généralement basée sur deux procédures : l'échantillonnage et l'estimation. Dans cette polycopie nous allons concentrer sur la deuxième procédure. Plus précisément nous nous intéressons à la théorie de l'estimation statistique. Ces cours sont destinés aux étudiants de la première année Master Mathématiques Statistique Appliqués à Economie et Finance. Cette polycopie contient cinq chapitres dont on traite les deux types des modèles (modèles paramétrique et les modèles non paramétrique).

Dans le premier chapitre on donne les notions fondamentales de l'estimation statistique, telles la structure statistique, l'information statistique, les différents propriétés nécessaires pour sélectionner le meilleur et les méthodes d'estimation.

Nous traitons l'estimation fonctionnelle dans le deuxième chapitre. Nous monterons dans ce chapitre que l'estimation de la fonction de répartition est un outil indispensable pour l'estimation de n'import paramètre fonctionnel. Ainsi, nous étudions les propriétés asymptotiques de la fonction de répartition empirique ainsi que l'estimateur à noyau de la fonction de répartition. Comme application de cette estimation nous présentons l'estimateur des quartiles et nous déduisons ces propriétés asymptotiques.

Dans le chapitre trois on va étudier l'estimation de la densité. Ce modèle non paramétrique est présenté en six sections. Nous construisons l'estimateur à noyau de la densité dans la première Section, puis nous étudions le biais et la variance, respectivement dans les Sections deux et trois. La quatrième section est consacrée à la convergence en probabilité. On trouvera dans les derniers section ( cinq et six) quelques applications telle l'estimation du mode et le choix du paramètre de lissage.

Le problème de prévision est l'un des applications importantes de l''estimation non paramétrique. Elle est modélisée comme une relation de régression entre deux variables aléatoires. Ce modèle non paramétrique est considéré dans le quatrième chapitre. Ce chapitre contient la définition de l'estimateur et les propriétés asymptotiques de cet estimateur telle la convergence presque complète et la convergence en moyenne quadratique. On donne dans le dernier chapitre une séries d'exercices ainsi quelques projets de travaux pratique.

Dédicace

*À tous ceux qui me sont proches.*

# Chapitre 1

# Estimation paramétrique

Ce chapitre est divisé en quatre sections. Dans la première section, on présente la définition de la structure statistique. La notion d'estimateur statistique ainsi que les critères du choix sont données dans la deuxième section. La troisième section est consacrée à l'information au sens de Fisher. Nous citons dans la dernière section les méthodes usuelle d'estimation.

## 1.1 Structure statistique

**Définition 1** *Soit* $(\Omega, \mathcal{A})$ *un espace probabilisable et soit*

$$\mathcal{P} = \left\{P_\theta; \quad \theta \in \Theta \qquad \Theta \subset \mathbb{R}^k\right\}$$

*une famille des lois de probabilités. Le triplet* $(\Omega, \mathcal{A}, \mathcal{P})$ *est appelé structure statistique.*

**Exemple 1**

La structure gaussienne est $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, f_{m,\sigma})$ avec

$$f_{m,\sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}.$$

Ici $\mathcal{P}$ est la famille des lois définies par la densité $f_{m,\sigma}$, $\theta = (m, \sigma)$, $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ et $k = 2$.

**Définition 2** *Soit* $X$ *une variable aléatoire sur* $(\Omega, \mathcal{A}, \mathcal{P})$*. Une suite des variables aléatoires* $X_1, \ldots X_n$ *est appelée échantillon de* $X$ *s'elles sont deux à deux indépendantes et elles ont la même loi de probabilité que* $X$*.*

*Remarque 1*

Si la variable $X$ admet une densité $f(\theta, x)$, alors la fonction

$$L(\theta, x_1, x_2, \dots x_n) = \prod_{i=1}^{n} f(\theta, x_i)$$

est une densité pour le vecteur aléatoire $(X_1, \dots X_n)$. Cette fonction est appelée vraisemblance de $X_1, \dots X_n$.

**Définition 3** *Soit $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$. Toute fonction mesurable $T$ $X_1, \dots X_n$ est appelée statistique. Par la suite cette application mesurable est notée $T(X_1, \dots X_n)$.*

**Définition 4** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ une statistique sur cet échantillon. On dit que la statistique $T(X_1, \dots X_n)$ est exhaustive si $P_\theta(A/T(X_1, \dots X_n))$ ne dépend pas de $\theta$.*

*Remarque 2*

Si la variable $X$ et la statistique $T(X_1, X_2, \dots X_n)$, ont respectivement des densité $f(\theta, x)$ et $g(\theta, x)$ alors la statistique est exhaustive si et seulement si

$$\frac{L(\theta, x_1, x_2, \dots x_n)}{g(\theta, T(x_1, x_2, \dots x_n)} = h(x_1, x_2, \dots x_n) \quad \text{ne dépend pas de } \theta.$$

## 1.2 Estimateur statistique

**Définition 5** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$. Un estimateur de $\theta$ est une statistique $T(X_1, \dots X_n)$ qui prend ces valeurs dans $\Theta$.*

**Définition 6** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ un estimateur $\theta$. On dit que l'estimateur $T(X_1, \dots X_n)$ est sans biais si et seulement si*

$$\mathbb{E}\left[T(X_1, \dots X_n)\right] = \theta.$$

**Définition 7** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ un estimateur $\theta$. On dit que cet estimateur est asymptotiquement sans biais si et seulement si*

$$\lim_{n \to \infty} \mathbb{E}\left[T(X_1, \dots X_n)\right] = \theta.$$

**Définition 8** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ un estimateur $\theta$. On dit que cet estimateur converge en probabilité vers $\theta$ si et seulement si*

$$\forall \epsilon > 0 \qquad \mathbb{P}\left\{|T(X_1, \dots X_n) - \theta| > \epsilon\right\} \to 0.$$

**Définition 9** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ un estimateur $\theta$. On dit que cet estimateur converge presque surement vers $\theta$ si et seulement si*

$$\mathbb{P}\left\{T(X_1, \dots X_n) \to \theta\right\} = 1.$$

**Définition 10** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, \dots X_n)$ un estimateur $\theta$. On dit que cet estimateur converge presque complètement vers $\theta$ si et seulement si*

$$\forall \epsilon > 0 \qquad \sum_{n=1}^{\infty} \mathbb{P}\left\{|T(X_1, \dots X_n) - \theta| > \epsilon\right\} < \infty.$$

**Définition 11** *On dit que l'estimateur $T(X_1, \dots X_n)$ converge en moyenne quadratique vers $\theta$ si*

$$\mathrm{E}\Big[T(X_1, \dots X_n) - \theta\Big]^2 \longrightarrow 0 \qquad \textit{quand} \qquad n \longrightarrow +\infty.$$

**Proposition 1** *Soit $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$. L'estimateur de variance minimale est unique parmi les estimateur sans biais.*

**Démonstration**
On va faire la démonstration par l'absurde. On suppose que nous avons deux estimateurs $T(X_1, X_2, \dots X_n)$ et $S(X_1, X_2, \dots X_n)$ tel que

$$T(X_1, X_2, \dots X_n) \neq S(X_1, X_2, \dots X_n)$$

$$\text{mais} \quad \mathbb{E}\left[T(X_1, X_2, \dots X_n)\right] = \mathbb{E}\left[S(X_1, X_2, \dots X_n)\right] = \theta.$$

De plus les deux estimateurs ont de variance minimale. Autrement dit

$$Var\left[T(X_1, X_2, \dots X_n)\right] = Var\left[S(X_1, X_2, \dots X_n)\right]$$

est la plus petite variance de tous les estimateurs. On considère l'estimateur

$$L(X_1, X_2, \dots X_n) = \frac{T(X_1, X_2, \dots X_n) + S(X_1, X_2, \dots X_n)}{2}.$$

Cet estimateur est sans biais pour le paramètre $\theta$, car

$$\mathbb{E}\left[L(X_1, X_2, \dots X_n)\right] = \frac{1}{2}\left(\mathbb{E}\left[S(X_1, X_2, \dots X_n)\right] + \mathbb{E}\left[T(X_1, X_2, \dots X_n)\right]\right) = \frac{2\theta}{2} = \theta.$$

D'autre part,

$$
\begin{aligned}
&Var\left[L(X_1, X_2, \ldots X_n)\right] \\
&\quad = \frac{1}{4} Var\left(\left[S(X_1, X_2, \ldots X_n)\right] + \left[T(X_1, X_2, \ldots X_n)\right]\right) \\
&\quad = \frac{1}{4}\left(Var\left[S(X_1, X_2, \ldots X_n)\right] + Var\left[T(X_1, X_2, \ldots X_n)\right]\right. \\
&\qquad \left. + 2Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right)\right) \\
&\quad = \frac{1}{4}\left(2Var\left[S(X_1, X_2, \ldots X_n)\right] + 2Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right)\right).
\end{aligned}
$$

L'inégalité de Cauchy-Schwarz permet d'écrire

$$
\begin{aligned}
&Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right) \\
&\qquad \leq \sqrt{Var\left[S(X_1, X_2, \ldots X_n)\right] Var\left[T(X_1, X_2, \ldots X_n)\right]} = Var\left[S(X_1, X_2, \ldots X_n)\right].
\end{aligned}
$$

Plus précisément nous avons deux cas :
cas où

$$
\begin{aligned}
&Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right) \\
&\qquad < \sqrt{Var\left[S(X_1, X_2, \ldots X_n)\right] Var\left[T(X_1, X_2, \ldots X_n)\right]}
\end{aligned}
\tag{1.1}
$$

et le cas où

$$
\begin{aligned}
&Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right) \\
&\qquad = \sqrt{Var\left[S(X_1, X_2, \ldots X_n)\right] Var\left[T(X_1, X_2, \ldots X_n)\right]}.
\end{aligned}
\tag{1.2}
$$

Si (1.1) est vérifié alors

$$
\begin{aligned}
Var\left[L(X_1, X_2, \ldots X_n)\right] &< \frac{1}{4}\left(2Var\left[S(X_1, X_2, \ldots X_n)\right]\right. \\
&\left. + 2Cov\left(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)\right)\right) < Var\left[S(X_1, X_2, \ldots X_n)\right]
\end{aligned}
$$

Ce qui est en contradiction avec $S(X_1, X_2, \ldots X_n)$ est un estimateur de variance minimale. Cependant si c'est l'autre cas (1.2) qui est vérifié, on aura

$$
\rho_{S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n)} = \frac{Cov(S(X_1, X_2, \ldots X_n), T(X_1, X_2, \ldots X_n))}{\sqrt{Var[T(X_1, X_2, \ldots X_n)] Var[S(X_1, X_2, \ldots X_n)]}} = 1
$$

Ce qui implique

$$T(X_1, X_2, \dots X_n) = aS(X_1, X_2, \dots X_n) + b, \quad \text{avec} \quad a > 0.$$

Mais comme

$$Var[T(X_1, X_2, \dots X_n)] = Var[S(X_1, X_2, \dots X_n)]$$

et

$$\mathbb{E}[T(X_1, X_2, \dots X_n)] = \mathbb{E}[S(X_1, X_2, \dots X_n)] = \theta.$$

D'où $a$ et $b$ sont solutions du système

$$\begin{cases} a^2 = 1 \\ \theta = a\theta + b. \end{cases}$$

Donc $a = 1$ et $b = 0$. Ainsi,

$$T(X_1, X_2, \dots X_n) = S(X_1, X_2, \dots X_n).$$

Ce qui est en contradiction avec la première condition

$$T(X_1, X_2, \dots X_n) \neq S(X_1, X_2, \dots X_n).$$

■

**Proposition 2** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$ et $T(X_1, X_2, \dots X_n)$ un estimateur sans biais de $\theta$, alors $T(X_1, X_2, \dots X_n)$ est de variance minimale si et seulement si*

$$\forall \textit{ statistique} \quad U \textit{ telle que } \mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$$

$$\textit{on a} \quad \mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] = 0.$$

**Démonstration**

– ($\Rightarrow$) On suppose que $T(X_1, X_2, \dots X_n)$ est de variance minimale et on démontre que

$$\forall \text{ statistique} \quad U \text{ telle que } \mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$$

$$\text{on a} \quad \mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] = 0.$$

En effet, soit $U(X_1, X_2, \dots X_n)$ une statistique telle que $\mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$. On considère un autre estimateur

$$S(X_1, X_2, \dots X_n) = T(X_1, X_2, \dots X_n) + aU(X_1, X_2, \dots X_n),$$

avec $a > 0$. Il est évident que l'estimateur $S(X_1, X_2, \dots X_n)$ est sans biais pour $\theta$. On calcule la variance de cet estimateur, dont, on obtient

$$Var[S(X_1, X_2, \dots X_n)] = Var[T(X_1, X_2, \dots X_n)]{+}a^2 Var[U(X_1, X_2, \dots X_n)]$$
$$+2aCov(T(X_1, X_2, \dots X_n), U(X_1, X_2, \dots X_n)).$$

En utilisant le faite que $\mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$, on démontre que

$$Cov(T(X_1, X_2, \dots X_n), U(X_1, X_2, \dots X_n)) = \mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)].$$

Il résulte que

$$Var[S(X_1, X_2, \dots X_n)] = Var[T(X_1, X_2, \dots X_n)]{+}a^2 Var[U(X_1, X_2, \dots X_n)]$$
$$+2a\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)].$$

Maintenant, comme $T(X_1, X_2, \dots X_n)$ est de variance minimale alors

$$Var[S(X_1, X_2, \dots X_n)] = Var[T(X_1, X_2, \dots X_n)]{+}a^2 Var[U(X_1, X_2, \dots X_n)]$$
$$+2a\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] > Var[T(X_1, X_2, \dots X_n)].$$

Donc,

$$a^2 Var[U(X_1, X_2, \dots X_n)]{+}2a\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] > 0$$

implique

$$a > \frac{-2\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)]}{Var[U(X_1, X_2, \dots X_n)]}$$

Comme $a$ est quelconque et $Var[U(X_1, X_2, \dots X_n)] > 0$ alors

$$\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] \geq 0 \tag{1.3}$$

On reprend les mêmes idées avec l'estimateur $S'(X_1, X_2, \dots X_n) = T(X_1, X_2, \dots X_n) - aU(X_1, X_2, \dots X_n)$, on obtient

$$a > \frac{2\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)]}{Var[U(X_1, X_2, \dots X_n)]}$$

Ce qui est vérifié seulement lorsque

$$\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] \leq 0 \tag{1.4}$$

De (1.4) and (1.3)), on a

$$\mathbb{E}[T(X_1, X_2, \dots X_n)U(X_1, X_2, \dots X_n)] = 0$$

– ($\Leftarrow$) Maintenant, on suppose que l'estimateur $T(X_1, X_2, \dots X_n)$ vérifie

$$\forall \text{ statistique} \quad U \text{ telle que } \mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$$
$$\text{on a} \quad \mathbb{E}[T(X_1, X_2, \dots X_n) U(X_1, X_2, \dots X_n)] = 0$$

et on démontre que $T(X_1, X_2, \dots X_n)$ est de variance minimale. En effet, soit $S(X_1, X_2, \dots X_n)$ un autre estimateur sans biais pour $\theta$ on a

$$\begin{aligned}
&Var[T(X_1, X_2, \dots X_n)] \\
&\quad = \mathbb{E}[T(X_1, X_2, \dots X_n) - \mathbb{E}[T(X_1, X_2, \dots X_n)]]^2 \\
&\quad = \mathbb{E}[T(X_1, X_2, \dots X_n) - \theta]^2 \\
&\quad = \mathbb{E}[(T(X_1, X_2, \dots X_n) - \theta)(T(X_1, X_2, \dots X_n) - S(X_1, X_2, \dots X_n) \\
&\qquad + S(X_1, X_2, \dots X_n) - \theta)] \\
&\quad = \mathbb{E}[(T(X_1, X_2, \dots X_n) - \theta)(T(X_1, X_2, \dots X_n) - S(X_1, X_2, \dots X_n)] \\
&\qquad + \mathbb{E}[(T(X_1, X_2, \dots X_n) - \theta)(S(X_1, X_2, \dots X_n) - \theta)] \\
&\quad = \mathbb{E}[(T(X_1, X_2, \dots X_n) - \theta)(T(X_1, X_2, \dots X_n) - S(X_1, X_2, \dots X_n)] \\
&\qquad + \mathbb{E}[(T(X_1, X_2, \dots X_n) - \mathbb{E}[T(X_1, X_2, \dots X_n)])(S(X_1, X_2, \dots X_n) \\
&\qquad\qquad - \mathbb{E}[S(X_1, X_2, \dots X_n)])].
\end{aligned}$$

Posons $U(X_1, X_2, \dots X_n) = T(X_1, X_2, \dots X_n) - S(X_1, X_2, \dots X_n)$. Il est clair que $\mathbb{E}[U(X_1, X_2, \dots X_n)] = 0$, donc par hypothèse

$$\mathbb{E}[T(X_1, X_2, \dots X_n) U(X_1, X_2, \dots X_n)] = 0.$$

Alors

$$\begin{aligned}
Var[T(X_1, X_2, \dots X_n)] &= Cov(T(X_1, X_2, \dots X_n), S(X_1, X_2, \dots X_n)) \\
&\leq \sqrt{Var[T(X_1, X_2, \dots X_n)] Var[S(X_1, X_2, \dots X_n)]}.
\end{aligned}$$

Finalement

$$\sqrt{Var[T(X_1, X_2, \dots X_n)]} \leq \sqrt{Var[S(X_1, X_2, \dots X_n)]} \Rightarrow$$
$$Var[T(X_1, X_2, \dots X_n)] \leq Var[S(X_1, X_2, \dots X_n)]$$

■

## 1.3 Information au sens de Fisher

**Définition 12** *Soit $X_1, \dots X_n$ un n-échantillon de $X$ de loi $P_\theta$. On suppose que $P_\theta$ admet une densité $f(\theta, \cdot)$. On appelle quantité d'information au sens de Fisher de $X_1, \dots X_n$, notée $I_n(\theta)$, la quantité suivante*

$$I_n(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log L(\theta, X_1, X_2, \dots, X_n)\right]$$

**Proposition 3** *Soit $X_1, \dots X_n$ un n-échantillon de $X$ de densité $f(\theta, \cdot)$. Si la fonction $L(\theta, x_1, x_2, \dots x_n)$ est de classe $C^2$ par rapport à $\theta$ alors*

$$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta}\log L(X_1, X_2, \dots, X_n)\right]$$

**Démonstration** Comme la fonction $L(\theta, x_1, x_2, \dots x_n)$ est une densité de probabilité alors

$$\int_{\mathbb{R}^n} L(\theta, x_1, x_2, \dots x_n) dx_1 \dots dx_n = 1$$

D'où

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n} L(\theta, x_1, x_2, \dots x_n) dx_1 \dots dx_n \\
&\qquad = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} L(\theta, x_1, x_2, \dots x_n) dx_1 \dots dx_n \\
&\qquad = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}\frac{L(\theta, x_1, x_2, \dots x_n)}{L(\theta, x_1, x_2, \dots x_n)} L(\theta, x_1, x_2, \dots x_n) dx_1 \dots dx_n \\
&\qquad = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log L(X_1, X_2, \dots, X_n)\right] = 0
\end{aligned}$$

De même

$$\frac{\partial^2}{\partial\theta}\int_{\mathbb{R}^n} L(\theta, x_1, x_2, \ldots x_n)dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}\left(\log L(\theta, x_1, x_2, \ldots x_n) L(\theta, x_1, x_2, \ldots x_n)\right) dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n) L(\theta, x_1, x_2, \ldots x_n)$$

$$+ \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n) \frac{\partial}{\partial\theta} L(\theta, x_1, x_2, \ldots x_n) dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n) L(\theta, x_1, x_2, \ldots x_n)$$

$$+ \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n) \frac{\partial}{\partial\theta} L(\theta, x_1, x_2, \ldots x_n) \frac{L(\theta, x_1, x_2, \ldots x_n)}{L(\theta, x_1, x_2, \ldots x_n)} dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n) L(\theta, x_1, x_2, \ldots x_n)$$

$$+ \int_{\mathbb{R}^n} \left(\frac{\partial}{\partial\theta} \log L(\theta, x_1, x_2, \ldots x_n)\right)^2 L(\theta, x_1, x_2, \ldots x_n) dx_1 \ldots dx_n = 0.$$

Ainsi

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta} \log L(X_1, X_2, \ldots, X_n)\right] = -\mathbb{E}\left[\frac{\partial}{\partial\theta} \log L(X_1, X_2, \ldots, X_n)\right]^2$$

ce qui nous permet d'écrire

$$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta} \log L(X_1, X_2, \ldots, X_n)\right].$$

■

**Définition 13** *Soit $X_1, \ldots X_n$ un n-échantillon de $X$ de loi $P_\theta$. On appelle quantité d'information au sens de Fisher portée par la statistique $T(X_1, \ldots X_n)$, notée $I_T(\theta)$, la quantité suivante*

$$I_T(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta} \log g(\theta, T(X_1, X_2, \ldots, X_n))\right]$$

*où $g$ est la densité de la statistique $T(X_1, \ldots X_n)$*

**Proposition 4** *Soint $X_1, \dots X_n$ un n-échantillon de $X$ et $T(X_1, \dots X_n)$ une statistique de densité $g(\theta, \cdot)$. Si la densité $g$ est de classe $C^2$ par rapport à $\theta$ alors*

$$I_T(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta}\log g(\theta, T(X_1, X_2, \dots, X_n))\right]$$

**Démonstration** La démonstration est similaire à la proposition précédente

**Proposition 5** *Soint $X_1, \dots X_n$ un n-échantillon de $X$ de densité $f(\theta, \cdot)$ et $T(X_1, \dots X_n)$ un estimateur pour $\theta$, alors*

$$Var[T(X_1, X_2, \dots, X_n)] \geq \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}[T(X_1, X_2, \dots, X_n)]\right]}{I_n(\theta)}.$$

*Cette inégalité s'appelle inégalité de Cramer-Rao*

**Démonstration**

Par définition, on a

$$\mathbb{E}[T(X_1, X_2, \dots, X_n)]] = \int_{\mathbb{R}^n} T(x_1, x_2, \dots x_n)L(\theta, x_1, x_2, \dots x_n)dx_1 \dots dx_n$$

alors

$$\frac{\partial}{\partial\theta}\mathbb{E}\left[T(X_1, X_2, \dots, X_n)\right]]$$

$$= \int_{\mathbb{R}^n} T(x_1, x_2, \dots x_n)\frac{\partial}{\partial\theta}L(\theta, x_1, x_2, \dots x_n)dx_1 \dots dx_n = 0$$

$$= \int_{\mathbb{R}^n} T(x_1, x_2, \dots x_n)\frac{\partial}{\partial\theta}L(\theta, x_1, x_2, \dots x_n)\frac{L(\theta, x_1, x_2, \dots x_n)}{L(\theta, x_1, x_2, \dots x_n)}dx_1 \dots dx_n$$

$$= \int_{\mathbb{R}^n} \left(T(x_1, x_2, \dots x_n)\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \dots x_n)\right) L(\theta, x_1, x_2, \dots x_n)dx_1 \dots dx_n$$

$$= \mathbb{E}\left[T(X_1, X_2, \dots, X_n)\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \dots x_n)\right]\right].$$

En utilisant $\mathbb{E}\left[\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \dots x_n))\right] = 0$ pour écrire

$$Cov(T(X_1, X_2, \dots, X_n), \frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \dots x_n))$$

$$= \mathbb{E}\left[T(X_1, X_2, \ldots, X_n)\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \ldots x_n)]\right].$$

D'où

$$\frac{\partial}{\partial\theta}\mathbb{E}[T(X_1, X_2, \ldots, X_n)]] = Cov(T(X_1, X_2, \ldots, X_n), \frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \ldots x_n).$$

On applique l'inégalité de Cauchy-Schwarz

$$\frac{\partial}{\partial\theta}\mathbb{E}[T(X_1, X_2, \ldots, X_n)]] \leq \sqrt{Var[T(X_1, X_2, \ldots, X_n)]Var[\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \ldots x_n)]}$$

Du monument que

$$I_n(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log L(\theta, X_1, X_2, \ldots, X_n)\right] = Var[\frac{\partial}{\partial\theta}\log L(\theta, x_1, x_2, \ldots x_n)]$$

on obtient

$$\frac{\partial}{\partial\theta}\mathbb{E}[T(X_1, X_2, \ldots, X_n)]] \leq \sqrt{Var[T(X_1, X_2, \ldots, X_n)]I_n(\theta)}$$

Par conséquent

$$Var[T(X_1, X_2, \ldots, X_n)] \geq \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}[T(X_1, X_2, \ldots, X_n)]\right]}{I_n(\theta)}.$$

■

*Remarque 3*

Si $T(X_1, X_2, \ldots, X_n)$ est un estimateur sans biais alors

$$Var[T(X_1, X_2, \ldots, X_n)] \geq \frac{1}{I_n(\theta)}.$$

**Définition 14** *Soient $X_1, \ldots X_n$ un n-échantillon de $X$ de densité $f(\theta, \cdot)$ et $T(X_1, X_2, \ldots, X_n)$ un estimateur sans bisais de $\theta$. On dit que $T(X_1, X_2, \ldots, X_n)$ est efficace si et seulement si*

$$Var[T(X_1, X_2, \ldots, X_n)] = \frac{1}{I_n(\theta)}.$$

## 1.4 Méthodes d'estimation

**Définition 15** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de densité $f(\theta, \cdot)$. On appelle estimateur du maximum de vraisemblance de $\theta$, notée $\hat{\theta_{MV}}$, la statistique définie par*

$$\hat{\theta_{MV}}(X_1, X_2, \dots, X_n) = \arg\max_{\theta} L(\theta, X_1, X_2, \dots, X_n).$$

*Remarque 4*

Si la densité $f(\theta, \cdot)$ est de classe $C^2$ alors l'estimateur $\hat{\theta_{MV}}$ est solution de l équation

$$\frac{\partial}{\partial\theta} \log L(\theta, X_1, X_2, \dots, X_n) = 0$$

avec

$$\frac{\partial^2}{\partial\theta} \log L(\hat{\theta_{MV}}, X_1, X_2, \dots, X_n) \leq 0$$

**Définition 16** *Soient $X_1, \dots X_n$ un n-échantillon de $X$ de densité $f(\theta, \cdot)$. On suppose que le paramètre $\theta = h(\mathbb{E}[X], \mathbb{E}[X^2], \dots \mathbb{E}[X^k])$, alors, l'estimateur de la méthode du moments, notée $\hat{\theta_{MM}}$, est la statistique définie par*

$$\hat{\theta_{MM}}(X_1, X_2, \dots, X_n) = h\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} X_i^2, \dots \frac{1}{n}\sum_{i=1}^{n} X_i^k\right)$$

Cette construction est motivé par le théorème suivant

**Théorème 1** *Soit $X_1, \dots X_n$ un n-échantillon de $X$ telle que $\mathbb{E}[X^k] < \infty$, alors*

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \to \mathbb{E}[X^k] \quad \textit{en probabilité.}$$

# Chapitre 2

# Estimation fonctionnelle

Ce chapitre contient quatre sections. Dans la première section on donne les notions fondamentales de l'estimation fonctionnelle. La fonction de répartition empirique comme premier estimateur de la fonction de répartition est donné dans la deuxième section. Un estimateur alternatif basé sur la méthode du noyau est étudié à la troisième section. La dernière section est consacrée à l'estimation des quantiles

## 2.1 Paramètre fonctionnel

**Définition 17** *On dit que $\theta$ est un paramètre fonctionnel si $\theta = T(F)$ où*

$$\begin{array}{rccc} T: & \mathcal{F} & \longrightarrow & \Theta \\ & F & \longmapsto & T(F) = \theta \end{array}$$

*$\mathcal{F}$ : l'ensemble des fonctions de répartition.*
*$\Theta$ : l'ensemble des paramètres.*
*$T$ : s'appelle fonctionnelle statistique.*

**Exemple 2**

**(i)** La densité d'une variable aléatoire est un paramètre fonctionnel tel que $\theta = f = d(F)$, dans ce cas $T$ est l'opérateur dérivée. $d$

**(ii)** $\mathrm{E}[X] = \displaystyle\int xdF$ est un paramètre fonctionnel tel que $T(F) = \displaystyle\int xdF$ est le fonctionnel statistique.

**(iii)** $V[X] = \displaystyle\int x^2dF - \Big[\int xdF\Big]^2$ est un paramètre fonctionnel tel que :

$$T(F) = \int x^2dF - \Big[\int xdF\Big]^2$$

**(iv)** La fonction de hasard

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{dF(x)}{1 - F(x)}$$

est un paramètre fonctionnel.

## 2.2 La fonction de répartition empirique

**Définition 18** *Soient $X_1$, $X_2$, ..., $X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, on appelle fonction de répartition empirique la fonction $\hat{F}$ définie par :*

$$\begin{array}{rcl} \hat{F}(x) & : & \mathbb{R} \to [0,1] \\ & & x \to \hat{F}(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} I_{]-\infty,x]}(x_i) \end{array} .$$

**Proposition 6** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, alors $\hat{F}$ est un estimateur sans biais pour $F(x)$.*

**Démonstration** Notre objectif est de verifier que $\mathrm{E}\Big[\hat{F}(x)\Big] \longrightarrow F(x)$ quand $n \longrightarrow \infty$.
Par définition de la fonction de repartition :

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i \leq x]}.$$

Ceci implique :

$$\begin{array}{rcl} \mathrm{E}\Big[\hat{F}(x)\Big] & = & \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathrm{E}\Big[I_{\{X_i \leq x\}}\Big] \\ & = & \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{P}\Big(X_i \leq x\Big) \\ & = & \dfrac{n}{n}F(x) \\ & = & F(x). \end{array}$$

Ainsi $\hat{F}(x)$ est asymptotiqument sans biais. ■

**Proposition 7** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, alors $\hat{F}$ converge presque complètement vers $F$.*

**Démonstration** L'objectif est de verifier que :

$$\forall\epsilon > 0, \sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\hat{F}(x) - F(x)\Big| > \epsilon\Big] < \infty.$$

En effet,

$$\begin{array}{rcl}\mathbb{P}\Big[\Big|\hat{F}(x) - F(x)\Big| > \epsilon\Big] & = & \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big] + \mathrm{E}\Big[\hat{F}(x)\Big] - F(x)\Big| > \epsilon\Big] \\ & \leq & \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \dfrac{\epsilon}{2}\Big] + \mathbb{P}\Big[\Big|\mathrm{E}\Big[\hat{F}(x)\Big] - F(x)\Big| > \dfrac{\epsilon}{2}\Big]\end{array}.$$

Pour que :

$$\forall\epsilon > 0, \quad \sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\hat{F}(x) - F(x)\Big| > \epsilon\Big] < +\infty.$$

Il suffit de vérifier que :

$$\sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon/2\Big] < +\infty \tag{2.1}$$

et

$$\sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\mathrm{E}\Big[\hat{F}(x)] - F(x)\Big| > \epsilon/2\Big] < +\infty. \tag{2.2}$$

D'aprés la proposition précédente $\mathrm{E}[\hat{F}(x)] = F(x)$ ce qui est équivalent à

$$\forall\epsilon > 0, \sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\mathrm{E}\Big[\hat{F}(x)\Big] - F(x)\Big| > \epsilon/2\Big] = 0.$$

Maintenant il reste de montrer que :

$$\sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon/2\Big] < +\infty.$$

En effet, on considère la suite

$$x_n = \sqrt{\frac{\log n}{n}} \to 0 \textit{ quand } n \to \infty$$

et on démontre qu'il existe $\epsilon_0$ tel que :

$$\sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon_0\sqrt{\frac{\log n}{n}}\Big] < +\infty$$

Comme la suite $x_n \to 0$

$$\forall \frac{\epsilon}{2}, \exists N \textit{ tel que } \forall n > N, \epsilon_0 |x_n| < \frac{\epsilon}{2}$$

Si $n > N$

$$\mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon/2\Big] \leq \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon_0 |x_n|\Big].$$

En effet pour

$$\sum_{n=1}^{\infty} \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon_0 \sqrt{\frac{\log n}{n}}\Big] < +\infty$$

on a

$$\begin{aligned} &\sum_{n=N}^{\infty} \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} I_{]-\infty,x]}(X_i) - \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^{n} I_{]-\infty,x]}(X_i)\Big] \geq \epsilon_0 \sqrt{\frac{\log n}{n}}\Big] \\ &= \sum_{n=N}^{\infty} \mathbb{P}\Big[\sum_{i=1}^{n} I_{]-\infty,x]}(X_i) - \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^{n} I_{]-\infty,x]}(X_i)\Big] \geq \epsilon_0 \sqrt{n \log n}\Big] \end{aligned}.$$

On applique l'inégalité de Bernstein :

$$\begin{array}{rcl} X_i & \leftrightarrow & I_{]-\infty,x]}(X_i) \\ \alpha_i & = & 0 \\ \beta_i & = & 1 \\ t & = & \epsilon_0 \Big[\sqrt{n \log n}\Big] \end{array}.$$

Donc :

$$\begin{aligned} \sum_{n=N}^{\infty} \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon_0 \sqrt{\frac{\log n}{n}}\Big] &\leq 2\sum_{n=N}^{\infty} \exp\Big(\frac{-2\epsilon_0^2 n \log n}{n}\Big) \\ &\leq 2\sum_{n=N}^{\infty} n^{-2\epsilon_0^2} \end{aligned}$$

pour un $\epsilon_0$ tel que $2\epsilon_0^2 > 1$,

$$\sum_{n=N}^{\infty} \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon/2\Big] < \infty.$$

Par conséquent :

$$\sum_{n=1}^{\infty} \mathbb{P}\Big[\Big|\hat{F}(x) - \mathrm{E}\Big[\hat{F}(x)\Big]\Big| > \epsilon/2\Big] < \infty.$$

Finalement, $\hat{F}(x)$ est un estimateur qui converge presque complètement. ■

**Proposition 8** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, alors $\hat{F}$ converge en moyenne quadratique vers $F$.*

**Démonstration** Pour montrer que $\hat{F}$ converge en moyenne quadratique vers $F$ il suffit de montrer que : que $V[\hat{F}(x)] \to 0$ car :

$$\Big[\mathrm{E}\Big[\hat{F}(x) - F(x)\Big]\Big]^2 = 0$$

et

$$V[\hat{F}(x) - F(x)] = V[\hat{F}(x)]$$

car

$$V\Big[\hat{F}(x) - F(x)\Big] = \mathrm{E}\Big[\hat{F}(x) - F(x)\Big]^2 - \Big[\mathrm{E}\Big[\hat{F}(x) - F(x)\Big]\Big]^2.$$

Pour ce faire, On re-ordonne les observations :

$$X_1, \ldots, X_n = X_{(1)}, \ldots, X_{(n)}.$$

On obtient donc la definition suivante de $\hat{F}(x)$ :

$$n\hat{F}(x) = \begin{cases} 0 & x \leq X_{(1)} \\ i & X_{(i)} \leq x \leq X_{(i+1)} \\ n & x \geq X_{(n)} \end{cases}$$

Les valeurs possibles pour $n\hat{F}(x)$ sont $\{0, 1, \ldots, n\}$.

$$\begin{aligned} \mathbb{P}\Big[n\hat{F}(x) = 0\Big] &= \mathbb{P}\Big[\forall i, X_i > x\Big] \\ &= \mathbb{P}\Big[\bigcap_{i=1}^{n} X_i > x\Big] \\ &= \prod_{i=1}^{n} \mathbb{P}\Big[X_i > x\Big] \\ &= \prod_{i=1}^{n} \Big\{1 - \mathbb{P}\Big(X_i \leq x\Big)\Big\} \\ &= \Big(1 - F(x)\Big)^n \end{aligned}$$

$$\mathbb{P}\Big[n\hat{F}(x) = i\Big] = C_n^i F(x)^i \Big(1 - F(x)\Big)^{n-i} C_n^i.$$

Donc :

$$n\hat{F}(x) \hookrightarrow B\Big(n, p = F(x)\Big).$$

$$\mathrm{E}\Big[nF(x)\Big] = nF(x).$$

$$V\Big[n\hat{F}(x)\Big] = nF(x)\Big(1 - F(x)\Big).$$

$$V\Big[n\hat{F}(x)\Big] = n^2 V\Big[\hat{F}(x)\Big].$$

$$\begin{aligned} V\Big[\hat{F}(x)\Big] &= \frac{nF(x)(1-F(x))}{n^2} \\ &= \frac{F(x)(1-F(x))}{n} \to 0 \end{aligned}.$$

D'où $\hat{F}$ est un estimateur qui converge en moyenne quadratique. ∎

## 2.3 L'estimateur à noyau de la fonction de répartition

**Définition 19** *Soient $X_1, X_2, \ldots, X_n$ un n-échantillon de $X$ de fonction de répartition $F$, on appelle estimateur à noyau $H$ l'estimateur définit par :*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} H\Big(\frac{x - X_i}{h_n}\Big).$$

*Où $(h_n)_n$ est une suite convergente vers 0 quand $n$ tend vers $\infty$ et $H$ est une fonction de répartition. La suite $h_n$ est appelé paramètre de lissage.*

**Théorème 2** *Soient $X_1,X_2, \ldots, X_n$ un n-échantillon de $X$ de fonction de répartition $F$, si*

$$\int_{-\infty}^{+\infty} tH'(t)\,dt < +\infty$$

*alors l'estimateur à noyau $F_n$ est asymptotiquement sans biais.*

**Démonstration** Notre objectif est de vérifier que $\mathrm{E}\Big[F_n(x)\Big] \longrightarrow F(x)$ quand $n \longrightarrow \infty$.

$$\begin{aligned} \mathrm{E}\Big[F_n(x)\Big] &= \mathrm{E}\Big[\frac{1}{n}\sum_{i=1}^{n} H\Big(\frac{x - X_i}{h_n}\Big)\Big] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big] \\ &= \mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big] \\ &= \int_{-\infty}^{+\infty} H\Big(\frac{x - z}{h_n}\Big) f(z)\,dz. \end{aligned}$$

On utilise l'intégration par parties, posons

$$U = H\Big(\frac{x-z}{h_n}\Big) \Rightarrow \; dU = -\frac{1}{h_n} H'\Big(\frac{x-z}{h_n}\Big) dz$$
$$dV = f(z)\, dz \Rightarrow V = F(z).$$

D'où

$$\begin{aligned}
\mathrm{E}\Big[F_n(x)\Big] &= \Big[H\Big(\frac{x-z}{h_n}\Big)F(z)\Big]_{-\infty}^{+\infty} + \frac{1}{h_n}\int_{-\infty}^{+\infty} H'\Big(\frac{x-z}{h_n}\Big)F(z)\, dz \\
&= \frac{1}{h_n}\int_{-\infty}^{+\infty} H'\Big(\frac{x-z}{h_n}\Big)F(z)\, dz.
\end{aligned}$$

On considère maintenant le changement de variable suivant :

$$t = \frac{x-z}{h_n} \;\Rightarrow\; dt = -\frac{1}{h_n}\, dz$$

Alors

$$\begin{aligned}
\mathrm{E}[F_n(x)] &= \frac{1}{h_n}\int_{-\infty}^{+\infty} H'(t)F(x-h_n t)h_n\, dt \\
&= \int_{-\infty}^{+\infty} H'(t)F(x-h_n t)\, dt.
\end{aligned}$$

Maintenant on applique le développement de Taylor à l'ordre 1 de la fonction $F$, on obtient :

$$F(x) = F(x_0) + (x - x_0)F'(\xi) \quad \text{où} \quad \xi \in ]x, x_0[.$$

$$\text{avec :} \quad \begin{array}{ccc} x & \leftrightarrow & x - h_n t \\ x_0 & \leftrightarrow & x \end{array}$$

donc on aurra :

$$\begin{aligned}
F(x-h_n t) &= F(x) + (x - h_n t - x)F'(\xi) \\
&= F(x) - h_n t F'(\xi) \quad \text{où} \quad \xi \in ]x - h_n t, x[.
\end{aligned}$$

D'où

$$\begin{aligned}
\mathrm{E}\Big[F_n(x)\Big] &= \int_{-\infty}^{+\infty} H'(t)\Big(F(x) - h_n t F'(\xi)\Big)dt \\
&= F(x)\int_{-\infty}^{+\infty} H'(t)\, dt - h_n \int_{-\infty}^{+\infty} tH'(t)F'(\xi)t \\
&= F(x) - h_n \int_{-\infty}^{+\infty} tH'(t)F'(\xi)t.
\end{aligned}$$

Au passage à la limite, on a $h_n$ tend vers 0 quand $n$ tend vers $+\infty$, donc $\mathrm{E}\Big[F_n(x)\Big]$ tend vers $F(x)$ quand $n$ tend vers $+\infty$. Ainsi $F_n(x)$ est asymptotiquement sans biais. ■

**Proposition 9** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, alors sous les mêmes conditions du théorème précédent l'estimateur à noyau $F_n$ converge presque complètement vers $F$*

**Démonstration** Pour montrer que $F_n$ converge presque complètement vers $F$ il faut montrer que :

$$\forall \epsilon > 0, \qquad \sum_{n=1}^{\infty} \mathbb{P}\Big[\Big|F_n(x) - F(x)\Big| > \epsilon\Big] < +\infty$$

Soit $\epsilon > 0$ on a

$$\begin{aligned} \mathbb{P}\Big[\Big|F_n(x) - F(x)\Big| > \epsilon\Big] &= \mathbb{P}\Big[\Big|F_n(x) - \mathrm{E}\Big[F_n(x)\Big] + \mathrm{E}\Big[F_n(x)\Big] - F(x)\Big| > \epsilon\Big] \\ &\leq \mathbb{P}\Big[\Big|F_n(x) - \mathrm{E}\Big[F_n(x)\Big]\Big| > \epsilon/2\Big] + \mathbb{P}\Big[\Big|\mathrm{E}\Big[F_n(x)\Big] - F(x)\Big| > \epsilon/2\Big]. \end{aligned}$$

Donc il suffit de vérifier que :

$$\sum_{n=1}^{\infty} \mathbb{P}\Big[\Big|F_n(x) - \mathrm{E}[F_n(x)]\Big| > \epsilon/2\Big] < +\infty \tag{2.3}$$

et

$$\sum_{n=1}^{\infty} \mathbb{P}\Big[\Big|\mathrm{E}[F_n(x)] - F(x)\Big| > \epsilon/2\Big] < +\infty \tag{2.4}$$

Pour la première inégalité on a pour $\epsilon_0 > 0$

$$\mathbb{P}\Big[\Big|F_n(x) - \mathrm{E}\Big[F_n(x)\Big]\Big| > \epsilon_0\sqrt{\frac{\log n}{n}}\Big]$$

$$= \mathbb{P}\Big[\Big|\frac{1}{n}\sum_{n=1}^{\infty}\Big[H\Big(\frac{x - X_i}{h_n}\Big) - \mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big]\Big]\Big| > \epsilon_0\sqrt{\frac{\log n}{n}}\Big]$$

$$= \mathbb{P}\Big[\Big|\sum_{n=1}^{\infty}\Big[H\Big(\frac{x - X_i}{h_n}\Big) - \mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big]\Big]\Big| > \epsilon_0\sqrt{n\log n}\Big]$$

Or, $0 \leq H\Big(\frac{x-X_i}{h_n}\Big) \leq 1$ avec $\alpha = 0$ et $\beta = 1$, donc on peut appliquer l'inégalité de BERNSTEIN, on trouve :

$$\begin{aligned}
\mathbb{P}\Big[\Big|\sum_{n=1}^{\infty}\Big[H\Big(\frac{x-X_i}{h_n}\Big)-\mathrm{E}\Big[H\Big(\frac{x-X_i}{h_n}\Big)\Big]\Big]\Big| > \epsilon_0\sqrt{n\log n}\Big] &\leq 2\exp\Big(\frac{-2\epsilon_0^2 n\log n}{n}\Big)\\
&\leq 2\exp\Big(-2\epsilon_0^2\log n\Big)\\
&\leq 2n^{-2\epsilon_0^2}.
\end{aligned}$$

Il suffit de prendre $2\epsilon_0^2 > 1$ pour conclure que :

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\Big|F_n(x)-\mathrm{E}\Big[F_n(x)\Big]\Big| > \frac{\epsilon}{2}\right] < +\infty.$$

Pour la deuxième inégalité, on a d'après le théorème précédent $\mathrm{E}\Big[F_n(x)\Big]-F(x)$ tend vers 0 quand $n$ tend vers $+\infty$ ce qui est équivalent à

$$\forall\epsilon > 0, \exists N \in \mathbb{N} \,/\, \forall n,\ n \geq N \Longrightarrow \left|\mathrm{E}\Big[F_n(x)\Big]-F(x)\right| < \epsilon$$

donc :

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbb{P}\Big[\Big|\mathrm{E}\Big[\mathrm{F_n(x)}\Big]-F(x)\Big| > \frac{\epsilon}{2}\Big] &= \sum_{n=1}^{N}\mathbb{P}\Big[\Big|\mathrm{E}\Big[F_n(x)\Big]-F(x)\Big| > \frac{\epsilon}{2}\Big]\\
&+ \sum_{n=N}^{\infty}\underbrace{\mathbb{P}\Big[\Big|\mathrm{E}\Big[F_n(x)\Big]-F(x)\Big| > \frac{\epsilon}{2}\Big]}_{0}\\
&= \sum_{n=1}^{N}\mathbb{P}\Big[\Big|\mathrm{E}\Big[F_n(x)\Big]-F(x)\Big| > \frac{\epsilon}{2}\Big] \ < \ \infty.
\end{aligned}$$

■

**Proposition 10** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$, alors sous les mêmes conditions du théorème précédent l'estimateur à noyau $F_n$ converge en moyenne quadratique vers $F$.*

**Démonstration** Pour montre que $F_n$ converge en moyenne quadratique vers $F$ il faut montrer que :

$$\mathrm{E}\Big[F_n(x)-F(x)\Big]^2 \quad \text{tend vers 0 quand } n \text{ tend vers } \infty$$

On a :

$$V\Big[F_n(x)\Big] = V\Big[F_n(x) - F(x)\Big] = \mathrm{E}\Big[F_n(x) - F(x)\Big]^2 - \Bigg(\mathrm{E}\Big[F_n(x) - F(x)\Big]\Bigg)^2$$

Donc il suffit de montrer que

$$V\Big[F_n(x)\Big] \to 0 \text{ } quand \text{ } n \to \infty$$

Puisque le deuxième terme tend vers 0. En effet,

$$\begin{aligned} V\Big[F_n(x)\Big] &= V\Big[\frac{1}{n}\sum_{i=1}^{n} H\Big(\frac{x - X_i}{h_n}\Big)\Big] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} V\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big] \\ &= \frac{1}{n} V\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big] \end{aligned}$$

et

$$V\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big] = \mathrm{E}\Big[H^2\Big(\frac{x - X_i}{h_n}\Big)\Big] - \Bigg(\mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big]\Bigg)^2$$

D'une part, une intégration par parties et un changement de variable donnent

$$\begin{aligned} \mathrm{E}\Big[H^2\Big(\frac{x - X_i}{h_n}\Big)\Big] &= \int_{-\infty}^{+\infty} H^2\Big(\frac{x - z}{h_n}\Big) f(z)\, dz \\ &= 2\int_{-\infty}^{+\infty} H(t)H'(t)F(x - h_n t)\, dt. \end{aligned}$$

D'autre part le développement de Taylor à l'ordre 1 nous permet d'écrire

$$\mathrm{E}\Big[H^2\Big(\frac{x - X_i}{h_n}\Big)\Big] = 2F(x)\int_{-\infty}^{+\infty} H(t)H'(t)\, dt$$

$$-2h_n \int_{-\infty}^{+\infty} tH(t)H'(t)f(\xi)\, dt, \quad \xi \in ]x - h_n t, x[.$$

Au passage à la limite quand $n$ tend vers $\infty$, $\mathrm{E}\Big[H^2\Big(\frac{x - X_i}{h_n}\Big)\Big]$ tend vers $2F(x)\int_{-\infty}^{+\infty} H(t)H'(t)\, dt < \infty$. Or, $\mathrm{E}\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big]$ tend vers $F(x)$.

Ainsi, $V\Big[F_n(x)\Big] = \frac{1}{n} V\Big[H\Big(\frac{x - X_i}{h_n}\Big)\Big]$ tend vers 0 quand $n$ tend vers $\infty$.

D'où la convergence en moyenne quadratique de l'estimateur à noyau $F_n$ vers $F$ . Ceci achève la démonstration de la proposition. ■

## 2.4 L'estimation des quantiles

**Définition 20** *Soit $X$ une variable aléatoire de fonction de répartition $F$. On appelle quantile d'ordre $p \in ]0,1[$ la quantité :*

$$t_p = F^{-1}(p).$$

*Remarque 5*

Le quantile est un paramètre fonctionnel. Si $F$ n'est pas inversible on prend

$$t_p = \inf\{c \,/\, F(c) > p\}.$$

**Définition 21** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$. L'estimateur $\hat{t_p} = F_n^{-1}(p)$ est un estimateur à noyau de quantile $t_p$ où*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} H\Big(\frac{x - X_i}{h_n}\Big)$$

.

*Remarque 6*

Si $F_n$ n'est pas inversible on prend l'estimateur :

$$\hat{t_p} = \inf\{c \,/\, F_n(c) > p\}.$$

**Théorème 3** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de fonction de répartition strictement croissante de classe $C^2$. Si $H$ est strictement croissante de classe $C^2$ vérifiant :*

$$\int_{-\infty}^{+\infty} t^2 H'(t)\, dt < \infty$$

*alors presque complétement on a*

$$\hat{t_p} - t_p = O(h^2) - O\Big(\sqrt{\frac{\log n}{nh_n}}\Big).$$

# Chapitre 3

# Estimation de la densité

Ce chapitre est divisé en six sections. Dans la première section on présente notre méthode d'estimation. Le biais de l'estimateur construit est traité dans la deuxième section. Tandis que sa variance est étudié dans la troisième section. La convergence en probabilité est démontrée dans la quatrième section. Une application directe de cet estimateur est l'estimation du mode qu'on va le présenté dans la cinquième section. La dernière section est consacrée au problème du choix du paramètre de lissage

## 3.1 Estimateur à noyau de la densité

**Définition 22** *Soit $X_1, X_2, \ldots, X_n$ un n-échantillon de $X$ de densité $f$. On appelle estimateur à noyau pour la densité $f$ la variable aléatoire définie par :*

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h_n}\Big).$$

*Où $K$ est un noyau borné intégrable et $(h_n)_n$ est une suite des réels avec $h_n$ (resp. $nh_n$) tend vers 0 (resp. $+\infty$) quand $n$ tend vers $\infty$.*

## 3.2 Biais de l'estimateur

**Proposition 11** *Soient $X_1$,$X_2, \ldots, X_n$ un n-échantillon de $X$ de densité $f$ de classe $C^1$, alors, si $K$ est tel que :*

$$\int_{-\infty}^{+\infty} tK(t)dt < \infty$$

*l'estimateur $f_n$ est asymptotiquement sans biais.*

**Démonstration** On a

$$\begin{aligned}
\mathrm{E}\Big[f_n(x)\Big] &= \mathrm{E}\Big[\frac{1}{nh_n}\sum_{i=1}^{n} K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{nh_n}\sum_{i=1}^{n} \mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{h_n}\mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{h_n}\int_{-\infty}^{+\infty} K\Big(\frac{x-X_i}{h_n}\Big) f(z)\,dz.
\end{aligned}$$

On considère maintenant le changement de variable suivant

$$t = \frac{x-z}{h_n} \Rightarrow dt = -\frac{1}{h_n}\,dz.$$

$$\begin{aligned}
\mathrm{E}\Big[f_n(x)\Big] &= \frac{1}{h_n}\int_{-\infty}^{+\infty} K(t) f(x-h_n t) h_n\,dt \\
&= \int_{-\infty}^{+\infty} K(t) f(x-h_n t)\,dt.
\end{aligned}$$

Maintenant on applique le développement de Taylor à l'ordre 1, on trouve

$$\begin{aligned}
\mathrm{E}\Big[f_n(x)\Big] &= \int_{-\infty}^{+\infty} K(t)\Big(f(x) - h_n t f'(\xi)\Big)\,dt \\
&= f(x)\int_{-\infty}^{+\infty} K(t)\,dt - h_n \int_{-\infty}^{+\infty} tK(t)f'(\xi)\,dt \\
&= f(x) - h_n \int_{-\infty}^{+\infty} tK(t)f'(\xi)\,dt.
\end{aligned}$$

où $\xi \in ]x-h_n t, x[$. Comme $h_n$ tend vers 0 quand $n$ tens vers $+\infty$ alors $\mathrm{E}\Big[f_n(x)\Big] \to f(x)$. D'où $f_n$ est asymptotiquement sans biais. ∎

## 3.3 Variance de l'estimateur

**Théorème 4** *Soient $X_1, X_2, \ldots, X_n$ un $n$-échantillon de $X$ de densité $f$ de classe $C^1$, alors, si $K$ est tel que :*

$$\int_{-\infty}^{+\infty} K^2(t)\,dt < +\infty \qquad et \qquad \int_{-\infty}^{+\infty} tK^2(t)\,dt < +\infty$$

*on a*

$$V\Big[f_n(x)\Big] = O\left(\frac{1}{nh_n}\right).$$

**Démonstration** On a

$$\begin{aligned}
V\Big[f_n(x)\Big] &= V\Big[\frac{1}{nh_n}\sum_{i=1}^{n} K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{n^2h_n^2}\sum_{i=1}^{n} V\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{nh_n^2} V\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big] \\
&= \frac{1}{nh_n^2}\Big[\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big] - \Big(\mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big]\Big)^2\Big] \\
&= \underbrace{\frac{1}{nh_n}\Big[\frac{1}{h_n}\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big]\Big]}_{A} - \underbrace{\frac{1}{n}\Big[\frac{1}{h_n}\mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big]\Big]^2}_{B}
\end{aligned}$$

D'une part on a

$$\frac{1}{h_n}\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big] = \frac{1}{h_n}\int_{-\infty}^{+\infty} K^2\Big(\frac{x-z}{h_n}\Big) f(z)\,dz.$$

On considère le changement de variable suivant :

$$t = \frac{x-z}{h_n} \Rightarrow dt = \frac{-1}{h_n}\,dz.$$

$$\frac{1}{h_n}\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big] = \int_{-\infty}^{+\infty} K^2(t) f(x-h_nt)\,dt.$$

L'application du développement de Taylor à l'ordre 1 donne :

$$f(x) = f(x_0) + (x-x_0)f'(\xi) \quad \text{où} \quad \xi \in ]x, x_0[.$$

ou encore :

$$\begin{aligned}
f(x-h_nt) &= f(x) + (x-h_nt-x)f'(\xi) \\
&= f(x) - h_ntf'(\xi).
\end{aligned}$$

Il résulte que :

$$\begin{aligned}
\frac{1}{h_n}\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big] &= \int_{-\infty}^{+\infty} K^2(t)\Big(f(x) - h_ntf'(\xi)\Big)\,dt \\
&= f(x)\int_{-\infty}^{+\infty} K^2(t)\,dt - h_n\int_{-\infty}^{+\infty} tK^2(t)f'(\xi)\,dt.
\end{aligned}$$

On voit clairement que $\frac{1}{h_n}\mathrm{E}\Big[K^2\Big(\frac{x-X_i}{h_n}\Big)\Big]$ tend vers $f(x)\int_{-\infty}^{+\infty}K^2(t)\,dt<\infty$ quand $n$ tend vers $\infty$. D'autre part

$$\begin{aligned}\frac{1}{h_n}\mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big] &= \frac{1}{h_n}\int_{-\infty}^{+\infty}K\Big(\frac{x-z}{h_n}\Big)f(z)\,dz\\ &= f(x)\int_{-\infty}^{+\infty}K(t)\,dt-h_n\int_{-\infty}^{+\infty}tK(t)f'(\xi)\,dt.\end{aligned}$$

D'où $\frac{1}{h_n}\mathrm{E}\Big[K\Big(\frac{x-X_i}{h_n}\Big)\Big]$ tend vers $f(x)$ quand $n$ tend vers $\infty$. Finalement,

$$V\Big[f_n(x)\Big]=O\left(\frac{1}{nh_n}\right).$$

■

## 3.4 Convergence de l'estimateur

**Proposition 12** *Sous les mêmes conditions de la proposition précédente $f_n$ converge en moyenne quadratique vers $f$.*

**Démonstration** La méme démonstration que le cas de la fonction de répartition il suffit d'écrire

$$V\Big[f_n(x)\Big]=V\Big[f_n(x)-f(x)\Big]=\mathrm{E}\Big[f_n(x)-f(x)\Big]^2-\left(\mathrm{E}\Big[f_n(x)-f(x)\Big]\right)^2$$

et

$$\mathrm{E}\Big[f_n(x)-f(x)\Big]^2=V\Big[f_n(x)\Big]+\mathrm{E}\Big[f_n(x)-f(x)\Big]^2$$

De même, les deux théorème précédents, montrent que les deux terms $V\Big[f_n(x)\Big]$ et $\mathrm{E}\Big[f_n(x)-f(x)\Big]^2$ tends vers 0 quand $n$ tend vers l'infinie. ■

**Proposition 13** *Sous les memes condition de la proposition precedente, si le noyau $K$ est borné et est vérifie*

$$\forall(x_1,x_2)\in\mathbb{R}^2\ \ |K(x_1)-K(x_2)|\le C|x_1-x_2|$$

*et si*

$$\exists\alpha>0\ \lim_{n\to\infty}n^{\alpha-1/2}h_n^{3/2}=\infty\ et\quad lim_{n\to\infty}\frac{nh_n}{\log n}=\infty.$$

*Alors,*

$$\sup_{y\in(\alpha,\beta)} |\widehat{f}(y) - f(y)| \to 0 \text{ en probabilité} \tag{3.1}$$

**Démonstration** On a

$$f_n(x) - f(x) = f_n(x) - E[f_n(x)] + E[f_n(x)] - f(x)$$

D'après le théorème précédent on a indepedament de $x$ $E[f_n(x)] - f(x)$ tend vers 0 donc, il suffit de démontrer

$$\forall \epsilon > 0 \qquad P\left(\sup_{y\in(\alpha,\beta)} |\mathbb{E}f_n(x) - f_n(x)| > \epsilon\right) \to 0.$$

Pour cela on démontre d'abord que

$$\exists \epsilon_0 > 0 \qquad P\left(\sup_{y\in(\alpha,\beta)} |\mathbb{E}f_n(x) - f_n(x)| > \epsilon_0\sqrt{\frac{\log n}{nh_n}}\right) \to 0.$$

L'idée de cette démonstration est de recouvrir le compact $(\alpha, \beta)$ par un nombre fini d'intervalles $B_k$ de longueurs égales $l_n$. On note $\tau_n$ le nombre de ces intervalles. On a

$$(\alpha, \beta) \subset \bigcup_{k=1}^{k=\tau_n} B_k = ]t_k - l_n; t_k + l_n].$$

On a $\tau_n = l_n^{-1}$, on pose $t_x$ est le points le plus proche de $x$ d'où

$$\mathbb{P}\left(\sup_{x\in(\alpha,\beta)} |\mathbb{E}\widehat{f}(x) - \widehat{f}(x)| > \epsilon\right) \leq \underbrace{\mathbb{P}\left[(\sup_{x\in(\alpha,\beta)} |\mathbb{E}\widehat{f}(x) - \mathbb{E}\widehat{f}(t_x)| > \frac{\epsilon}{3}\right]}_{T(1)} +$$

$$\underbrace{\mathbb{P}\left[\sup_{x\in(\alpha,\beta)} |\mathbb{E}\widehat{f}(t_x) - \widehat{f}(t_x)| > \frac{\epsilon}{3}\right]}_{T(2)} + \underbrace{\mathbb{P}\left[(\sup_{x\in(\alpha,\beta)} |\widehat{f}(t_x) - \widehat{f}(x)| > \frac{\epsilon}{3}\right]}_{T(3)}.$$

Pour T(2)

$$\begin{aligned}
\mathbb{P}\left(\sup_{x\in(\alpha,\beta)} |\mathbb{E}\widehat{f}(t_x) - \widehat{f}(t_x)| < \epsilon\right) &= P\left(\max_{t_i\in 0,\ldots,\tau_n} |\mathbb{E}\widehat{f}(t_k) - \widehat{f}(t_k)| \geq \epsilon\right) \\
&\leq \sum_{k=1}^{\tau_n} \mathbb{P}\left[|\mathbb{E}\widehat{f}(t_k) - \widehat{f}(t_k)| > \epsilon\right] \\
&\leq \tau_n \max_{t_i\in 0,\ldots,\tau_n} \mathbb{P}\left[|\mathbb{E}\widehat{f}(t_k) - \widehat{f}(t_k)| > \epsilon\right]
\end{aligned}$$

Il suffit de calculer :

$$P\left[|\mathbb{E}\widehat{f}(t_k) - \widehat{f}(t_k)| > \epsilon\right], \forall t_i \in 0, ..., \tau_n$$

En effet, on va appliquer l'inégalité de Hoeffding et on exprime

$$\mathbb{E}\widehat{f}(t_k) - \widehat{f}(t_k) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i,$$

oú

$$\Delta_i = \frac{1}{h_n}\left\{K\left(\frac{t_k - X_i}{h_n}\right) - \mathbb{E}K\left(\frac{t_k - X_i}{h_n}\right)\right\}.$$

Pour appliquer l'inégalité de Hoeffding au variable $\Delta_i$ il faut évaluer la moyenne d'ordre deux de $\Delta_i$ et trouver un majorant pour $|\Delta_i|$, en effet, comme le noyau K est borné, donc il existe une constante C telle que

$$|(\Delta_i)| \leq \frac{C}{h_n} = d$$

Pour $\mathbb{E}(\Delta_i)^2$ on a

$$\mathbb{E}(\Delta_i)^2 = \mathbb{E}\left[\left(\frac{1}{h_n}\right)^2 \left[K\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}K\left(\frac{x - X_i}{h_n}\right)\right]^2\right].$$

Il vient que :

$$\mathbb{E}(\Delta_i)^2 = var(\omega_i) \leq \mathbb{E}(\omega_i^2),$$

oú

$$\omega_i = \frac{1}{h_n}K\left(\frac{x - X_i}{h_n}\right).$$

Donc

$$\begin{aligned}\mathbb{E}(\omega_i^2) &= \mathbb{E}\frac{1}{(h_n)^2}K^2\left(\frac{x - X_i}{h_n}\right) \\ &= \frac{1}{(h_n)^2}\int K^2\left(\frac{x - u}{h_n}\right) f(u)f(v)dudv\end{aligned}$$

On considére le changement de variable $z = x - u/h_n$ et $t = y - v/h_H$, on obtient, donc

$$\begin{aligned}\mathbb{E}(\omega_i^2) &\leq \frac{C}{h_n}\int K^2(z)f(x - zh_n)dz \\ &\leq \frac{C}{h_n} \text{ car f une fonction continue et le noyau K est borné}\end{aligned}$$

Il vient que

$$\mathbb{E}(\Delta_i)^2 = var(\omega_i) \leq \mathbb{E}(\omega_i^2) \leq \frac{C}{h_n} = \delta^2.$$

D'après l'inégalité de Hoeffding (Lemme 1.3.1) on a :

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\Delta_i\right| > \epsilon\right) \leq 2\exp^{\frac{-n\epsilon^2 h_n}{4C}}.$$

Si en prend

$$\epsilon = \epsilon_0\sqrt{\frac{\log n}{nh_n}}$$

$$\begin{aligned} P\left[|E\widehat{f}(t_x) - \widehat{f}(t_y)| > \epsilon_0\sqrt{\tfrac{\log n^2}{4Ch_n}}\right] &\leq 2\exp^{\frac{-\epsilon_0^2\log n}{C^2}} \\ &\leq 2n^{\frac{-\epsilon_0^2}{4C}} \end{aligned}$$

donc

$$\mathbb{P}\left(\sup_{y\in(\alpha,\beta)}|E\widehat{f}(t_y) - \widehat{f}(t_y)| > \epsilon_0\sqrt{\frac{\log n}{nh_n}}\right) \leq \tau_n n^{\frac{-\epsilon_0^2}{4C}},$$

sachant que $\tau_n = \frac{c}{l_n}$. Il suffit, donc, de choisir $l_n = n^{-\alpha}$ pour que

$$\mathbb{P}\left(\sup_{y\in(\alpha,\beta)}|E\widehat{f}(t_y) - \widehat{f}(t_y)| > \epsilon_0\sqrt{\frac{\log n}{nh_n}}\right) \to 0.$$

Ceci implique

$$\sup_{y\in(\alpha,\beta)}|\mathbb{E}\widehat{f}(t_y) - \widehat{f}(t_y)| \to 0, \quad \text{en probabilité.}$$

Pour T(3)

On a :

$$\begin{aligned}
|\widehat{f}(t_x) - \widehat{f}(x)| &= \left| \tfrac{1}{nh_n} \textstyle\sum_{i=1}^n K\left(\tfrac{t_x - X_i}{h_n}\right) - \tfrac{1}{nh_n} \textstyle\sum_{i=1}^n K\left(\tfrac{x - X_i}{h_n}\right) \right| \\
&\leq C \frac{1}{nh_n} \sum_{i=1}^n \left| \left(\frac{t_x - X_i}{h_n}\right) - \left(\frac{x - X_i}{h_n}\right) \right| \\
&\leq C \frac{M}{h_n^2} |t_x - x| \\
&\leq C' \frac{l_n}{h_n^2} \\
&\leq C' \frac{n^{-\alpha}}{h_n^2} = o\left(\sqrt{\frac{\log n}{nh_n}}\right),
\end{aligned}$$

d'où

$$\mathbb{P}\left[\sup_{x \in (\alpha,\beta)} |\widehat{f}(t_x) - \widehat{f}(x)| > \frac{\epsilon}{3}\right] = 0$$

Pour T(1), de même que T(3) on a

$$\begin{aligned}
\sup_{y \in (\alpha,\beta)} |\mathbb{E}\widehat{f}(t_y) - \mathbb{E}\widehat{f}(x)| &= \sup_{y \in (\alpha,\beta)} \mathbb{E}|\widehat{f}(t_y) - \widehat{f}(x)| \\
&\leq C' \tfrac{n^{-\alpha}}{h_n^2} = o\left(\sqrt{\tfrac{\log n}{nh_n}}\right)
\end{aligned}$$

On arrive à

$$\mathbb{P}\left(\sup_{x \in (\alpha,\beta)} |\mathbb{E}\widehat{f}(t_x) - \mathbb{E}\widehat{f}(x)| \geq \epsilon_0 \sqrt{\frac{\log n}{nh_n}}\right) = 0.$$

■

## 3.5 Estimation non paramétrique du mode

**Définition 23** *Soit $(\alpha, \beta)$ un compcat de $\mathbb{R}$. Le mode $\theta$ est le point le plus probable. C'est à dire $\theta$ est tel que*

(3.2) $$f(\theta) = \sup_{y \in (\alpha,\beta)} f(x).$$

**Définition 24** *Soit $X_1, X_2, \ldots, X_n$ un n-échantillon de $X$ de densité $f$. L'estimateur $\hat{\theta}$ du mode $\theta$ est*

$$f_n(\widehat{\theta}) = \sup_{y\in(\alpha,\beta)} f_n(x). \tag{3.3}$$

En général, cet estimateur n'est pas unique. Ainsi, dans toute la suite de ce chapitre $\widehat{\theta}$ désigne n'importe quelle variable aléatoire vérifiant (3.3).

**Théorème 5** *Sous les memes condition du Théorème precedent, et si* $\inf_{y\in(\alpha,\beta)} f^2(\xi) > C > 0$ *alors $\widehat{\theta}$ converge en probabilité vers $\theta$.*

**Démonstration** Le développement de Taylor d'ordre 2 de la fonction $f$ au voisinage de $\theta$, en particulier pour le point $\widehat{\theta}$, est

$$f(\widehat{\theta}) = f(\theta) + (\widehat{\theta} - \theta) f^{(2)}(\xi) \tag{3.4}$$

oú $\xi$ est entre $\widehat{\theta}$ et $\theta$, il est donc nécessairement dans le compact $(\alpha, \beta)$. Ainsi :on obtient de la formule 3.4 :

$$|f(\widehat{\theta}) - f(\theta)| = |(\widehat{\theta} - \theta) f^{(2)}(\xi)|$$

On retranche et on ajoute $f_n(\widehat{\theta})$ dans la formule présidente :

$$\begin{aligned} |f(\widehat{\theta}) - f(\theta)| &= |f(\widehat{\theta}) - f_n(\widehat{\theta}) + f_n(\widehat{\theta}) - f(\theta)| \\ &\leq |f(\widehat{\theta}) - f_n(\widehat{\theta})| + |f_n(\widehat{\theta}) - f(\theta)| \end{aligned}$$

Par définition du $\theta$ et $\widehat{\theta}$, on a

$$\begin{aligned} |f(\widehat{\theta}) - f(\theta)| &\leq \sup_{y\in(\alpha,\beta)} |f(y) - f_n(y)| + |\sup_{y\in(\alpha,\beta)} f_n(y) - \sup_{y\in(\alpha,\beta)} f(y)| \\ &\leq \sup_{y\in(\alpha,\beta)} |f_n(y) - f(y)| + \sup_{y\in(\alpha,\beta)} |f_n(y) - f(y)| \\ &\leq 2 \sup_{y\in(\alpha,\beta)} |f_n(y) - f(y)| \end{aligned}$$

Ce qui donne ;

$$|\widehat{\theta} - \theta| \leq \frac{2 \sup_{y\in(\alpha,\beta)} |f_n(y) - f(y)|}{\inf_{y\in(\alpha,\beta)} f^{(2)}\xi)}.$$

En appliquant le théorème précédent on obtient

$$\sup_{y\in(\alpha,\beta)} |f_n(y) - f(y)| \to 0 \quad inprobabilit$$

Ainsi

$$\widehat{\theta} - \theta \to 0 \quad inprobabilit$$

■

## 3.6 Choix du paramètre de lissage

Dans cette section, on donne les principales méthodes de calcul de paramètre de lissage. En général, les méthodes de calcul du paramètre de lissage sont basées sur la minimisation de l'erreur quadratique

$$\mathrm{MSE}_g(x) = \mathrm{E}\Big[g(x) - \hat{g}(x)\Big]^2$$

ou l'erreur quadratique intégrée :

$$\mathrm{IMSE}_g(x) = \int_{-\infty}^{+\infty} \Big[g(x) - \hat{g}(x)\Big]^2 dx.$$

qui est souvent calculé à partir de la fonction de densité $f$. Cependant cette méthode dépend du paramètre inconnu $f$ donc il faut chercher des méthodes plus pratique

# Chapitre 4

# Estimation de la régression non-parametrique

Ce chapitre est divisé en trois sections. Dans la première section on présente notre modèle et son estimateur. Les deux sections contient respectivement la convergen

## 4.1 La régression non parmétrique

**Définition 25** *Soient $X$ et $Y$ deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$. La régression non paramétrique est une fonction notée $r$ vérifie*

$$Y = r(X) + \varepsilon \tag{4.1}$$

*où $\varepsilon$ est une variable aléatoire réelle centrée et indépendente de $X$.*

**Définition 26** *La variable aléatoire $X$ s'appelle variable explicative et la variable aléatoire $Y$ s'appelle variable réponse.*

*Remarque 7*

La fonction de régression nous permet de prévoir la variable $Y$ sachant la variable explicative $X$. Cette fonction est explicitement donnée par

$$r(x) = \mathbb{E}(Y/X = x) \tag{4.2}$$

**Définition 27** *Soit $\{(X_i, Y_i), i = 1, \cdots, n\}$ un $n$- échantillon de (X,Y).On notera $f$ la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$de la variable* *explicative $X$ . L'estimateur à noyau de la fonction $r$ est défini par*

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)} \tag{4.3}$$

*où $K$ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, $h = h_n$ est un paramètre réel strictement positifs.*

*Remarque 8*

Cette estimateur peut être exprimer par

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)} \tag{4.4}$$

avec

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)$$

et

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

## 4.2 Convergence presque complètement

**Théorème 6** *Sous les hypothèses*

*(H1) Les fonctions $r$ et $f$ sont $k$-fois continûment différentiables au voisinage de $x$.*

*(H2) La densité $f$ de la variable explicative est strictement positive au point $x$.*

*(H3) La variable réponse est telle que : $|Y| < M < \infty$*

*(H4) Le noyau $K$ est supposé borné intégrable et à support compact. tel que*

$$\int t^j K(t)dt = 0, \quad \forall j = 1, \ldots, k-1 \quad et \quad 0 < \left| \int t^k K(t)dt \right| < \infty$$

*(H5) Le paramètre de lissage est tel que :*

$$\lim_{n\to\infty} \frac{nh}{\log n} = \infty.$$

*on a*

$$\hat{r}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \ p.co. \tag{4.5}$$

**Démonstration**

Nous avons

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}$$

où

$$\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)$$

et

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

On considère la décomposition suivante :

$$\begin{aligned}
\hat{r}(x) - r(x) \quad &= \frac{1}{\hat{f}(x)}\left(\hat{g}(x) - g(x)\right) + \frac{r(x)}{\hat{f}(x)}\left(f(x) - \hat{f}(x)\right) \\
&= \frac{1}{\hat{f}(x)}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x) + \mathbb{E}\hat{g}(x) - g(x)\right) \\
&+ \frac{r(x)}{\hat{f}(x)}\left(f(x) - \mathbb{E}\hat{f}(x) + \mathbb{E}\hat{f}(x) - \hat{f}(x)\right)
\end{aligned}$$

Ainsi, le théorème est une conséquence directe des lemmes suivants : ■

**Lemme 1** *Sous les hypothèses (H1)-(H4), on a*

$$\mathbb{E}\hat{g}(x) - g(x) = \bigcirc(h^k) \tag{4.6}$$

$$\mathbb{E}\hat{f}(x) - f(x) = \bigcirc(h^k) \tag{4.7}$$

**Démonstration du lemme (1)**

Pour le premier résultat

$$
\begin{aligned}
E\left[\widehat{f}(x)\right] &= E\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{1}{nh_n}E\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{1}{h_n}E\left[K\left(\frac{x-X_1}{h_n}\right)\right] \\
&= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right) f(z)dz \\
&= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right) f(z)dz.
\end{aligned}
$$

le changement des variables $t = \frac{x-z}{h_n}$ donc

$$
E\left[\widehat{f}(x)\right] = \int K(t) f(x - h_n t)dt
$$

on utilise le développement de Taylor

$$
f(x - h_n t) = f(x) + h_n t f^{(1)}(x) + \frac{h_n^2}{2} t^2 f^{(2)}(x) + \ldots + o(h_n^k),
$$

d'où

$$
E\left[\widehat{f}(x)\right] = f(x) + \frac{h_n^k}{2}\int t^k K(t) f^{(k)}(x) + o(h_n^k).
$$

Pour la deuxième équation

$$\begin{aligned}
E\left[\widehat{g}(x)\right] &= E\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)Y_i\right] \\
&= \frac{1}{nh_n}E\left[\sum_{i} = 1^n K\left(\frac{x-X_i}{h_n}\right)Y_i\right] \\
&= \frac{1}{h_n}E\left[K\left(\frac{x-X_1}{h_n}\right)Y_1\right] \\
&= \frac{1}{h_n}E\left[K\left(\frac{x-X_1}{h_n}\right)E\left[Y_1|X=X_1\right]\right] \\
&= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right)f(z)r(z)dz \\
&= \frac{1}{h_n}\int K\left(\frac{x-z}{h_n}\right)g(z)dz.
\end{aligned}$$

le changement des variables $t = \dfrac{x-z}{h_n}$ donc

$$E\left[\widehat{g}(x)\right] = \int K(t)g(x-h_nt)dt.$$

De même, on utilise le développement de Taylor

$$g(x-h_nt) = g(x) + h_ntg^{(1)}(x) + \frac{h_n^2}{2}t^2g^{(2)}(x) + \ldots + o(h_n^k).$$

Ainsi,

$$E\left[\widehat{g}(x)\right] = g(x) + \frac{h_n^k}{2}\int t^kK(t)g^{(k)}(x) + o(h_n^k).$$

■

**Lemme 2** *Sous les hypothèses (H1)-(H5), on a*

$$\hat{g}(x) - \mathbb{E}\hat{g}(x) = \bigcirc\left(\sqrt{\frac{\log n}{nh}}\right) \tag{4.8}$$

$$\hat{f}(x) - \mathbb{E}\hat{f}(x) = \bigcirc\left(\sqrt{\frac{\log n}{nh}}\right) \tag{4.9}$$

**Démonstration du lemme (2)**

On a

$$\begin{aligned}
\hat{g}(x) - \mathbb{E}\hat{g}(x) &= \frac{1}{h}\sum_{i=1}^{n} YK\left(\frac{x-X_i}{h}\right) - \mathbb{E}\frac{1}{h}\sum_{i=1}^{n} YK\left(\frac{x-X_i}{h}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h}\left(Y_iK\left(\frac{x-X_i}{h}\right) - \mathbb{E}Y_iK\left(\frac{x-X_i}{h}\right)\right)
\end{aligned}$$

on pose

$$\Delta_i = \frac{1}{h}\left[Y_iK\left(\frac{x-X_i}{h}\right) - \mathbb{E}Y_iK\left(\frac{x-X_i}{h}\right)\right].$$

Comme K et Y sont borné alors

$$|\Delta_i| \leq \frac{C}{h}.$$

Alors il suffit de calculer

$$Var\Delta_i = Var\left(\frac{1}{h}Y_iK\left(\frac{x-X_i}{h}\right)\right)$$

et

$$\mathbb{E}(\Delta_i^2) \leq \mathbb{E}(\mu_i^2)$$

Où
$\mu_i = \frac{1}{h}\left[Y_iK\left(\frac{x-X_i}{h}\right)\right]$, alors,

$$\begin{aligned}
\mathbb{E}(\mu_i^2) &= \mathbb{E}\left(\frac{1}{h^{2p}}Y^2K^2\left(\frac{x-X_i}{h}\right)\right) \\
&= \mathbb{E}\left(\frac{1}{h^{2p}}\mathbb{E}(Y^2/X_1)K^2\left(\frac{x-X_i}{h}\right)\right) \\
&= \frac{1}{h^{2p}}\int \phi(u)f(u)K^2\left(\frac{x-u}{h}\right)du
\end{aligned}$$

oú $\phi(u) = \mathbb{E}\left(Y^2/X = u\right)$.

On considère le même changement des variables $z = \dfrac{(x-u)}{h}$ donc $dz = \dfrac{-du}{h}$ pour démontrer que :

$$\mathbb{E}(\mu_i^2) = \frac{1}{h}\int \phi(x-zh)f(x-zh)K^2(z)dz.$$

Puisque la fonction $\phi$ est borné, la densité $f$ continue et le noyau $K$ est à support compact, alors, il existe une constante $C$ telle que :

$$\mathbb{E}(\mu_i^2) \leq \frac{C}{h}.$$

La preuve de ce lemme est basée sur l'inégalité de d'Hoeffding, pour laquelle alors, pour tout $\varepsilon$ >0 et pour tout $\delta_2$ >1on a :

$$P\left[\frac{1}{n}\left|\sum_{i=1}^{n}\Delta_i\right| > \varepsilon\right] \leq 2\exp\left(\frac{-n\varepsilon^2}{4\delta_2}\right). \tag{4.10}$$

En prenant

$$\varepsilon = \varepsilon_0\sqrt{\frac{\log n}{nh}}$$

pour tout $\varepsilon_0 > 0$ à :

$$\begin{aligned} P\left[|\mathbb{E}\hat{g}(x) - \hat{g}(x)| > \varepsilon_0\sqrt{\frac{\log n}{nh}}\right] &\leq 2\exp\left(\frac{-n\varepsilon_0^2 h\log n}{4nhC}\right) \\ &\leq 2\exp\left(\frac{\varepsilon_0^2\log n}{4C}\right) \\ &\leq 2n^{\frac{-\varepsilon_0^2}{4C}} \end{aligned}$$

Donc

$$\sum_{i=1}^{n} P\left[|\mathbb{E}\hat{g}(x) - \hat{g}(x)| > \varepsilon_0\sqrt{\frac{\log n}{nh}}\right] \leq \sum_{i=1}^{n} 2n^{\frac{-\varepsilon_0^2}{4C}}.$$

Il suffit de choisir $\varepsilon_0^2/4C > 1$ pour que la série converge d'aprés la série de Rieman.

Pour $\hat{f}(x)$ la démonstration est similaire à celle qui précède

$$\hat{f}(x) - \mathbb{E}\hat{f}(x) = \bigcirc\left(\sqrt{\frac{\log n}{nh}}\right)$$

■

**Lemme 3** *Sous les hypothèses du lemme 2*

$$\exists \delta > 0, \quad \textit{tel que} \quad \sum_{i=1}^{n} P\left[|\hat{f}(x)| < \delta\right] < \infty \tag{4.11}$$

**Démonstration du lemme (3)**
Le lemme (2) entraîne en particulier la convergence presque complète de $\hat{f}(x)$ vers $f(x)$. Ainsi, pour tout $\varepsilon > 0$ on a

$$\sum_{i=1}^{n} P\left[|\hat{f}(x) - f(x)| > \varepsilon\right] < \infty.$$

On remarque que

$$\hat{f}(x) \leq \frac{f(x)}{2} \Rightarrow |\hat{f}(x) - f(x)| \geq \frac{f(x)}{2}.$$

On peut écrire, alors

$$P\left[|\hat{f}(x)| \leq \frac{f(x)}{2}\right] \leq P\left[|\hat{f}(x) - f(x)| > \frac{f(x)}{2}\right].$$

Donc

$$\sum_{i=1}^{n} P\left[|\hat{f}(x)| \leq \frac{f(x)}{2}\right] \leq \sum_{i=1}^{n} P\left[|\hat{f}(x)| \leq \frac{f(x)}{2}\right] < \infty.$$

Il suffit maintenant de prendre $\delta = \dfrac{f(x)}{2}$. ∎

## 4.3 Convergence en moyenne quadratique.

**Théorème 7** *Sous les conditions suivantes :*

- *(H1) Les fonctions $r$ et $f$ sont 2-fois continûment différentiables au voisinage de $x$.*
- *(H2) La densité $f$ de la variable explicative est strictement positive au point $x$.*
- *(H3) Le paramètre de lissage est tel que :*

$$lim_{n\to\infty} h_n = 0 \quad et \quad lim_{n\to\infty} nh_n = \infty.$$

- *(H4) Le noyau $K$ est d'order 2, borné et intégrable.*
- *(H5) La variable réponse est telle que : $|Y| < M < \infty$.*

*on a*

$$E[\widehat{r}(x) - r(x)]^2 = B^2(x)h_n^4 + V(x)\frac{1}{nh_n} + o(h_n^4) + o(\frac{1}{nh_n}) \tag{4.12}$$

*où*

$$B(x) = \frac{\int t^2 K(t)dt}{2}\left(\frac{g^{(2)} - r(x)f^{(2)}(x)}{f(x)}\right) \tag{4.13}$$

*et*

$$V(x) = \int K^2(t)dt\frac{\phi(x) - r^2(x)}{f(x)} \tag{4.14}$$

**Démonstration** La démonstration se fait par un calcul séparé des deux parties : partie biais et partie variance. En effet, l'erreur quadratique peut être exprimer

$$E\left[\widehat{r}(x) - r(x)\right]^2 = \left[E\left(\widehat{r}(x)\right) - r(x)\right]^2 + Var\left[\widehat{r}(x)\right].$$

On considère le développement usuel de $1/z$ :

$$\frac{1}{z} = 1 - (z-1) + \ldots + (-1)^p(z-1)^p + (-1)^{p+1}\frac{(z-1)^{p+1}}{z}, \forall z \neq 0, \forall p \in \mathbb{N}^*.$$

Une application de ce développement pour $z = \widehat{f}(x)/E(\widehat{f}(x))$ et $p = 1$ nous permet d'écrire

$$\begin{aligned}
\widehat{r}(x) - r(x) = \; & \left((E\widehat{f}(x))^{-1}\widehat{g}(x) - r(x)\right) \\
& -(E\widehat{f}(x))^{-2}\left(\widehat{g}(x) - E\widehat{g}(x)\right)\left(\widehat{f}(x) - E\widehat{f}(x)\right) \\
& -(E\widehat{f}(x))^{-2}(E\widehat{g}(x))\left(\widehat{f}(x) - E\widehat{f}(x)\right) \\
& +(E\widehat{f}(x))^{-2}\left(\widehat{f}(x) - E\widehat{f}(x)\right)^2\widehat{r}(x).
\end{aligned} \tag{4.15}$$

Ceci implique que,

$$\begin{aligned}
E\left[\widehat{r}(x)\right] - r(x) = & \left((E\widehat{f}(x))^{-1}E\widehat{g}(x) - r(x)\right) - (E\widehat{f}(x))^{-2}Cov(\widehat{g}(x), \widehat{f}(x)) \\
& +(E\widehat{f}(x))^{-2}E\left(\widehat{f}(x) - E\widehat{f}(x)\right)^2\widehat{r}(x).
\end{aligned}$$

Comme la variable $Y$ est borné, on peut trouver une constante $C > 0$ telle que $\widehat{r}(x) \leq C$. D'où,

$$E\left[\widehat{r}(x)\right] - r(x) = \left((E\widehat{f}(x))^{-1}E\widehat{g}(x) - r(x)\right) - (E\widehat{f}(x))^{-2}Cov(\widehat{g}(x), \widehat{f}(x))$$

$$+(E\widehat{f}(x))^{-2}Var\left(\widehat{f}(x)\right)^2 O(1).$$

Pour la partie dispersion nous inspirons de techniques de Sarda et Vieu (2000) et de Bosq et Lecoutre (1987) et en vertu de l'expression (4.15), nous déduisons que

$$Var\left[\widehat{r}(x)\right] = (E\widehat{f}(x))^{-2}Var\left[\widehat{g}(x)\right] - 2(E\widehat{f}(x))^{-3}(E\widehat{g}(x))Cov(\widehat{g}(x), \widehat{f}(x))$$

$$+(E\widehat{f}(x))^{-4}(E\widehat{g}(x))^2 Var(\widehat{f}(x)) + o\left(\frac{1}{nh_n}\right).$$

Finalement, le Théorème (7) est une conséquence des lemmes suivants : ■

**Lemme 4** *Sous les conditions du Théorème (7) on a*

$$E\left[\widehat{f}(x)\right] = f(x) + h_n^2 \frac{f^{(2)}(x)}{2!} \int t^2 K(t) + o(h_n^2).$$

$$E\left[\widehat{g}(x)\right] = g(x) + h_n^2 \frac{g^{(2)}(x)}{2!} \int t^2 K(t) + o(h_n^2).$$

**Démonstration du Lemme 4** On a

$$\begin{aligned}
E\left[\widehat{f}(x)\right] &= E\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \frac{1}{nh_n} E\left[\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \frac{1}{h_n} E\left[K\left(\frac{x - X_1}{h_n}\right)\right] \\
&= \frac{1}{h_n} \int K\left(\frac{x - z}{h_n}\right) f(z)dz \\
&= \frac{1}{h_n} \int K\left(\frac{x - z}{h_n}\right) f(z)dz.
\end{aligned}$$

le changement des variables $t = \frac{x-z}{h_n}$ donc

$$E\left[\widehat{f}(x)\right] = \int K(t)f(x - h_n t)dt$$

on utilise le développement de Taylor

$$f(x - h_n t) = f(x) + h_n t f^{(1)}(x) + \frac{h_n^2}{2} t^2 f^{(2)}(x) + o(h_n^2),$$

d'où

$$E\left[\widehat{f}(x)\right] = f(x) + \frac{h_n^2}{2} \int t^2 K(t) f^{(2)}(x) + o(h_n^2).$$

Pour la deuxième équation

$$\begin{aligned}
E\left[\widehat{g}(x)\right] &= E\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\
&= \frac{1}{nh_n} E\left[\sum_i = 1^n K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\
&= \frac{1}{h_n} E\left[K\left(\frac{x - X_1}{h_n}\right) Y_1\right] \\
&= \frac{1}{h_n} E\left[K\left(\frac{x - X_1}{h_n}\right) E\left[Y_1 | X = X_1\right]\right] \\
&= \frac{1}{h_n} \int K\left(\frac{x - z}{h_n}\right) f(z) r(z) dz \\
&= \frac{1}{h_n} \int K\left(\frac{x - z}{h_n}\right) g(z) dz.
\end{aligned}$$

le changement des variables $t = \dfrac{x-z}{h_n}$ donc

$$E\left[\widehat{g}(x)\right] = \int K(t) g(x - h_n t) dt$$

on utilise le développement de Taylor

$$g(x - h_n t) = g(x) + h_n t g^{(1)}(x) + \frac{h_n^2}{2} t^2 g^{(2)}(x) + o(h_n^2).$$

Ainsi,

$$E\left[\widehat{g}(x)\right] = g(x) + \frac{h_n^2}{2} \int t^2 K(t) g^{(2)}(x) + o(h_n^2).$$

■

**Corollary 1** *Sous les conditions du Théorème (7) on a*

$$(E\widehat{f}(x))^{-1}E\left[\widehat{g}(x)\right]-r(x)=B(x)h_n^2+o(h_n^2).$$

**Démonstration du Corollaire 1** On a

$(E\widehat{f}(x))^{-1}E\left[\widehat{g}(x)\right]-r(x)$

$$=\frac{g(x)+\frac{h_n^2}{2}\int t^2K(t)g^{(2)}(x)+o(h_n^2)}{f(x)+\frac{h_n^2}{2}\int t^2K(t)f^{(2)}(x)+o(h_n^2)}-r(x)$$

$$=\frac{g(x)+\frac{h_n^2}{2}\int t^2K(t)g^{(2)}(x)+o(h_n^2)-r(x)\left(f(x)+\frac{h_n^2}{2}\int t^2K(t)f^{(2)}(x)+o(h_n^2)\right)}{f(x)+\frac{h_n^2}{2}\int t^2K(t)f^{(2)}(x)+o(h_n^2)}$$

$$=\frac{g(x)+\frac{h_n^2}{2}\int t^2K(t)g^{(2)}(x)+o(h_n^2)-r(x)\left(f(x)+\frac{h_n^2}{2}\int t^2K(t)f^{(2)}(x)+o(h_n^2)\right)}{f(x)+O(h_n^2)}$$

$$=\frac{g(x)+\frac{h_n^2}{2}\int t^2K(t)g^{(2)}(x)-r(x)\left(f(x)+\frac{h_n^2}{2}\int t^2K(t)f^{(2)}(x)\right)}{f(x)}+O(h_n^2)$$

$$=\frac{h_n^2}{2}\int t^2K(t)\left(\frac{g^{(2)}(x)-r(x)f^{(2)}(x)}{f(x)}\right)+O(h_n^2)$$

■

**Lemme 5** *Sous les conditions du Théorème (7) on a*

$$Var\left[\widehat{g}(x)\right]=\frac{1}{nh}f(x)\phi(x)\int K^2(t)dt+o\left(\frac{1}{nh_n}\right),$$

*avec* $\phi(x)=E[Y^2|X=x]$

**Démonstration du Lemme 5** On a

$$\begin{aligned}
Var\left[\widehat{g}(x)\right] &= Var\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)Y_i\right]\\
&= \frac{1}{n^2h_n^2} Var\left[\sum_{=1}^{n} K\left(\frac{x-X_i}{h_n}\right)Y_i\right]\\
&= \frac{1}{nh_n^2}\left(E\left[K^2\left(\frac{x-X_1}{h_n}\right)Y_1^2\right] - \left(E\left[K\left(\frac{x-X_1}{h_n}\right)Y_1\right]\right)^2\right)
\end{aligned}$$

Il est clair que

$$\frac{1}{h_n^2}\left(E\left[K\left(\frac{x-X_1}{h_n}\right)Y_1\right]\right)^2 = \left(E\left[\widehat{g}(x)\right]\right)^2 = O(h^4).$$

D'où

$$\begin{aligned}
Var\left[\widehat{g}(x)\right] &= \frac{1}{nh_n^2} E\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1^2|X=X_1\right]\right] + o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1^2/X=X_i\right]\right] + o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1^2/X=X_i\right]\right] + o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n^2}\int K^2\left(\frac{x-z}{h_n}\right)f(z)\phi(z)dz + o\left(\frac{1}{nh_n}\right).
\end{aligned}$$

le changement de variable $t = \dfrac{x-z}{h_n}$ et on utilise la continuité de la fonction $f(x)\phi(x)$ pour arriver à

$$Var\left[\widehat{g}(x)\right] = \frac{1}{nh_n}f(x)\phi(x)\int K^2(t)dt + o\left(\frac{1}{nh_n}\right).$$

∎

**Lemme 6** *Sous les conditions du Théorème (7) on a*

$$Cov(\widehat{g}(x),\widehat{f}(x)) = \frac{1}{nh_n}g(x)\int K^2(t)dt + o\left(\frac{1}{nh_n}\right).$$

**Démonstration du lemme 6** On a

$$
\begin{aligned}
Cov(\widehat{g}(x),\widehat{f}(x)) &= Cov\left(\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)Y_i,\frac{1}{nh_n}\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right)\right)\\
&= \frac{1}{n^2h_n^2}Cov\left(\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)Y_i,\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right)\right)\\
&= \frac{1}{nh_n^2}\left(E\left[K^2\left(\frac{x-X_1}{h_n}\right)Y_1\right]\right.\\
&- \left.\left(E\left[K\left(\frac{x-X_1}{h_n}\right)Y_1\right]E\left[K\left(\frac{x-X_1}{h_n}\right)\right]\right)\right)
\end{aligned}
$$

En vertu du calcule précèdent, (voir le lemme 1) on peut conclure

$$
\frac{1}{h_n^2}\left(E\left[K\left(\frac{x-X_1}{h_n}\right)Y_1\right]E\left[K\left(\frac{x-X_1}{h_n}\right)\right]\right)=O(1).
$$

Ainsi,

$$
\begin{aligned}
Cov(\widehat{g}(x),\widehat{f}(x)) &= \frac{1}{h_n}E\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1|X=X_1\right]\right]+o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1/X=X_i\right]\right]+o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)E\left[Y_1/X=X_i\right]\right]+o\left(\frac{1}{nh_n}\right)\\
&= \frac{1}{nh_n}\int K^2\left(\frac{x-z}{h_n}\right)g(z)dz+o\left(\frac{1}{nh_n}\right).
\end{aligned}
$$

On considère le changement des variables usuelle $t=\dfrac{x-z}{h_n}$ et on utilise la continuité de la fonction $g(x)$ pour arriver à

$$
Cov(\widehat{g}(x),\widehat{f}(x))=\frac{1}{nh_n}g(x)\int K^2(t)dt+o\left(\frac{1}{nh_n}\right).
$$

■

**Lemme 7** *Sous les hypothèses (H1)-(H2) et (H4) on a,*

$$
Var\left[\widehat{f}(x)\right]=\frac{1}{nh_n}f(x)\int K^2(t)dt+o\left(\frac{1}{nh_n}\right).
$$

**Démonstration du Lemme 7** On a

$$\begin{aligned}
Var\left[\widehat{f}(x)\right] &= Var\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{1}{n^2h_n^2} Var\left[\sum_{i} = 1^n K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{1}{nh_n^2}\left(E\left[K^2\left(\frac{x-X_1}{h_n}\right)\right] - \left(E\left[K\left(\frac{x-X_1}{h_n}\right)\right]\right)^2\right)
\end{aligned}$$

Il est clair que

$$\frac{1}{h_n^2}\left(E\left[K\left(\frac{x-X_1}{h_n}\right)\right]\right)^2 = (E\left[\widehat{f}(x)\right])^2 = O(1).$$

D'où

$$\begin{aligned}
Var\left[\widehat{f}(x)\right] &= \frac{1}{nh_n^2} E\left[K^2\left(\frac{x-X_1}{h_n}\right)\right] + o\left(\frac{1}{nh_n}\right) \\
&= \frac{1}{nh_n^2} \mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)\right] + o\left(\frac{1}{nh_n}\right) \\
&= \frac{1}{nh_n^2} \mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)\right] + o\left(\frac{1}{nh_n}\right) \\
&= \frac{1}{nh_n}\int K^2\left(\frac{x-z}{h_n}\right) f(z)dz + o\left(\frac{1}{nh_n}\right).
\end{aligned}$$

Le changement des variables $t = \dfrac{x-z}{h_n}$ et si on utilise la continuité de la fonction $f(x)$ pour arriver à

$$Var\left[\widehat{f}(x)\right] = \frac{1}{nh_n} f(x)\int K^2(t)dt + o\left(\frac{1}{nh_n}\right).$$

# Chapitre 5

# Exercices

## 5.1 Exercices sur l'estimation parmétrique

### Exercice 1

Dans chacun des modèles suivants, dire quel est le paramètre et l'ensemble de ses valeurs

- Un échantillon d''une loi de Bernounlli $X_1, X_2, \ldots X_n \approx B(1,p)$,
- Un échantillon d''une loi de Poisson $X_1, X_2, \ldots X_n \approx P(\lambda)$,
- Un échantillon d''une loi de Exponentielle $X_1, X_2, \ldots X_n \approx \exp(\lambda)$,
- Un échantillon d''une loi de Normale $X_1, X_2, \ldots X_n \approx N(m,\sigma)$
- Un échantillon d''une loi de Uniforme $X_1, X_2, \ldots X_n \approx U(0,\theta)$

### Exercice 2

Soit $X$ une variable de densité

$$f(x) = \begin{cases} 1/\theta & x \in ]0,\theta[ \\ 0 & \text{sinon} \end{cases}$$

Montrer que l'estimateur $T = \frac{2}{n}(X_1 + X_2 + \ldots X_n)$ est convergente

### Exercice 3

Soit $X_1, \ldots X_n$ un $n$- échantillon de taille $n$ de loi normale $N(m,\sigma)$ où $m$ et $\sigma$ sont inconnus. Trouver un estimateur sans biais $\sigma^2$

### Exercice 4

Soit $X_1, X_2, \ldots X_n$ un $n$-échantillon de densité

$$f(x) = \frac{1}{a} e^{-x/a} \qquad x > 0$$

- Calculer $E(X)$ et $V(X)$
- Trouver l'estimateur du maximum de vraisemblance de $a$
- Montrer que cet estimateur est sans biais

**Exercice 5**

Soit $X_1, X_2, \dots X_n$ un $n$-échantillon de densité

$$f(x) = \theta x^{\theta-1} \qquad 0 < x < 1$$

Donner l'estimateur du maximum de vraisemblance de $\theta$

**Exercice 6**

Soit $X$ une variable de densité

$$f(x) = \left\{ \begin{array}{cc} \theta x^2 & x \in [-1, 0] \\ 1 - \theta x^2 & x \in [0, 1] \\ 0 & x \notin [-1, 1] \end{array} \right.$$

- Calculer $E(X)$ et $V(X)$
- Estimer par la méthode du moments le paramètre $\theta$
- Est-il sans bias
- Est-il convergent

**Exercice 7**

Soit $X_1, \dots X_n$ un echantillon de taille $n$ de densité

$$f(x, a) = \frac{3x^2}{a} \exp(-x^3/a) \; quuad a > 0$$

- Determiner l'estimateur du maximum de vraisemblance de $a$ noté $\hat{a}$
- Calculer $\frac{\partial \log f(x,a)}{\partial a}$ et en déduire $E[X^3]$
- Calculer la quantite d'information de Fisher de $a$ et en déduire $V(X^3)$
- Montrer que $\hat{a}$ est un estimateur sans biais et convergent de $a$
- Est-il efficace
- On considère la variable $Y = X^3/a$. Calculer la densité de $Y$ et retrouver $E[X^3]$ et $V(X^3)$

**Exercice 8**

Soit $X_1, \dots X_n$ un $n$- échantillon de $X$ de loi $\Gamma(a, 1/\theta)$

a) Calculer les moments de $X$, la fonction caractéristique de $Y = \log X$, la moyenne et la variance de $Y$.
b) $a$ est supposé connu ; estimer $\theta$ par la méthode du maximum de vraisemblance ; quelles sont les qualités de l'estimateur obtenu

**Exercice 9**

Soit $P$ la loi de probabilité de densité $\exp(-x)$ sur $\mathbb{R}^+$ ; on désigne par $P_\theta$ la loi déduite de $P$ par la translation $\theta$.

Déterminer l'estimateur de maximum de vraisemblance de $\theta$ au vu d'un échantillon de taille $n$ de la loi $P_\theta$ ; étudier ses qualités.

**Exercice 10**

Soit $X_1, \dots X_n$ un $n$- échantillon de taille $n$ de loi normale $N(m, \sigma)$ où $m$ et $\sigma$ sont inconnus.

$k$ étant un réel positif, trouver un estimateur sans biais $\sigma^{2k}$

**Exercice 11**

Soit $X_1, \dots X_n$ un $n$- échantillon de loi $P_\theta$. On considère $G_1$ et $G_2$ deux estimateurs, sans biais pour le paramètre $\theta$ de carré sommable.
a) Déterminer une forme linéaire de $G_1$ et $G_2$ qui soit un estimateur sans biais de $\theta$ et qui ait la plus petite variance possible.
b) Que devient le résultat de la question precedente si $G_2$ est un estimateur de variance minimum. En particulier, calculer le coefficient de corrélation de $G_1$ et $G_2$.

**Exercice 12**

Soit $X_1, \dots X_n$ un $n$- échantillon de $X$ de loi $\Gamma(a, 1/\theta)$
a) Calculer les moments de $X$, la fonction caractéristique de $Y = \log X$, la moyenne et la variance de $Y$.
b) $a$ est supposé connu ; estimer $\theta$ par la méthode du maximum de vraisemblance ; quelles sont les qualités de l'estimateur obtenu
c) $\theta$ est supposé connu ; estimer $a$ par la méthode du maximum de vraisemblance ;
d) Comment pourrait-on estimer $a$ quand les deux paramètres sont inconnus

**Exercice 13**

Soit $X_1, \dots X_n$ un $n$- échantillon de $X$ de densité $f_\theta$ définie par

$$f_\theta(x) = \theta(1-x)^{\theta-1} \quad \text{si} \quad 0 \leq x \leq 1 \quad f(x) = 0 \quad \text{sinon}$$

où $\theta$ est un paramètre positif inconnu.

a) Déterminer l'estimateur de maximum de vraisemblance de $\theta$

b) Introduire la variable aléatoire $Z = -\log(1-X)$ et explique le résultat précèdent.

c) L'estimateur déterminer à la question précèdent est -il sans biais Exhaustive

**Exercice 14**

Le cout d'un certain type de sinistre peut être considéré comme une variable aléatoire $X$ suivant la loi normale $N(m, \sigma)$. On observe dans une compagnie d'assurance $n$ dossiers de sinistres indépendants.

- On suppose que l'écart-type $\sigma$ est connu, égal à 15 Euros.
  (a) Calculer l'intervalle bilatéral de confiance 98% pour m. Application numérique : pour 20 dossiers, la moyenne des coûts observée est 120 Euros : dans quelle Fourchete palcer vous $m$
  (b) Combien de dossiers doit-on examiner pour que la la longueur de l'intervalle de 98% soit inférieur ou égale à 10
- L'écart-type $\sigma$ n'est en fait pas connu.
  (a) Comment l'intervalle est-il modifié Application numérique : pour 20 dossiers, la moyenne des coûts observée est 120 Euros et l'estimation sans biais de $sigma^2$ est égale à 15 Euros

**Exercice 15**

Avant une élection, un candidat affirme qu'il est sûr de dépasser le score des 30%. Pour savoir s'il a raison, on interroge un nombre suffisant de personnes en souhaitant que le risque de dire qu'il a raison alors qu'il se trompe n'excède pas 0.05.

- Quelle est l'hypothèse $H_0$ 2. Quelle est l'hypothèse $H_1$
- Quel risque de $1^{er}$ espèce choisissez-vous
- Si on interroge 500 personnes, est-ce que 32% d'intentions de votes dans l'échantillon sera considéré comme significatif

**Exercice 16**

Un appareil de télécommunications reçoit un signal stocké à chaque minute dans une suite de variables $(X_n)$. Cet appareil doit détecter un signal effectif, en le différenciant d'un bruit. On suppose que le bruit est une suite de variables indépendantes de loi normale de moyenne nulle et de variance 1. Pour un signal, la moyenne n'est pas nulle. Aujourd'hui on a observé une suite de 40 variables $(X_1, \dots X_{40})$, supposées indépendantes de variance 1. La moyenne empirique vaut 0.6. S'agit-il de bruit Construire un test de niveau $\alpha = 0.05$ pour répondre à cette question.

**Exercice 17**

On utilise une nouvelle variété de pommes de terre dans une exploitation agricole. Le rendement de l'ancienne variété était de 41.5 tonnes /h . La nouvelle est cultivée sur 100 h, avec un rendement moyen de 45 tonnes /h et un écart-type de 11.25. Faut-il, au vu de ce rendement ( supposé de loi normale) , favoriser la culture de la nouvelle variété

## 5.2 Exercices sur l'estimation non parmétrique

**Exercice 18**

Soit $X_1, X_2, \dots X_n$ un $n$-échantillon de $X$ de densité $f$. On considère l'estimateur à noyau de $f$, noté $f_n$ défini par

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

1. Supposons que $f$ est de classe $C^2$ et $K$ est une noyau symétrique, paire, continue et à support compact, calculer le biais de $f_n(x)$
2. Montrer que cet estimateur converge en moyenne quadratique et préciser son erreur quadratique
3. Trouver la valeur de $h_n$ qui minimise la partie dominante de cette erreur
4. Trouver la valeur exacte de $h_n$ lorsque l'échantillon suit la loi normal.
5. Refaire les questions précédentes pour les estimateurs des dérivée d'ordre $j > 0$, noté $f_n^{(j)}$ et étudier sa converegnce presque complète.

**Exercice 19**

Soit $X_1, X_2, \dots X_n$ un $n$-échantillon de $X$ de densité $f$. On considère l'estimateur à noyau de $f$, noté $f_n$ défini par

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

1. Supposons que $f$ est de classe $C^j$ et $K$ est une noyau à support compact de classe $C^j$, trouver un estimateur à noyau de $f^{(j)}(x)$.
2. Etudier sa converge en moyenne quadratique
3. Etudier sa converegnce presque complète.

**Exercice 20**

On dispose d'un $n$-échantillon $X_1, X_2, \dots X_n$ de $X$ de densité $f$ et un compact $S$ de $\mathbb{R}$

1. Montrer que le mode $\theta = \arg\max_{x \in S} f(x)$ est un paramètre fonctionnel
2. En utilisant l'estimateur à noyau de la densité $f$, déterminer l'estimateur $\theta_n$ pour $\theta$
3. Montrer que
$$|f(\theta) - f(\theta_n)| \leq 2 \sup_{x \in S} |f(x) - f_n(x)|$$
4. En utilisant l'inversibilité a gauche et a droite de la densité $f$ au voisinage de $\theta$, montrer que l'estimateur $\theta_n$ converge presque complètement si
$$\sup_{x \in S} |f_n(x) - f(x)|$$

**Exercice 21**

Soit $X_1, X_2, \dots X_n$ un $n$-échantillon de $X$ de fonction de répartition $F$ et de densité $f$ et soit $x \in \mathbb{R}$ tel que $F(x) < 1$

1. Montrer que la fonction de hasard
$$\lambda(x) = \frac{f(x)}{1 - F(x)}$$
est un paramètre fonctionnel.
2. En déduire un estimateur $\lambda_n(x)$ pour $\lambda(x)$ par la méthode du noyau
3. Montrer que
$$\lambda_n(x) - \lambda(x) = \frac{f_n(x) - f(x)}{1 - F_n(x)} + (F_n(x) - F(x)) \frac{\lambda(x)}{1 - F_n(x)}$$
4. En utilisant la convergence presque complète de $F_n(x)$ vers $F(x)$ montrer qu'il existe $\delta > 0$ tel que :
$$\sum_n P\left((1 - F_n(x)) < \delta\right) < \infty.$$

5. Etudier la convergence presque complète de l'estimateur $\lambda_n(x)$

**Exercice 22**

Soit $(X_1, Y_1) \dots (X_n, Y_n)$ un $n$- échantillon de $(X, Y)$ dans $\mathbb{R}^2$ et soit $\psi$ une fonction mesurable bornée. On considère la fonction de régression de $\psi(Y)$, sachant $X$, définie par

$$\forall x \in \mathbb{R} \qquad r_\psi(x) = E(\psi(Y)|X = x)$$

et on suppose que la densité de la variable explicative $X$ et la fonction $r_\psi$ vérifient la condition suivante :

$$(5.1)\ \exists k > 0,\ \exists C < \infty,\ \exists \epsilon > 0,\ \forall y \in ]x-\epsilon, x+\epsilon[,\ |\phi(x)-\phi(y)| \leq C|x-y|^k,$$

où $\phi$ désigne indifféremment $f$ ou $r_\psi$.

1. Estimer la fonction $r_\psi$ par la méthode du noyau
2. Montrer que, si :
   (a) $\lim_{n\to\infty} h_n = 0$ et $\lim\limits_{n\to\infty} \dfrac{nh_n}{\log n} = \infty$,
   (b) Le noyau $K$ est borné, intégrable et à support compact,
   (c) La fonction $f$ est telle que
   $$f(x) > 0,$$
   alors,
   $$|r_{\psi n}(x) - r_\psi(x)| = O(h_n^k) + O\left(\sqrt{\frac{\log n}{nh_n}}\right) \quad \text{en p.co.}$$

## 5.3 Projets

**Projet 1**
Le gérant d'une entreprise qui commercialise un produit a observé que le processus de vente suit une loi normale de paramètre inconnue. Le fournisseur de ce produit a proposé à l'entreprise une quantité $q$ avec une réduction à négocier. Le gérant a demandé au statisticien de l'entreprise une fonction en $q$ et d'observations $x$ qui donne

1. La probabilité de vendre $q$

2. La varie quantité qu'il faut demander
3. La réduction qu'il doit proposer pour accepter $q$
4. Application numérique : Les observations sont x=(99.99, 105.108, 104.95, 101.82, 98.987, 98.15, 105.07, 102.016, 102.61, 99.65, 97.47, 100.30, 105.97, 98.02, 104.37, 98.89, 105.05, 98.41, 103.33, 104.94)
   q=110
5. Refaire la même question si la loi de probabilité est gamma lorsque x(102.66, 97.17, 96.87, 111.63, 91.86 , 104.72, 90.80, 89.91, 109.79, 104.90, 105.77, 84.43, 124.42, 95.76, 82.35, 90.76, 84.35, 102.82, 98.51, 105.44)
   q=110

**Projet 2**

Le gérant de votre entreprise désire modéliser le comportement des acheteurs d'un produit. Afin de trouver l'emballage la plus adéquate. Il pense que la fréquence d'achat dépend (entre autre) du nombre de paquets achetés à chaque passage. Plus précisément il suppose que le nombre journalier d'acheteur d'un seul paquet est une variable aléatoire $X$ suivant une loi de Poisson $P(\lambda_1)$, alors que le nombre journalier d'acheteurs de deux paquet est un variable aléatoire $Z$ suivant une loi de Poisson $P(\lambda_2)$. Nous supposons qu'il n'y a pas d'acheteur de plus de deux paquets à la fois et que $X$ et $Z$ sont indépendante . La seule observation disponible chaque jour est le nombre total de paquets vendus : Trouver en fonction du coûts de l'emballage la combinaison la plus approprié et le couts de la mauvaise décision.

**Application numérique**

Les observations sont ( 12 18 14 13 17 10 17 16 13 17 14 13 14 13 23 21 12 13 11 19 19 5 17 10 22 13 6 12 15 13)
Couts d'un paquet=50
Couts de deux paquets=70

Même question si il y encore des acheteur de trois paquets donné par une variable aléatoire $Z$ qui suit aussi loi de Poisson ($P(\lambda_3)$)

**Application numérique**

Les observations sont( 21 20 20 14 16 19 29 18 9 22 19 27 20 14 21 18 17 13 14 10 19 12 16 24 14 25 22 20 12 19) Couts de trois paquet=90

**Projet 3**

Le service de cardiologie de CHU de sidi bel abbes a étudié la durée $T$ d'incubation d'un virus sur une population de $n$ individus. Ce virus présente un grand risque inobservable, mais elle suit une loi gamma de paramètres $(a, \lambda)$ et la loi de la durée d'incubation conditionnellement à un facteur de risque $\lambda_r$ est $exp(\lambda_r)$. Afin de préparer la durée d'hospitalisation des patients, ils ont demandé au statisticien du service de programmer une fonction que donne

1. La durée d'incubation la plus probable
2. Le risque moyen
3. La durée d'un incubation pour le risque moyen
4. La durée d'un incubation pour le risque le plus probable.

**Application :** $T =$ (14.793366 13.729592 6.978536 8.976544 3.894525 3.044546 7.889189 9.226947 32.083980 4.121468 3.164000 3.853124 24.220555 3.354880 46.103705 34.544855 9.914948 3.048532 3.4277395.592706).