

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

1	Estimation paramétrique	7
1.1	Structure statistique	7
1.2	Estimateur statistique	8
1.3	Information au sens de Fisher	14
1.4	Méthodes d'estimation	18
2	Estimation fonctionnelle	19
2.1	Paramètre fonctionnel	19
2.2	La fonction de répartition empirique	20
2.3	L'estimateur à noyau de la fonction de répartition	24
2.4	L'estimation des quantiles	29
3	Estimation de la densité	31
3.1	Estimateur à noyau de la densité	31
3.2	Biais de l'estimateur	31
3.3	Variance de l'estimateur	32
3.4	Convergence de l'estimateur	34
3.5	Estimation non paramétrique du mode	38
3.6	Choix du paramètre de lissage	40
4	Estimation de la régression non-paramétrique	41
4.1	La régression non paramétrique	41
4.2	Convergence presque complètement	42
4.3	Convergence en moyenne quadratique.	48
5	Exercices	57
5.1	Exercices sur l'estimation paramétrique	57
5.2	Exercices sur l'estimation non paramétrique	61
5.3	Projets	63

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 Appliquées 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 montrerons dans ce chapitre que l'estimation de la fonction de répartition est un outil indispensable pour l'estimation de n'importe quel 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 dernières sections (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 série d'exercices ainsi que quelques projets de travaux pratiques.

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 (Ω, \mathcal{A}) un espace probabilisable et soit

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

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, \dots, 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_θ . 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_θ 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 θ .

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_θ . Un estimateur de θ est une statistique $T(X_1, \dots, X_n)$ qui prend ces valeurs dans Θ .

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

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

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

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

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

$$\forall \epsilon > 0 \quad \mathbb{P}\{|T(X_1, \dots, X_n) - \theta| > \epsilon\} \rightarrow 0.$$

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

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

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

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

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

$$\mathbb{E} \left[T(X_1, \dots, X_n) - \theta \right]^2 \longrightarrow 0 \quad \text{quand} \quad n \longrightarrow +\infty.$$

Proposition 1 Soit X_1, \dots, X_n un n -échantillon de X de loi P_θ . L'estimateur de variance minimale est unique parmi les estimateurs 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

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

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

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

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

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 θ , car

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

D'autre part,

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

L'inégalité de Cauchy-Schwarz permet d'écrire

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

Plus précisément nous avons deux cas :
cas où

$$\begin{aligned}
(1.1) \quad & Cov (S(X_1, X_2, \dots X_n), T(X_1, X_2, \dots X_n)) \\
&< \sqrt{Var [S(X_1, X_2, \dots X_n)] Var [T(X_1, X_2, \dots X_n)]}
\end{aligned}$$

et le cas où

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

Si (1.1) est vérifié alors

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

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

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

Ce qui implique

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

Mais comme

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

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_θ et $T(X_1, X_2, \dots, X_n)$ un estimateur sans biais de θ , alors $T(X_1, X_2, \dots, X_n)$ est de variance minimale si et seulement si

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

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

Démonstration

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

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

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

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 θ . On calcule la variance de cet estimateur, dont, on obtient

$$\begin{aligned} \text{Var}[S(X_1, X_2, \dots, X_n)] &= \text{Var}[T(X_1, X_2, \dots, X_n)] + a^2 \text{Var}[U(X_1, X_2, \dots, X_n)] \\ &\quad + 2a \text{Cov}(T(X_1, X_2, \dots, X_n), U(X_1, X_2, \dots, X_n)). \end{aligned}$$

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

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

Il résulte que

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

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

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

Donc,

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

implique

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

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

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

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

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

Ce qui est vérifié seulement lorsque

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

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

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

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

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 θ on a

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

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

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

Finalement

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

■

1.3 Information au sens de Fisher

Définition 12 Soit X_1, \dots, X_n un n -échantillon de X de loi P_θ . On suppose que P_θ 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 θ alors

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

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

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

De même

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

Ainsi

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

ce qui nous permet d'écrire

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

■

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

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

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

Proposition 4 Soient 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 θ alors

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

Démonstration La démonstration est similaire à la proposition précédente

Proposition 5 Soient 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 θ , alors

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

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

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

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

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

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

D'où

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

On applique l'inégalité de Cauchy-Schwarz

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

Du moment que

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

on obtient

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

Par conséquent

$$\text{Var}[T(X_1, X_2, \dots, X_n)] \geq \frac{[\frac{\partial}{\partial \theta} \mathbb{E}[T(X_1, X_2, \dots, X_n)]]^2}{I_n(\theta)}.$$

■

Remarque 3

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

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

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

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

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 θ , 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

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

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

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

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 θ est un paramètre fonctionnel si $\theta = T(F)$ où

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

\mathcal{F} : l'ensemble des fonctions de répartition.

Θ : 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) $E[X] = \int x dF$ est un paramètre fonctionnel tel que $T(F) = \int x dF$ est le fonctionnel statistique.
- (iii) $V[X] = \int x^2 dF - \left[\int x dF \right]^2$ est un paramètre fonctionnel tel que :

$$T(F) = \int x^2 dF - \left[\int x dF \right]^2$$

(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, \dots, X_n un n -échantillon de X de fonction de répartition F , on appelle fonction de répartition empirique la fonction \hat{F} définie par :

$$\begin{aligned} \hat{F}(x) &: \mathbb{R} \rightarrow [0, 1] \\ x \rightarrow \hat{F}(x) &= \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(x_i) \end{aligned}$$

Proposition 6 Soient X_1, X_2, \dots, 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 vérifier que $E[\hat{F}(x)] \rightarrow F(x)$ quand $n \rightarrow \infty$.

Par définition de la fonction de répartition :

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

Ceci implique :

$$\begin{aligned} E[\hat{F}(x)] &= \frac{1}{n} \sum_{i=1}^n E[I_{\{X_i \leq x\}}] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}(X_i \leq x) \\ &= \frac{n}{n} F(x) \\ &= F(x). \end{aligned}$$

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

Proposition 7 Soient X_1, X_2, \dots, 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 vérifier que :

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

En effet,

$$\begin{aligned} \mathbb{P} \left[\left| \hat{F}(x) - F(x) \right| > \epsilon \right] &= \mathbb{P} \left[\left| \hat{F}(x) - \mathbb{E} \left[\hat{F}(x) \right] + \mathbb{E} \left[\hat{F}(x) \right] - F(x) \right| > \epsilon \right] \\ &\leq \mathbb{P} \left[\left| \hat{F}(x) - \mathbb{E} \left[\hat{F}(x) \right] \right| > \frac{\epsilon}{2} \right] + \mathbb{P} \left[\left| \mathbb{E} \left[\hat{F}(x) \right] - F(x) \right| > \frac{\epsilon}{2} \right]. \end{aligned}$$

Pour que :

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

Il suffit de vérifier que :

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

et

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

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

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

Maintenant il reste de montrer que :

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

En effet, on considère la suite

$$x_n = \sqrt{\frac{\log n}{n}} \rightarrow 0 \text{ quand } n \rightarrow \infty$$

et on démontre qu'il existe ϵ_0 tel que :

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

Comme la suite $x_n \rightarrow 0$

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

Si $n > N$

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

En effet pour

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

on a

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

On applique l'inégalité de Bernstein :

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

Donc :

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

pour un ϵ_0 tel que $2\epsilon_0^2 > 1$,

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

Par conséquent :

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

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

Proposition 8 Soient X_1, X_2, \dots, 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 $V[\hat{F}(x)] \rightarrow 0$ car :

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

et

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

car

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

Pour ce faire, On re-ordonne les observations :

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

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 $\{0, 1, \dots, n\}$.

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

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

Donc :

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

$$E[nF(x)] = nF(x).$$

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

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, \dots, X_n un n -échantillon de X de fonction de répartition F , on appelle estimateur à noyau H l'estimateur défini par :

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

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

Théorème 2 Soient X_1, X_2, \dots, 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 $E[F_n(x)] \rightarrow F(x)$ quand $n \rightarrow \infty$.

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

On utilise l'intégration par parties, posons

$$\begin{aligned} U = H\left(\frac{x-z}{h_n}\right) &\Rightarrow dU = -\frac{1}{h_n} H'\left(\frac{x-z}{h_n}\right) dz \\ dV = f(z) dz &\Rightarrow V = F(z). \end{aligned}$$

D'où

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

On considère maintenant le changement de variable suivant :

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

Alors

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

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

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

$$x \leftrightarrow x - h_n t$$

avec :

$$x_0 \leftrightarrow x$$

donc on aura :

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

D'où

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

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

Proposition 9 *Soient X_1, X_2, \dots, 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 :

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

Soit $\epsilon > 0$ on a

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

Donc il suffit de vérifier que :

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

et

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

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

$$\begin{aligned} &\mathbb{P}\left[\left|F_n(x) - E[F_n(x)]\right| > \epsilon_0 \sqrt{\frac{\log n}{n}}\right] \\ &= \mathbb{P}\left[\left|\frac{1}{n} \sum_{i=1}^n \left[H\left(\frac{x - X_i}{h_n}\right) - E\left[H\left(\frac{x - X_i}{h_n}\right)\right]\right]\right| > \epsilon_0 \sqrt{\frac{\log n}{n}}\right] \\ &= \mathbb{P}\left[\left|\sum_{i=1}^n \left[H\left(\frac{x - X_i}{h_n}\right) - E\left[H\left(\frac{x - X_i}{h_n}\right)\right]\right]\right| > \epsilon_0 \sqrt{n \log n}\right] \end{aligned}$$

2.3. L'ESTIMATEUR À NOYAU DE LA FONCTION DE RÉPARTITION 27

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

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

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

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

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

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

donc :

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left[\left|\mathbb{E}[F_n(x)] - F(x)\right| > \frac{\epsilon}{2}\right] &= \sum_{n=1}^N \mathbb{P}\left[\left|\mathbb{E}[F_n(x)] - F(x)\right| > \frac{\epsilon}{2}\right] \\ &+ \underbrace{\sum_{n=N}^{\infty} \mathbb{P}\left[\left|\mathbb{E}[F_n(x)] - F(x)\right| > \frac{\epsilon}{2}\right]}_0 \\ &= \sum_{n=1}^N \mathbb{P}\left[\left|\mathbb{E}[F_n(x)] - F(x)\right| > \frac{\epsilon}{2}\right] < \infty. \end{aligned}$$

■

Proposition 10 Soient X_1, X_2, \dots, 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 montrer que F_n converge en moyenne quadratique vers F il faut montrer que :

$$\mathbb{E}\left[F_n(x) - F(x)\right]^2 \text{ tend vers } 0 \text{ quand } n \text{ tend vers } \infty$$

On a :

$$V[F_n(x)] = V[F_n(x) - F(x)] = E[F_n(x) - F(x)]^2 - \left(E[F_n(x) - F(x)]\right)^2$$

Donc il suffit de montrer que

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

Puisque le deuxième terme tend vers 0. En effet,

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

et

$$V\left[H\left(\frac{x - X_i}{h_n}\right)\right] = E\left[H^2\left(\frac{x - X_i}{h_n}\right)\right] - \left(E\left[H\left(\frac{x - X_i}{h_n}\right)\right]\right)^2$$

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

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

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

$$\begin{aligned} E\left[H^2\left(\frac{x - X_i}{h_n}\right)\right] &= 2F(x) \int_{-\infty}^{+\infty} H(t) H'(t) dt \\ &\quad - 2h_n \int_{-\infty}^{+\infty} t H(t) H'(t) f(\xi) dt, \quad \xi \in]x - h_n t, x[. \end{aligned}$$

Au passage à la limite quand n tend vers ∞ , $E\left[H^2\left(\frac{x - X_i}{h_n}\right)\right]$ tend vers

$$2F(x) \int_{-\infty}^{+\infty} H(t) H'(t) dt < \infty. \text{ Or, } E\left[H\left(\frac{x - X_i}{h_n}\right)\right] \text{ tend vers } F(x).$$

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

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, \dots, 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

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

.

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, \dots, 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 complètement on a

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

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, \dots, X_n un n -échantillon de X de densité f . On appelle estimateur à noyau pour la densité f la variable aléatoire définie par :

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

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 ∞ .

3.2 Biais de l'estimateur

Proposition 11 Soient X_1, X_2, \dots, 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

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

On considère maintenant le changement de variable suivant

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

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

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

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

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

3.3 Variance de l'estimateur

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

$$\int_{-\infty}^{+\infty} K^2(t) dt < +\infty \quad \text{et} \quad \int_{-\infty}^{+\infty} t K^2(t) dt < +\infty$$

on a

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

Démonstration On a

$$\begin{aligned} V[f_n(x)] &= V\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{n^2 h_n^2} \sum_{i=1}^n V\left[K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n^2} V\left[K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n^2} \left[E\left[K^2\left(\frac{x - X_i}{h_n}\right)\right] - \left(E\left[K\left(\frac{x - X_i}{h_n}\right)\right]\right)^2 \right] \\ &= \underbrace{\frac{1}{nh_n} \left[\frac{1}{h_n} E\left[K^2\left(\frac{x - X_i}{h_n}\right)\right] \right]}_A - \underbrace{\frac{1}{n} \left[\frac{1}{h_n} E\left[K\left(\frac{x - X_i}{h_n}\right)\right] \right]^2}_B \end{aligned}$$

D'une part on a

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

On considère le changement de variable suivant :

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

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

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 :

$$\begin{aligned} f(x - h_n t) &= f(x) + (x - h_n t - x)f'(\xi) \\ &= f(x) - h_n t f'(\xi). \end{aligned}$$

Il résulte que :

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

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

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

D'où $\frac{1}{h_n} \mathbb{E} \left[K \left(\frac{x - X_i}{h_n} \right) \right]$ tend vers $f(x)$ quand n tend vers ∞ . Finalement,

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

■

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 même démonstration que le cas de la fonction de répartition il suffit d'écrire

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

et

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

De même, les deux théorème précédents, montrent que les deux terms $V[f_n(x)]$ et $\mathbb{E} [f_n(x) - f(x)]^2$ 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*

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

et si

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

Alors,

$$(3.1) \quad \sup_{y \in (\alpha, \beta)} |\hat{f}(y) - f(y)| \rightarrow 0 \text{ en probabilité}$$

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 indépendamment de x $E[f_n(x)] - f(x)$ tend vers 0 donc, il suffit de démontrer

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

Pour cela on démontre d'abord que

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

L'idée de cette démonstration est de recouvrir le compact (α, β) par un nombre fini d'intervalles B_k de longueurs égales l_n . On note τ_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

$$\begin{aligned} \mathbb{P} \left(\sup_{x \in (\alpha, \beta)} |\mathbb{E}\hat{f}(x) - \hat{f}(x)| > \epsilon \right) &\leq \underbrace{\mathbb{P} \left[\left(\sup_{x \in (\alpha, \beta)} |\mathbb{E}\hat{f}(x) - \mathbb{E}\hat{f}(t_x)| > \frac{\epsilon}{3} \right) \right]}_{T(1)} + \\ &\quad \underbrace{\mathbb{P} \left[\left(\sup_{x \in (\alpha, \beta)} |\mathbb{E}\hat{f}(t_x) - \hat{f}(t_x)| > \frac{\epsilon}{3} \right) \right]}_{T(2)} + \underbrace{\mathbb{P} \left[\left(\sup_{x \in (\alpha, \beta)} |\hat{f}(t_x) - \hat{f}(x)| > \frac{\epsilon}{3} \right) \right]}_{T(3)}. \end{aligned}$$

Pour T(2)

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

Il suffit de calculer :

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

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

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

où

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

Pour appliquer l'inégalité de Hoeffding au variable Δ_i il faut évaluer la moyenne d'ordre deux de Δ_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 :

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

où

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

Donc

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

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

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

Il vient que

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

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

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

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

donc

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

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

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

Ceci implique

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

Pour T(3)

On a :

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

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

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

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 (α, β) un compcat de \mathbb{R} . Le mode θ est le point le plus probable. C'est à dire θ est tel que

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

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

$$(3.3) \quad f_n(\hat{\theta}) = \sup_{y \in (\alpha, \beta)} f_n(y).$$

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

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

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

$$(3.4) \quad f(\hat{\theta}) = f(\theta) + (\hat{\theta} - \theta)f'(\xi)$$

où ξ est entre $\hat{\theta}$ et θ , il est donc nécessairement dans le compact (α, β) . Ainsi : on obtient de la formule 3.4 :

$$|f(\hat{\theta}) - f(\theta)| = |(\hat{\theta} - \theta)f'(\xi)|$$

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

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

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

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

Ce qui donne ;

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

En appliquant le théorème précédent on obtient

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

Ainsi

$$\hat{\theta} - \theta \rightarrow 0 \quad \text{inprobabilit}$$

■

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

$$\text{MSE}_g(x) = \text{E} \left[g(x) - \hat{g}(x) \right]^2$$

ou l'erreur quadratique intégrée :

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

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-paramétrique

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

4.1 La régression non paramétrique

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

$$(4.1) \quad Y = r(X) + \varepsilon$$

où ε est une variable aléatoire réelle centrée et indépendante 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

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

Définition 27 Soit $\{(X_i, Y_i), i = 1, \dots, n\}$ 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

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

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

Remarque 8

Cette estimateur peut être exprimé par

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

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*

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

(H5) Le paramètre de lissage est tel que :

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

on a

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

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 :

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

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

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

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

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

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

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

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

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

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

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

De même, on utilise le développement de Taylor

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

Ainsi,

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

■

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

$$(4.8) \quad \widehat{g}(x) - \mathbb{E}\widehat{g}(x) = \mathcal{O}\left(\sqrt{\frac{\log n}{nh}}\right)$$

$$(4.9) \quad \widehat{f}(x) - \mathbb{E}\widehat{f}(x) = \mathcal{O}\left(\sqrt{\frac{\log n}{nh}}\right)$$

Démonstration du lemme (2)

On a

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

on pose

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

Comme K et Y sont bornés alors

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

Alors il suffit de calculer

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

et

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

Où

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

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

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

On considère le même changement des variables $z = \frac{(x-u)}{h}$ donc $dz = \frac{-du}{h}$ 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 ϕ 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 $\delta_2 > 0$ on a :

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

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'après la série de Riemann.

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

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

■

Lemme 3 *Sous les hypothèses du lemme 2*

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

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

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

Donc

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

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

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 :*

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

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

on a

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

où

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

et

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

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

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

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

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

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

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

Ceci implique que,

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

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

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

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

$$\begin{aligned} \text{Var}[\widehat{r}(x)] &= (E\widehat{f}(x))^{-2} \text{Var}[\widehat{g}(x)] - 2(E\widehat{f}(x))^{-3} (E\widehat{g}(x)) \text{Cov}(\widehat{g}(x), \widehat{f}(x)) \\ &\quad + (E\widehat{f}(x))^{-4} (E\widehat{g}(x))^2 \text{Var}(\widehat{f}(x)) + o\left(\frac{1}{nh_n}\right). \end{aligned}$$

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

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

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

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

Démonstration du Lemme 4 On a

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

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

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

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

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

Pour la deuxième équation

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

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

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

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,

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

■

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

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

Démonstration du Corollaire 1 On a

$$\begin{aligned} & (E\widehat{f}(x))^{-1}E[\widehat{g}(x)] - r(x) \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2)}{f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)} - r(x) \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2) - r(x) \left(f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2) \right)}{f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)} \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2) - r(x) \left(f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2) \right)}{f(x) + O(h_n^2)} \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) - r(x) \left(f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) \right)}{f(x)} + O(h_n^2) \\ &= \frac{h_n^2}{2} \int t^2 K(t) \left(\frac{g^{(2)}(x) - r(x)f^{(2)}(x)}{f(x)} \right) + O(h_n^2) \end{aligned}$$

■

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

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

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

Démonstration du Lemme 5 On a

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

Il est clair que

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

D'où

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

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

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

■

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

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

En vertu du calcul précédent, (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,

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

On considère le changement des variables usuelle $t = \frac{x - z}{h_n}$ 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

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

Il est clair que

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

D'où

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

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

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

Chapitre 5

Exercices

5.1 Exercices sur l'estimation paramétrique

Exercice 1

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

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

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 $T = \frac{2}{n} (X_1 + X_2 + \dots + X_n)$ est convergente

Exercice 3

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

Exercice 4

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

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

- 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

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

Donner l'estimateur du maximum de vraisemblance de θ

Exercice 6

Soit X une variable de densité

$$f(x) = \begin{cases} \theta x^2 & x \in [-1, 0] \\ 1 - \theta x^2 & x \in [0, 1] \\ 0 & x \notin [-1, 1] \end{cases}$$

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

Exercice 7

Soit X_1, \dots, X_n un échantillon de taille n de densité

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

- Déterminer 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 quantité 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 θ 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_θ la loi déduite de P par la translation θ .

Déterminer l'estimateur de maximum de vraisemblance de θ au vu d'un échantillon de taille n de la loi P_θ ; 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 σ sont inconnus.

k étant un réel positif, trouver un estimateur sans biais σ^{2k}

Exercice 11

Soit X_1, \dots, X_n un n - échantillon de loi P_θ . On considère G_1 et G_2 deux estimateurs, sans biais pour le paramètre θ de carré sommable.

- a) Déterminer une forme linéaire de G_1 et G_2 qui soit un estimateur sans biais de θ et qui ait la plus petite variance possible.
- b) Que devient le résultat de la question précédente 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 θ par la méthode du maximum de vraisemblance ; quelles sont les qualités de l'estimateur obtenu
- c) θ 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_θ définie par

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

où θ est un paramètre positif inconnu.

- a) Déterminer l'estimateur de maximum de vraisemblance de θ
- b) Introduire la variable aléatoire $Z = -\log(1-X)$ et explique le résultat précédent.
- c) L'estimateur déterminé à la question précédente 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 σ 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 Fourchette 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 σ 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 σ^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 paramétrique

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 convergence 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 un noyau à support compact de classe C^j , trouver un estimateur à noyau de $f^{(j)}(x)$.
2. Etudier sa convergence en moyenne quadratique
3. Etudier sa convergence 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 θ_n pour θ
3. Montrer que

$$|f(\theta) - f(\theta_n)| \leq 2 \sup_{x \in S} |f(x) - f_n(x)|$$

4. En utilisant l'inversibilité à gauche et à droite de la densité f au voisinage de θ , montrer que l'estimateur θ_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 :

$$\sum_n P((1 - F_n(x)) < \delta) < \infty.$$

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 ψ une fonction mesurable bornée. On considère la fonction de régression de $\psi(Y)$, sachant X , définie par

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

et on suppose que la densité de la variable explicative X et la fonction r_ψ vérifient la condition suivante :

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

où ϕ désigne indifféremment f ou r_ψ .

1. Estimer la fonction r_ψ par la méthode du noyau
2. Montrer que, si :

$$(a) \quad \lim_{n \rightarrow \infty} h_n = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{nh_n}{\log n} = \infty,$$

(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)

Coûts d'un paquet=50

Coûts 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) Coûts 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, λ) et la loi de la durée d'incubation conditionnellement à un facteur de

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