République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

SAOUTHI Samia

Titre :

Estimation non paramétrique des quantiles

Membres du Comité d'Examen :

Pr.	CHERFAOUI Mouloud	Professeur	UMKB	Président
Dr.	KHEIREDDINE Souraya	M.C.B.	UMKB	Encadreur
Dr.	CHINE Amel	M.C.B.	UMKB	Examinateur

Juin 2023

Dédicace

Je dédie ce travail

à *mes chers père et mère, pour tous les sacrifices qu'ils ont consentis pour me permettre de pour suivre mes études dans les meilleures conditions possibles et pour n'avoir jamais cessé de m'encourager tout au long de mes années d'étude.*

à mes très chers frères et soeurs, à tout ma famille

à *tous les membres de ma promotion*

et tous mes professeurs.

Remerciements

*Mes remerciements vont tout premièrement à **ALLAH** le tout puissant pour la volonté, la santé, et la patience qu'il m'a donné durant ces longue années d'étude et le courage pour terminer ce modeste travail.*

Je tiens à remercier mon encadreur Dr. Kheireddine Souraya pour l'aide et le temps qu'il a vaulu nous consacrer et qu'on ne remerciera jamais assez pour ses lectures, ses conseils et sapartience.

Je remercie également les membres du Jury Pr. Cherfaoui Mouloud et Dr. Chine Amel pour avoir accepté d'évaluer, de juger ce modeste travail et de l'enrichir par leurs propositions.

Nous remercions sincèrement tous ceux qui ont contribué de près ou de lion, à la réaliation de ce project..

Résumé

Dans ce mémoire nous étudions l'estimateur non paramétrique de la fonction des quantiles. La construction de l'estimateur est basé sur l'utilisation d'une densité K appelée noyau et d'un paramètre de lissage h. Nous rappelons les propriétés asymptotique de l'estimateur. Nous parlons aussi du choix de noyau et de paramètre de lissage.

.

Table des matières

Liste des tableaux

Table des figures

Introduction

L'estimation des quantiles est une technique statistique essentielle dans de nombreux domaines tels que la finance, l'analyse de données, la modélisation en vironnementale, la fiabilité des systèmes, entre autres. Les méthodes d'estimation non paramétriques des quantiles offrent des avantages significatifs par rapport aux modèles paramétriques traditionnels, en particulier lorsque la distribution des données est non normale ou inconnue. ces méthodes sont basées sur des algorithmes de calcul direct qui n'exigent pas de suppositions sur la fonction de distribution sous-jacente.

L'utilisation de ces méthodes peut conduire à une meilleure précision et à une plus grande robustesse des estimations de quantiles, ce qui peut s'avérer crucial dans les applications où la fiabilité est facteur important. Cet article vise à presenter les principles techniques d'estimation non paramétrique des quantiles et à discuter de leurs avantages et des conditions d'utilisation optimales.

Dans ce mémoire, on s'intéresse à l'estimation à noyau de la fonction de quantile. L'estimateur à noyau de la fonction de quantile est basé sur l'estimateur de la fonction de répartition introdiut par Nadaraya (1964)[16]. *De nombreux chercheurs ont étudie cet estimateur et ces propriétés, comme Falk (1984)[6] qui a montré que la performance asymptotique de l'estimateur à noyau est meilleur par rapport à cella de l'estimateur empirique de la même fonction Yang (1985)[24] à établi la normalisté asymptotique et la consistance en moyenne quadratique du même estimateur. Sheather et Marron (1990)[21] ont donné l'expresion de l'erreur moyenne quadratique*

(MSE) *et d'autres comme Yamato (1973), Parzen (1979)[18], Azzalini (1981)[1], Harrell et Davis (1982)[8].*

Ce mémoire est composé de deux chapitres. Le premier chapitre est consacré aux quantiles. Tout d'abord, on a donné les définitions des quantiles, la fonction des quantile, quantile d'ordre p *et l'estimation des quantiles. Et après nous étudions les statistiques d'ordre, fonction de répartition empirique, nous présentons certaines modes de convergence de variable aléatoire.*

Le second chapitre nous mettons l'accent sur l'estimateur des quantiles, puis on a donné l'expression de l'estimateur à noyau de cette fonction, ensuite on a étudié les propriétés asymptotiques de l'estimateurs.

Chapitre 1

Estimation fonctionnelle

Dans ce chapitre, nous avons commencé par donner la définition des statistiques d'ordre et la fonction de répartition empirique. Dans la deuxième section nous présentons les définitions des quantiles, la fonction des quantile, quantile d'ordre p, puis nous présentons certaines modes de convergence de variable aléatoire.

1.1 Statistique d'ordre

Soit $(X_1, ..., X_n)$ une suite de variable aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition $F \quad (F(x) = P(X_n \leq x)\,,\ \text{pour } x \in \mathbb{R})$.Soit S_n l'ensemble des permutation de $\{1, ..., n\}$.

1.1.1 Définition de la statistique d'ordre

En classant les variables aléatoires $X_1, X_2, ..., X_n$ par ordre croissant, on obtient :

$$X_{1,n} \leq X_{2,n} ... \leq X_{n,n}.$$

on appelle $X_{1,n}, X_{2,n}, ..., X_{n,n}$ les statistiques d'ordre associées à $(X_1, X_2, ..., X_n)$, et on désigne par $X_{i,n}$ la $i^{ième}$ statistique d'ordre.

Intéressons nous plus praticulièrement au comportement d'extrêmes corespondant à $i = 1$ et $i = n$, i.e,

$$X_{1,n} = \min\left(X_1, X_2, ..., X_n\right),$$
$$X_{n,n} = \max\left(X_1, X_2, ..., X_n\right).$$

Lois des valeurs extrême $X_{1,n}$ et $X_{n,n}$

L'un des rôles importants des statistiques d'ordre en statistique est la détection des points aberrants dans un échantillon, on s'intéresse aux deux lois suivantes :

Loi de $X_{1,n}$

$$F_1(x) = P\left(X_{1,n} \leq x\right) = \sum_{i=1}^{n} C_n^i F^i(x)\left(1 - F(x)\right)^{n-i} = 1 - C_n^0\left(1 - F(x)\right)^n,$$

d'où

$$F_1(x) = 1 - \left(1 - F(x)\right)^n, \text{ et } f_1(x) = nf(x)\left(1 - F(x)\right)^{n-1}.$$

Loi de $X_{n,n}$

$$F_n(x) = P\left(X_{n,n} \leq x\right) = F^n(x),$$

d'où

$$f_n(x) = nf(x)\,F^{n-1}(x).$$

En utilisant la propriété d'indépendance des variables aléatoires $X_1, ..., X_n$ nous en déduisons que les lois de maximum $X_{n,n}$ et de minimum $X_{1,n}$ de la statistique d'ordre associées à l'échantillon $X_1, ..., X_n$ sont :

$$\begin{aligned}
F_{X_{n,n}}(x) &= P(X_{n,n} \leq x) = P(\max(X_1, ..., X_n) \leq x) \\
&= P\left[\bigcap_{i=1}^{n}(X_i \leq x)\right] \\
&= \prod_{i=1}^{n} P(X_i \leq x) \\
&= \prod_{i=1}^{n} F(x)
\end{aligned}$$

Alors

$$F_{X_{n,n}}(x) = [F_X(x)]^n .$$

Puis

$$\begin{aligned}
F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) = 1 - P(X_{1,n} > x) \\
&= P(\min(X_1, ..., X_n) > x) \\
&= 1 - P\left[\bigcap_{i=1}(X_i > x)\right] \\
&= 1 - \prod_{i=1}^{n}[1 - P(X_i \leq x)] \\
&= 1 - \prod_{i=1}^{n}[1 - F_X(x)] .
\end{aligned}$$

D'où on déduit :

$$F_{X_{1,n}}(x) = 1 - [1 - F_X(x)]^n .$$

De ces résultats, nous en tirons la conclusion que le maximum $F_{X_{n,n}}$ est une variable aléatoire dont la fonction de répartition correspond à F^n.

1.2 La fonction de répartition empirique

On appelle fonction de répartition empirique associée à un échantillon $X_1, ..., X_n$ la fonction F_n définie par :

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{\{X_k \leq x\}} = \frac{1}{n} Card\{k \in \{1,...,n\} : X_k \leq x\}$$

$$= \begin{cases} 0 & \text{si} \quad x \leq X_{1,n}, \\ \frac{1}{n} & \text{si} \quad X_{1,n} < x < X_{2,n}, \\ \frac{2}{n} & \text{si} \quad X_{2,n} < x < X_{3,n}, \\ . & \quad ... \\ . & \quad ... \\ \frac{k}{n} & \text{si} \quad X_{k,n} < x < X_{k+1,n}, \\ . & \quad ... \\ 1 & \text{si} \quad x \geq X_{n,n} \end{cases}$$

où $I_{\{X_k \leq x\}}$: c'est la fonction indicatrice, telle que :

$$I_{\{X_k \leq x\}} = \begin{cases} 0 & \text{si} \quad X_k > x, \\ 1 & \text{si} \quad X_k \leq x. \end{cases}$$

Convergence de la fonction de répartition empirique

Soit $(X_n, n \in N)$ une suite de variable aléatoire réelles, indépendantes de méme loi F alors :

$$\lim_{n \longrightarrow +\infty} \sup_{x \in \mathbb{R}} \mid F_n(x) - F(x) \mid = 0,$$

Convergence presque sûre.

1.3 Les Quantiles

Les quantiles d'une variable aléatoire univariée discrète (entière) ou continue (réelle) sont les valeurs que prend la variable pour des valeurs de probabilité p $(0 < p < 1)$ sous le quantile considéré. On les appelles encore fractile, et ce sont les valeurs réciproques de la fonction de répartion de la loi de probabilité considérée.

Les quantiles d'un échantillon statistique des nombres sont des valeurs remarquables permettant de diviser le jeu de ces données ordonnées c'est à dire triées en intervalle consécutifs contenant le même nombre de données.

Définition 1.3.1 ***(Définition générale des quantiles)*** *Soit X une variable aléatoire réelle. Pour tout nombre p strictement compris entre 0 et 1, on appelle p-**quantile** de X, ou de la loi de X, tout nombre x vérifiant les deux inégalités :*

$$\begin{cases} P\left[X < x\right] \leq p \\ P\left[X \leq x\right] \geq p \end{cases}.$$

L'ensemble des valeurs de x pour lesquelles $P\left[X < x\right] \leq p$ est un intervalle de la forme $\left]-\infty, M\right]$, contenant sa borne supérieure M. Ce nombre M est un quantile de niveau p de X parce que, pour tout $x > M$, ona $P\left[X \leq x\right] \geq p$, ce qui entraîne $P\left[X \leq M\right] \geq p$, par continuité à droite de la fonction de répartition F_X.

L'ensemble des valeurs de x pour lesquelles $P\left[X \leq x\right] \geq p$ est un intervalle de la forme $\left[m, +\infty\right[$, contenant sa borne inférieure m.ce nombre est un quantile de niveau p de X parce que, pour tout $x < m$, ona $P\left[X < x\right] \geq p$, ce qui entraîne $P\left[X < m\right] \leq p$, par continuité à gauche de la fonction $x \longrightarrow P\left[X < x\right]$.

Par conséquent, l'ensemble de quantiles de niveau p de la variable X est le segment $[m, M]$, qu'on notera désormais $\left[Q_X^-(p), Q_X^+(p)\right]$, où $Q_X^-(p)$ désigne donc le plus petit quantile de niveau p de X et $Q_X^+(p)$ le plus grand :

$$Q_X^-(p) = \inf\{x \in \mathbb{R}; P[X \leq x] \geq p\}$$
$$Q_X^+(p) = \sup\{x \in \mathbb{R}; P[X < x] \leq p\}$$

Exemple 1.3.1 *On suppose ici que la variable aléatoire X est discrète et ne prend que trois valeurs* 0, 1 *et* 2,*avec des probabilités respectives* $\frac{1}{4}, \frac{1}{4}$ *et* $\frac{1}{2}$. *Les quantiles d'ordre* $\frac{1}{4}$ *de X sont les éléments du segment* $[0,1]$ *et les médianes de X les éléments du segment* $[1,2]$.

Les autres quantiles sont uniques :

$$Q_X^-(p) = Q_X^+(p) = \begin{cases} 0 & \text{si } 0 < p < \frac{1}{4} \\ 1 & \text{si } \frac{1}{4} < p < \frac{1}{2} \\ 2 & \text{si } \frac{1}{2} < p < 1 \end{cases}$$

On observera pour l'anecdote que le troisième quantile est unique alors que le premier ne l'est pas. Ces choses-là arrivent.

1.3.1 La fonction des quantiles

Définition 1.3.2 *La fonction de quantile est définie par :*

$$Q(p) := F^{-1}(p) = \inf\{x : F(x) \geq p\}, 0 < p < 1.$$

1.3.2 Quantile d'ordre p

Soit X une variable aléatoire réelle de loi F, soit $p \in]0,1[$ .On appelle p-quantile de F toute réelle vérifiant :

$$Q_X(p) = x_p = F^{-1}(p) = \inf \{x \in \mathbb{R} : F(x) \geq p\}.$$

S'il existe un intervalle $[a, b[$ contenue dans $\mathbb{R}$ telle que :

$$\forall x \in [a, b[, F(x) = p.$$

Remarque 1.3.1 *pour certaines valeurs de p, il ya des noms spéciaux aux quantiles :*

1. La **médiane** est le quantile d'ordre $p = \frac{1}{2}$.
2. Le **premier quartile** est le quantile d'odre $p = \frac{1}{4}$.
3. Le **troisième quartile** est le quantile d'ordre $p = \frac{3}{4}$.
4. La $i^{ième}$ **quintile,** $1 \leq i \leq 4$ est le quantile d'ordre $p = \frac{i}{5}$.
5. La $i^{ième}$ **décile,** $1 \leq i \leq 9$ est le quantile d'ordre $p = \frac{i}{10}$.
6. La $i^{ième}$ **vingtile,** $1 \leq i \leq 19$ est le quantile d'ordre $p = \frac{i}{20}$.
7. La $i^{ième}$ **centile,** $1 \leq i \leq 99$ est le quantile d'ordre $p = \frac{i}{100}$.

1.3.3 Estimation des quantiles

L'estimation des quantiles ou l'estimation de la fonction des quantiles est un des problèmes fondamentaux en statistique, et représente un enjeu important pour les application.Pour estimer la fonction de quantile nous recensons dans la littérature deux types d'approches que sont : "l'approche paramétrique et l'approche non-paramétrique".

- **L'approche paramétrique :** On cherche à sa faire une idée de la valeur inconnue d'un paramètre, qui détermine la loi de probabilité d'une variable aléatoire.
- **L'approche non-paramétrique :** On cherche à estimer des fonctions qui sont inconnues lorsqu'aucune information ne précise sur la forme et la classe de la vraie fonction n'est disponible.

1.3.4 L'approche paramétrique

Supposons que la fonction F_X est continue et appartient à une famille de distribution paramétrique $F = \left\{F_\theta, \theta \in \Theta \in \mathbb{R}^k\right\}$. L'idée d'estimation paramétrique est de supposer que toute quantilé statistique peut être vu en fonction de θ et dons un estimateur naturel du quantile est obtenu en substituant un estimateur de paramêtre $\hat{\theta}$ pour θ, ensuite étant donnée par :

$$Q_X(p) = F_\theta^{-1}(p).$$

donc l'estimateur naturel est :

$$\hat{Q}_X(p) = F_{\hat{\theta}}^{-1}(p).$$

Cette méthode est pratique, car il existe plusieurs téchniques pour l'obtention $\hat{\theta}$ (méthodes de moments, maximum de vrisemblance,...), mais le choix de F est cruciale. Une idée naturelle(qui peut être trouvé dans les modèles financiers classiques) est de supposer la loi gaussienne :si $X \sim N(\mu, \sigma)$, le quantile $Q_X(p)$ est donné par :

$$Q_X(p) = \mu + \sigma\Phi^{-1}(p).$$

où Φ^{-1} est l'inverse d'une distribution normale.

Définition 1.3.3 *Soit $X_1, ...X_n$ une suite de n variables aléatoires indépendantes distribuées normalement avec la fonction de distribution Φ, l'estimateur paramétrique(Gaussien) de la quantile d'ordre p est :*

$$\hat{Q}_X\left(p\right) = \hat{\mu} + \hat{\sigma}\Phi^{-1}\left(p\right).$$

où

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \qquad \hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{\mu}\right)^2.$$

En effet, le modèle paramétrique est modèle gaussien ne correspond pas trés bien, il est toujours possible d'utiliser l'approximation gaussien. Si la variance est finie, $\left(X - E\left(X\right)\right) \diagup \sigma$ pourrait être plus proche de la distribution gaussienne, et donc, considérent l'approximation dite **Cornish-Fisher,** i.e :

$$Q_X\left(P\right) \sim E\left(X\right) + \sigma\hat{z}_p,$$

$$\hat{z}_p = \Phi^{-1}\left(p\right) + \frac{\hat{\zeta}_1}{6}\left(\left[\Phi^{-1}\left(p\right)\right]^2 - 1\right) + \frac{\hat{\zeta}_2}{24}\left(\left(\Phi^{-1}\left(p\right)\right)^3 - 3\Phi^{-1}\left(p\right)\right) + \frac{\hat{\zeta}_1^2}{36}\left(2\left(\Phi^{-1}\left(p\right)\right)^3 - 5\Phi^{-1}\left(p\right)\right),$$

où $\hat{\zeta}_1$ est l'estimateur naturel de paramètre d'asymétrie ζ_1 de X, et $\hat{\zeta}_2$ est l'estimateur naturel de paramètre d'aplatissement ζ_2 de X, c'est à dire :

$$\hat{\zeta}_1 = \frac{\sqrt{n\left(n-1\right)}}{n-2}\frac{\sqrt{n}\sum\limits_{i=1}^{n}\left(X_i - \hat{\mu}\right)^3}{\left(\sum\limits_{i=1}^{n}\left(X_i - \hat{\mu}\right)^2\right)^{\frac{3}{2}}},$$

et

$$\hat{\zeta}_2 = \frac{n-1}{(n-2)(n-3)} \left((n+1)\, \hat{\zeta}_2' + 6 \right),$$

où

$$\hat{\zeta}_2' = \frac{n \sum\limits_{i=1}^{n} (X_i - \hat{\mu})^4}{\left(\sum\limits_{i=1}^{n} (X_i - \hat{\mu})^2 \right)^2} - 3.$$

Définition 1.3.4 ***(L'estimation de Cornish-Fisher)***

Soit $X_1, ..., X_n$ un échantillon, l'estimation de Cornish-fisher du quantile d'ordre p est :

$$\hat{Q}_n(p) = \hat{\mu} + \sigma \hat{z}_p.$$

1.3.5 L'approche non-paramétrique

Dans la littérateur, plusieurs estimateurs de la fonction des quantiles, nous citons par exemple :

Quantile empirique

La fonction quantile empirique est donnée par la définition suivante :

Définition 1.3.5 *La fonction quantile empirique de l'échantillon $X_1, ..., X_n$ est donnée par :*

$$Q_n(p) = F_n^{-1}(p) = \inf \{ x \in \mathbb{R} : F_n(x) \geq p \}, \text{ avec } 0 < p < 1.$$

où $F_n(x)$ est la fonction de distribution empirique.

Définition 1.3.6 *Soit $(X_1, ..., X_n)$ un n échantillon issu d'une distribution F et $X_{1,n}, ..., X_{n,n}$ les statistques d'ordre associées. Soit $p \in \,]0,1[$. La statistique d'ordre*

$X_{([np]+1),n}$ *s'appelle le quantile empirique d'ordre* p *de l'échantillon.*

$$X_{([np]+1),n} = Q_n(p) = \inf\{x \in \mathbb{R} : F_n(x) \geq p\},$$

où $[np]$ désigne la partie entière de np. Soit $p_k = \frac{k}{(n+1)}$ et $q_k = 1 - p_k$, si nous utilisons $X_{k,n}$ pour estimer le quantile d'ordre p_k, alors le biais asymptotique et la variance sont :

$$ABiais\{X_{k,n}\} = \frac{p_k q_k Q''(p_k)}{2(n+2)} + \frac{p_k q_k}{(n+2)^2}\left\{\frac{1}{3}(q_k - p_k) Q'''_K + \frac{1}{8} Q''''_K\right\},$$

et

$$AVar\{X_{k,n}\} = \frac{p_k q_k}{(n+2)} Q'^2_K + \frac{p_k q_k}{(n+2)^2}\left\{2(q_k - p_k) Q'_K Q''_K + p_k q_k \left(Q'_K Q'''_K + \frac{1}{2} Q''_k\right)\right\}.$$

L'erreur quadratique moyenne asymptotique de $X_{k,n}$ doit être :

$$AMSE\{X_{k,n}\} = (ABiais\{X_{k,n}\})^2 + AVar\{X_{k,n}\}.$$

Remarque 1.3.2 *On peut représenter la fonction des quantiles emprique en fonction des statistique d'ordre comme suit :*

$$Q_n(p) = X_{k,n} \qquad \text{avec } \frac{k-1}{n} \leq p \leq \frac{k}{n}.$$

1.3.6 Noyaux

La définition du noyau est donnée par la définition suivante

Définition 1.3.7 *Soit $k : \mathbb{R} \longrightarrow \mathbb{R}$, on dit que k est un noyau si et seulement si :*

$$\int k(u)\, du = 1.$$

1. k est dit positive si $k(u) \geq 0 \ \forall u$.
2. k est symétrique si $k(u) = k(-u) \ \forall u$.

Exemples de noyaux usuels

Les noyaux les plus couramment utilisés en pratique sont (voire, Silverman, 2006) :

Noyau	**Fonction**	**Demaine**
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}$	$\mathbb{R}$
Triangulaire	$(1- \mid u \mid)$	$[-1, 1]$
d'Epanechnikov	$\frac{3}{4}(1-u^2)$	$[-1, 1]$
Biweight	$\frac{15}{16}(1-u^2)^2$	$[-1, 1]$

TAB. 1.1 – Quelques noyaux classiques.

La figure des ces noyaux sont présentées ci-dessous :

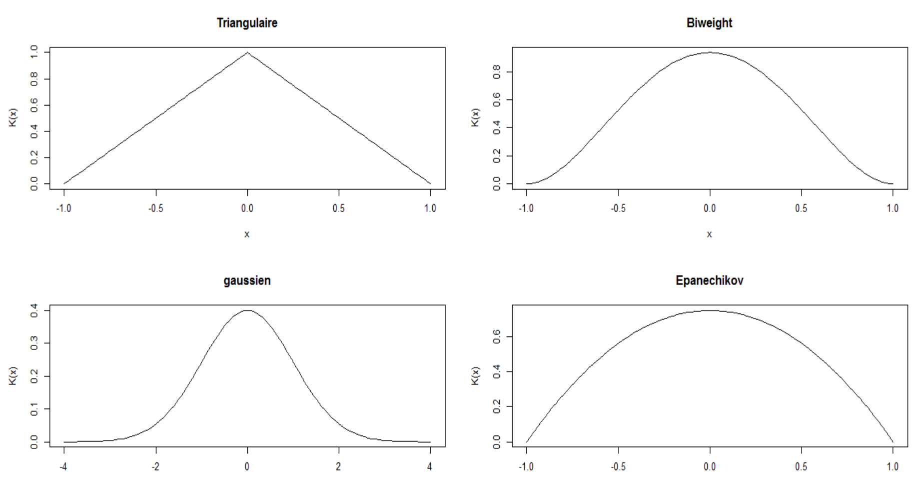

FIG. 1.1 – Allures des noyaux : Triangulaire, Biweight, Gaussien et Epanechnikov.

1.4 Théorèmes de convergences de variables aléatoires

Dans ce qui suit, nous présentons certaines modes de convergence de variable aléatoire :

1. **Convergence presque sûre :** Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$, définie sur $(\Omega, \mathcal{A}, P)$, converge presque sûrement (p.s.) vers la variable aléatoire X, définie sur $(\Omega, \mathcal{A}, P)$, si

$$P\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1$$

   Dans ce cas, on note $\lim\limits_{n\to\infty} X_n = X$ p.s. ou $X_n \to X$ p.s. lorsque $n \to \infty$.

2. **Convergence en probabilité :** Soient $X_n, n \in \mathbb{N}$, X, des variables aléatoires réelles sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X_n converge en probabilité vers X, et on note $X_n \overset{P}{\to} X$, ou $\lim\limits_{n\to\infty} X_n = X$ en probabilité, ou $P-\lim\limits_{n\to\infty} X_n = X$, si pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{|X_n - X| \geq \varepsilon\right\} = 0$$

3. **Convergence dans** L^p : Soient, $X_n, n \in \mathbb{N}$, X, des variables aléatoires réelles dans $L^p(\Omega, \mathcal{A}, \mathbb{P}), 0 < p < \infty$. On dit que X_n converge vers X dans L^p si $\lim\limits_{n\to\infty} \|X_n - X\|_p = 0$, ou de façon équivalente, $\lim\limits_{n\to\infty} \mathbb{E}\left(|X_n - X|_p\right) = 0$.

4. **Convergence en loi :** On dit que la suite de v.a. (X_n), de fonction de répartition F_n, converge en loi vers une v.a. X de fonction de répartition F, si la suite $(F_n(x))$ converge vers $F(x)$ en tout point x où F est continue : $X_n \xrightarrow{\mathcal{L}} X$, quand $n \longrightarrow \infty$.

5. **Convergence en moyenne quadratique :** Une suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ converge en moyenne quadratique vers une v.a.r. X si

$$\lim_{n\to\infty}\left(\mathbb{E}\left(X_n - X\right)^2\right) = 0,$$

et on note dans ce cas : $X_n \overset{m.q}{\to} X$.

Théorème de la limite centrale : Si $X_1, X_2, ..., X_n$ est une suit de *v.a. i.i.d.* d'espérance $\mu < \infty$ et de variance $\sigma^2 < \infty$, alors

$$\sqrt{n}\left(\bar{X} - \mu\right)/\sigma \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,1\right), \quad \text{quand} \quad n \longrightarrow \infty, \qquad \text{où } \bar{X} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i.$$

Théorème (Lois des grands nombres) : Si $(X_1, X_2, ..., X_n)$ est un échantillon provenant d'une v.a. X telle que $\mathbb{E}\left|X\right| < \infty$, alors :

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}\left(X\right) \quad \text{quand } n \longrightarrow \infty, \qquad \text{(loi faible)}$$

$$\bar{X}_n \xrightarrow{\mathbb{P}s.} \mathbb{E}\left(X\right) \quad \text{quand } n \longrightarrow \infty, \qquad \text{(loi forte).}$$

Définition (Biais d'un estimateur) : Un estimateur $\hat{\theta}_n$ de θ est dite sans biais si pour tout $\theta \in \Theta$ et tout entier positif n : $\mathbb{E}\left(\hat{\theta}_n\right) = \theta$.
De même, $\hat{\theta}_n$ est dite asymptotiquement sans biais si pour tout $\theta \in \Theta$:

$$\mathbb{E}\left(\hat{\theta}_n\right) \longrightarrow \theta \text{ quand } n \longrightarrow \infty.$$

La quantité : $\mathbb{E}\left(\hat{\theta}_n\right) - \theta$, est appelée le biais de l'estimateur $\hat{\theta}_n$.

Définition ($o_p(1)$ *et* $O_p(1)$) : La notation $o_p(1)$ segnifie qu'une suite de v.a's convergent vers 0 en probabilité. La notation $O_p(1)$ désigne une séquence qui est bornée en probabilité. Plus généralement, pour une suite donnée de v.a. R_n :

$$X_n = o_p(1) \Longleftrightarrow X_n = R_n Y_n \quad \text{et} \quad Y_n \xrightarrow{\mathbb{P}} 0,$$

$$X_n = O_p(1) \Longleftrightarrow X_n = R_n Y_n \quad \text{et} \quad Y_n = O_p(1).$$

Ces quantités vérifies les assertions :

$$\begin{aligned} &o_p(1) + O_p(1) = O_p(1), && o_p(1)\,O_p(1) = o_p(1), \\ &(1 + o_p(1))^{-1} = O_p(1), && o_p(O_p(1)) = o_p(1), \\ &o_p(R_n) = R_n o_p(1), && O_p(R_n) = R_n O_p(1). \end{aligned}$$

Chapitre 2

Estimateur à noyau des quantiles

Dans ce chapitre, on a étudié l'estimation par la méthode à noyau des quantiles et les propriétés asymptotiques de l'estimateur, le biais, la variance, l'erreur quadratique moyenne, et le paramètre de lissage optimal.

Définition 2.0.1 *On observe $X_1, X_2, ..., X_n$ variables aléatoires rélles indépendant et identiquement distribué (i.i.d) de fonction de répartition :*

$$F : x \longrightarrow F(x) = P(X_1 \leq X)$$

L'estimateur de la fonction de répartition F est la fonction de répartition empirique notée F_n et définie par :

$$F_n(x) = \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} = \begin{cases} 0 & si \quad x \leq X_1 \\ \frac{k}{n} & si \quad X_k \leq x \leq X_{k+1} \\ 1 & si \quad x \geq X_n \end{cases}$$

C'est un estimateur non paramétrique de fonction de répartition F.

Définition 2.0.2 *Soit $X_1, X_2, ..., X_n$ un échantillon de variables aléatoires (i.i.d) de* *fonction de répartition F ona :*

$$
\begin{aligned}
f(x) &= \lim_{h \longrightarrow 0} \frac{F(x+h) - F(x)}{2h} \\
f_n(x) &= \lim_{h \longrightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h}.
\end{aligned} \tag{2.1}
$$

L'estimation à noyau de Rosenblatt en 1956 notée $\hat{f}_n(x)$ et définie par :

$$
\begin{aligned}
\hat{f}_n(x) &= \lim_{h \longrightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h} \\
&= \frac{1}{2h}\left[\frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x+h\}} - \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x-h\}}\right] \\
&= \frac{1}{2nh}\left[\sum_{i=1}^{n} I_{\left\{\frac{X_i - x}{h} \leq 1\right\}} - \sum_{i=1}^{n} I_{\left\{\frac{X_i - x}{h} \leq -1\right\}}\right] \\
&= \frac{1}{nh}\left[\sum_{i=1}^{n} \frac{1}{2} I_{\left\{\frac{|X_i - x|}{h} \leq 1\right\}}\right]
\end{aligned} \tag{2.2}
$$

est le premier exemple d'estimateur à noyau construit à l'aide du noyau $k(u) = \frac{1}{2} I_{\{-1<u<1\}}$.

2.1 Estimation à noyau des quantiles

Soit $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ la statistique d'ordre des variables aléatoires i.i.d. $X_1, X_2, ..., X_n$ de fonction de répartition F absolument continue. Rappelons que la fonction de quantile Q_X est l'inverse de F_X définie par :

$$
Q_X(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, 0 < p < 1. \tag{2.3}
$$

estimateur lisse du quantile est fourni par Yang (1985)[24] et a été également retracé par Parzen (1979) [18], est donné par :

$$\tilde{Q}_n(p) = \frac{1}{h}\sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} k\left(\frac{x-p}{h}\right) dx \tag{2.4}$$

k le noyau, et $X_{(i)}$ la statistique d'ordre.

2.2 Propriété de l'estimateur à noyau des quantiles

Dans cette partie, Nous présentons les propriétés statistiques de l'estimateur $\tilde{Q}_n$.

2.2.1 Le biais

Supposant que le noyau k, la fenêtre h et Q satisfaisaient les hypothèes suivantes :

1. $h \longrightarrow 0$ quand $n \longrightarrow \infty$;
2. $\int_{-\infty}^{+\infty} u^2 k(u)\, du < \infty$;
3. Q''_X et Q'_X sont bornées ;
4. $\int_{-\infty}^{+\infty} u\ k(u)\, du < \infty$;

Proposition 2.2.1 *Sous les hypothèse 1-4, l'estimateur $\tilde{Q}_n$ de la fonction Q_X est asymptotiquement non biaisé. C'est-à-dire,* $\lim\limits_{n\longrightarrow\infty} biais\left(\tilde{Q}_n(p)\right) = 0$.

Démonstration :

ona :

$$\begin{aligned}
biais\left(\tilde{Q}_n(p)\right) &= E\left(\tilde{Q}_n(p) - Q_X(p)\right) \\
&= E\left(\sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x-p}{h}\right) dx\right) - Q_X(p) \\
&= \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x-p}{h}\right) dx E\left(X_{(i)}\right) - Q_X(p) \\
&= \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k\left(\frac{x-p}{h}\right) dx Q\left(\frac{i}{n+1}\right) - Q_X(p) \\
&= \int_{0}^{1} \frac{1}{h} k\left(\frac{x-p}{h}\right) Q_X(x)\, dx - Q_X(p).
\end{aligned}$$

Soit le changement de variable $y = \frac{x-p}{h}$ donc $x = yh + p$ et $dx = hdy$, par conséquent le biais de $\hat{Q}_n$ sera :

$$biais\left(\tilde{Q}_n(p)\right) = \int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\, Q(yh+p)\, hdy - Q_X(p).$$

On applique un développement de taylor d'ordre 2 pour la fonction $Q(yh+p)$, on obtient :

$$Q(yh+p) = Q_X(p) - \frac{yh}{1!} Q'_X(p) + \frac{y^2h^2}{2!} Q''_X(p) + O\left(y^2h^3\right).$$

$$biais\left(\tilde{Q}_n(p)\right) = \int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\left[Q_X(p) - \frac{yh}{1!}Q'_X(p) + \frac{y^2h^2}{2!}Q''_X(p) + O\left(y^2h^3\right)\right] - Q_X(p) \tag{2.5}$$

$$= \int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\,Q_X(p)\,dy - \int_{\frac{-p}{h}}^{\frac{i-p}{h}} yhQ'_X(p)\,k(y)\,dy + \int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\frac{y^2h^2}{2!}Q''_X(p)\,dy$$

$$+ O\left(y^2h^3\right)\int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\,dy - Q_X(p)$$

$$= Q_X(p)\int_{\frac{-p}{h}}^{\frac{i-p}{h}} k(y)\,dy - hQ'_X(p)\int_{\frac{-p}{h}}^{\frac{i-p}{h}} yk(y)\,dy + \frac{h^2}{2!}Q''_X(p)\int_{\frac{-p}{h}}^{\frac{i-p}{h}} y^2k(y)\,dy$$

$$+ O\left(y^2h^3\right) - Q_X(p).$$

Lorsque n tend vers l'infini, on obtient :

$$\begin{aligned}
\lim_{n\longrightarrow\infty} biais\left(\tilde{Q}_n(p)\right) &= Q_X(p)\int_{-\infty}^{+\infty} k(y)\,dy - hQ'_X(p)\int_{-\infty}^{+\infty} yk(y)\,dy \\
&+ \frac{h^2}{2!}Q''_X(p)\int_{-\infty}^{+\infty} y^2k(y)\,dy + O\left(y^2h^3\right) - Q_X(p) \\
&= \frac{h^2}{2!}Q''_X(p)\int_{-\infty}^{+\infty} y^2k(y)\,dy + O\left(y^2h^3\right).
\end{aligned}$$

$\lim_{h\longrightarrow\infty} biais\left(\tilde{Q}_n(p)\right) = 0$, alors $\tilde{Q}_n$ est asymptotiquement non biaisé.

2.2.2 La variance

Falk(1984)[6] a prouvé que la variance de $\tilde{Q}_n$ est minimale sous les hypothéses suivantes :

1. k est de support compact ;
2. Q'_X est dérivable ;
3. Q''_X est bornée ;

4. $\mu_2(k) = \int_{-\infty}^{+\infty} t^2 k(t)\,dt < \infty$;

5. $\int (q - ht)\, tk(t)\, J(t)\,dt < \infty$, tq $J(t) = \int_{-\infty}^{t} xk(x)\,dx$

En écrivant $\tilde{Q}_X$ sous la forme :

$$\tilde{Q}_n(p) = \int_0^1 F_n^{-1}(x)\, h^{-1} k\left(\frac{p-x}{h}\right) dx,$$

où F_n est la fonction de répartition empirique.

Sous les hypothèse 1-5, l'estimateur $\tilde{Q}_n$ de la fonction Q_X est a variance minimal asymptotique. C'est à dire la variance de $\tilde{Q}_n(p)$ défini par :

$$Var\left(\tilde{Q}_n(p)\right) = \frac{p(1-p)}{n}\left(Q'_X(p)\right)^2 - \frac{h}{n}\left(Q'_X(p)\right)^2 \int xk(x)\, K_h(x)\,dx + O\left(n^{-1}h\right) \tag{2.6}$$

est tend vers 0 lorsque n tend vers l'infini. et $K_h(x) = \frac{1}{h}\int_{-\infty}^{x} k\left(\frac{t}{h}\right) dt.$

2.2.3 Erreur Quadratique Moyenne (MSE)

Le théorème suivant donne l'expression de l'erreur quadratique moyenne asymptotique de $\tilde{Q}_n(p)$ basée sur l'expression de la variance calculée par Falk(1984).

Supposons que Q''_X est continue dans voisinage p et k est une densité (un noyau) symétrique par rapport à zéro, a support compact. L'erreur quadratique moyenne de $\tilde{Q}_n(p)$ est :

– Lorsque F_X est symétriques et $p \neq \frac{1}{2}$ (ou lorsque F_X est asymétriques) :

$$\begin{aligned} MSE\left(\tilde{Q}_n(p)\right) = {} & \frac{p(1-p)}{n}\left(Q'_X(p)\right)^2 + \frac{h^4}{4}\left(Q''_X(p)\right)^2 \mu_2^2(k) - \frac{h}{n}\left(Q'_X(p)\right)^2 \psi(k) \\ & + O\left(\frac{h}{n} + h^4\right). \end{aligned} \tag{2.7}$$

– Lorsque F_X est symétrique et $p = \frac{1}{2}$:

$$MSE\left(\tilde{Q}_n(p)\right) = \frac{1}{n}\left(Q'(0.5)\right)^2\left(0.25 - 0.5h\psi(k) + (nh)^{-1}R(k)\right) + O\left(n^{-1}h + (nh)^{-2}\right). \tag{2.8}$$

tel que $R(k) = \int_{\mathbb{R}} k^2(x)\,dx$, et $\psi(k) = 2\int_{\mathbb{R}} uK_h(u)\,k(u)\,du$.

Démonstration :

Découle directement des deux propositions précédentes.

Notez que pour une sélection raisonnable de h (c'est-à-dire tendant vers zéro plus rapidement que $n^{\frac{-1}{4}}$), le terme dominant pour MSE est la variance asymptotique de la quantité d'échantillon.

En s'appuyant sur *Falk (1984), Sheader et Marron (1990)* ont donné l'erreur quadratique moyenne asymptotique (AMSE) de $\tilde{Q}_n(p)$ comme suit :

i) Lorsque F_X est symétrique et $p \neq \frac{1}{2}$;

L'erreur quadratique moyenne asymptotique de $\tilde{Q}_n$ est :

$$AMSE\left(\tilde{Q}_n(p)\right) = \frac{p(1-p)}{n}\left(Q'(p)\right)^2 + \frac{h^4}{4}\left(Q''(p)\right)^2\mu_2^2(k) - \frac{h}{n}\left(Q'(p)\right)^2\psi(k). \tag{2.9}$$

– Si $Q'_X(p) > 0$, la fenêtre optimale asymptotique pour $AMSE\left(\tilde{Q}_n(p)\right)$ est :

$$\tilde{h}_{opt} = \left\{\frac{\left(Q'_X(p)\right)^2\psi(k)}{n\left(Q''_X(p)\right)^2\left(\mu_2(k)\right)^2}\right\}^{\frac{1}{3}}, \tag{2.10}$$

et à partir de cette valeur, on obtient l'erreur quadratique moyenne asymptotique optimale suivante :

$$\begin{aligned} AMSE_{opt}\left(\tilde{Q}_n(p)\right) &= \frac{1}{n}\left\{p(1-p)\left(Q'_X(p)\right)^2 + \frac{3}{4}\left(\frac{\left(Q'_X(p)\right)^8\left(\psi(k)\right)^4}{n\left(Q''_X(p)\right)^2\mu_2^2(k)}\right)^{\frac{1}{3}}\right\} \\ &= n^{-1}p(1-p)\left(Q'_X(p)\right)^2 + O\left(n^{-\frac{4}{3}}\right). \end{aligned} \tag{2.11}$$

ii) Lorque F_X est symétrique et $p = \frac{1}{2}$:

L'erreur quadratique moyenne asymptotique de $\tilde{Q}_n(p)$ est :

$$AMSE\left\{\tilde{Q}_n(p)\right\} = \frac{1}{n}\left(Q'_X(0.5)\right)^2\left\{0.25 - 0.5h\psi(k) + (nh)^{-1}R(k)\right\}.$$

$R(k)$: est précédemment connu.

Remarque 2.2.1

Si elle était $Q'_X = 0$, nous avons besion de termes d'ordre supérieur peut afficher $AMSE$ de $\tilde{Q}_n(p)$ par :

$$AMSE\left(\tilde{Q}_n(p)\right) = \left(\frac{1}{4} - \frac{1}{n}\right)h^4 Q''_X(p)^2 \mu_2^2(k) + 2n^{-1}h^2 Q''_X(p)^2 \int (p - ht)\, tk(t)\, J(t)\, dt,$$

tel que $J(t) = \int_{-\infty}^{t} xk(x)\,dx$. La preuve est fournie en (Nadaraya (1964) [16]).

2.2.4 Choix du paramètre de lissage h de $\tilde{\mathbf{Q}}_n(p)$

Nous intéressons au choix du paramètre de lissage h de $\tilde{Q}_n(p)$, pour tout p différent de $\frac{1}{2}$ et si F symétrique

Définition 2.2.1 *Un noyau est dit d'ordre m pour quelques $m \geq 2$ si :*

$$\int t^j k(t)\,dt = \begin{cases} 1 & si \quad j = 0 \\ 0 & si \quad j = 1, \dots, m-1 \\ \mu_m & si \quad j = m \end{cases} \quad où \quad \mu_m \neq 0.$$

Nous avons vu dans la formule 2.9 que le choix donné de k, la valeur asymptotiquement optimale de h dépend les première et seconde dérivées de la fonction des quantiles $Q_X(p)$. Ainis, les estimations de $Q'_X(p)$ et $Q''_X(p)$ sont nécessaires pour le choix de h. Si la première et la seconde dérivées de $\tilde{Q}_n(p)$, ce qui donne l'estimateur :

$$\tilde{Q}'_X(p) = \sum_{i=1}^{n} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} a^{-2} k' \left(a^{-1} (x-p) \right) dx \right] X_{i,n},$$

et

$$\tilde{Q}''_X(p) = \sum_{i=1}^{n} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} b^{-3} k'' \left(b^{-1} (x-p) \right) dx \right] X_{i,n},$$

où k est un noyau d'ordre $m.$, et a, le paramètre de lissage de $\tilde{Q}'_X$, et b le paramètre de lissage de $\tilde{Q}''_X$

Le paramètre de lissage optimal est donnée par :

$$\tilde{h}_{opt} = \left(\frac{\left(\tilde{Q}'_X(p)\right)^2 \psi(k)}{n\left(\tilde{Q}''_X(p)\right)^2 \mu_2^2(k)} \right)^{\frac{1}{3}}. \tag{2.12}$$

Le problème est alors de choisir des valeurs pour le paramètre de lissage a et b.

Théorème 2.2.1 *[21]Supposons que $Q^{(m+2)}$ est continue au voisinage de p, et que k est un noyau d'ordre un à support compact, symétrique par rapport à zéro. Alors, le paramètre de lissage optimal pour $\tilde{Q}'_X(p)$ est donnée par :*

$$a_{opt} = \left(\frac{Q'_X(p)}{Q_X^{(m+1)}(p)} \right)^{2/(m+2)} \left[\frac{(m!)^2 \int k^2(t)\, dt}{2m\left(\int t^m k(t)\, dt\right)^2} \right]^{1/(2m+1)} n^{-1/(2m+1)},$$

et le paramètre de lissage optimal pour $\tilde{Q}''_X(p)$ est alors donnée par :

$$b_{opt} = \left(\frac{Q'_X(p)}{Q_X^{(m+2)}(p)} \right)^{2/(m+3)} \left[\frac{3(m!)^2 \int k'^2(t)\, dt}{2m\left(\int t^m k(t)\, dt\right)^2} \right]^{1/(2m+3)} n^{-1/(2m+3)},$$

Discussion du comportement du biais et de la variance :

1. Le biais décroit si h diminue mais la variance augmente.
2. La variance diminue si h augmente mais le biais augmente.
3. Pour que la variance tend vers zéro, il faut que $nh \longrightarrow \infty$.

2.3 Normalité asymptotique

Théorème 2.3.1 *Si Q_X a une dérivée seconde bornée au voisinage de $q \in (0,1)$ et si $Q'_X > 0$, k a un support borné, tel que $\int k(x)\,dx = 1$ et $h_n \overset{n \to 0}{\longrightarrow} 0$ on a :*

$$\frac{n^{\frac{1}{2}}\left[\tilde{Q}_n(p) - E\left(\tilde{Q}_n(p)\right)\right]}{var\left[\tilde{Q}_n(p)\right]} \longrightarrow N(0,1).$$

Remarque 2.3.1 *Démonstration voir Falk, M. (1985)[7].*

Conclusion

L'estimation non paramétrique des quantiles est une méthode statistique couramment utilisée pour estimer les valeurs d'une distribution à une position donnée. Elle est utile lorsqu'il n'y a pas d'hypothèse claire sur la forme de la distribution.

On a donné un aperçu sur l'estimation à noyau des quantiles et leurs propriétés. Cette méthode d'estimation non paramétrique est basée sur l'utilisation de deux fonctions : un noyau et une fenêtre ou paramètre de lissage. On a vu que le choix de ce dernier, est basé sur l'utilisation du quantile et ces dérivés. Mais, l'inconvénient de cette méthode est que pratiquement ces quantités sont inconnus et le paramètre de lissage optimal de l'estimateur du quantile se détermine, dans ce cas, par la méthode de validation croisée.

Bibliographie

[1] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika, 68(1), 326-328.

[2] Capéraà, P., & Van Cutsem, B. (1988). Méthodes et modèles en statistique non paramétrique : exposé fondamental (Vol. 1). Presses Université Laval.

[3] De Haan, L. (1976). Sample extremes : an elementary introduction (No. 2099-2018-3109).

[4] Djaber, I. (2003). comportement asymptotique presque sûre de la moyenne empirique pour les distribution à queues lourdes (Doctoral dissertation)

[5] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. Theory of Probability & Its Applications, 14(1), 153-158.

[6] Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. The Annals of Statistics, 261-268.

[7] Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. The Annals of statistics, 428-433.

[8] Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. Biometrika, 69(3), 635-640.

[9] Izenman, A. J. (1991). Review papers : Recent developments in nonparametric density estimation. Journal of the american statistical association, 86(413), 205-224.

[10] Jones, M. C. (1992). Estimating densities, quantiles, quantile densities and density quantiles. Annals of the Institute of Statistical Mathematics, 44, 721-727.

[11] Khalifa, I. B. (2008). Estimation non-paramétrique par noyaux associés et données de panel en marketing. Projet de Fin d'Etude. Université du, 7.

[12] Lall, U. (1995). Recent advances in nonparametric function estimation : Hydrologic applications. Reviews of Geophysics, 33(S2), 1093-1102, U.S. National Report to Internationl Union of Geophysics 1991-1994.

[13] Larry Wasserman. All of Statistics : A Concise Course in Statistical Inference, Springer Berlin Heidelberg New York Barcelona Hong London Milan Paris Tokyo.

[14] Malet, J. (2017). cours quantiles et simulation. https : //www. institudesactuaires.com.

[15] Ming-Yen Cheng1 et Shan Sun2, Bandwidth Selction for kernel Quantile Estimation, JEL subject classification : C14,C13.

[16] Nadaraya, E. A. (1964). Some new estimates for distribution functions. Theory of Probability & Its Applications, 9(3), 497-500.

[17] Padgett, W. J. (1986). A kernel-type estimator of a quantile function from right-censored data. Journal of the American Statistical Association, 81(393), 215-222.

[18] Parzen, E. (1979). Non parametric Statistical Data Modelling. Journal of The American Statistical Association. 74,105-131.

[19] Sayah, A. (2015). Kernel quantile estimation for heavy-tailed distribution. Thèse de doctorat, Université Mohamed Khider, Biskra.

[20] Shankar, B. (1998). An optimal choice of bandwidth for perturbed sample quantiles.

[21] Sheather, S. J., & Marron, J. S. (1990). Kernel quantile estimators. Journal of the American Statistical Association, 85(410), 410-416.

[22] Silverman, B.W. (1986) *Density Estimation*. London : Chapman and Hall.

[23] Statistique d'ordre.https ://www.math.u-bordeaux.fr/mchabance/Agreg/probaAgreg 1213-COURS1-Statordre.pdf.

[24] Yang, S. S. (1985). A smooth nonparametric estimator of a quantile function. Journal of the American Statistical Association, 80(392), 1004-1011.

Annexe A : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$i.i.d$	:	Indépendants et identiquement distribuées
F	:	Fonction de distribution
F_n	:	Fonction de distribution empirique
F^{-1}	:	Inverse de la fonction de distribution
Q	:	Fonction quantile
Q_n	:	Fonction quantile empirique
$E[x]$	:	Espérance mathématique ou moyenne du v.a X
$Var[x]$	:	Variance de X
I_A	:	La fonction de l'indécatrice sur A
$\mathbb{R}$	:	Ensemble des nombres réels
$\mathbb{N}$	:	Ensemble des entiers naturels

Θ	:	Espace de paramètres
$\xrightarrow{ps}$	:	Convergence presque sûre
$\bar{x}$	:	La moyenne arithmétique de l'échantillon
σ^2	:	La variance de population
μ	:	La moyenne de la population
ζ_{np}	:	Le quantile empirique d'ordre p
$\hat{\zeta}_{np}$	:	Le quantile empirique de l'échantillon
k	:	Un noyau
h	:	La fenêtre, ou le paramètre de lissage
MSE	:	Mean Square Error (erreur quadratique moyenne)
$AMSE$	:	Asymptotic Mean Square Error (erreur quadratique moyenne asymptotique)
h_{opt}	:	La fenêtre optimale
Q', Q''	:	Le premier et la deuxième dérivé de la fonction quantile
k'	:	La dérivé de noyau k
$\hat{h}_{opt}$	:	L'estimateur de la fenêtre optimale de l'estimateur $\tilde{Q}_n$
$c - \grave{a} - d$	:	C'est-à-dire
$v.a$	:	Variable aléatoire

الملخـــــــــص

في هذه الرسالة ندرس المقدّر اللاوسيطي لدالة الكمية. يعتمد بناء المقدر على استخدام كثافة تسمى K ومعامل تجانس h. نتذكر خصائص المقاربة للمقدر. نتحدث أيضًا عن اختيار دالة النواة و معامل تجانس.

Abstract

In this thesis we study the nonparametric estimator of the quantile function. The construction of the estimator is based on the use of a density K called kernel and a smoothing parameter h. We recall the asymptotic properties of the estimator. We also talk about the choice of kernel and smoothing parameter.

Résumé

Dans ce mémoire nous étudions l'estimateur non paramétrique de la fonction des quantiles. La construction de l'estimateur est basé sur l'utilisation d'une densité K appelée noyau et d'un paramètre de lissage h. Nous rappelons les propriétés asymptotiques de l'estimateur. Nous parlons aussi du choix de noyau et de paramètre de lissage.