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

Menaceur Sabrina

Titre :

Estimation à noyau de la fonction des quantiles

Membres du Comité d'Examen :

Dr.	**MERAGHNI Djamel**	UMKB	Président
Dr.	**SAYAH Abdallah**	UMKB	Encadreur
Dr.	**KHEIREDDINE Souraya**	UMKB	Examinateur

Septembre 2020

Dédicace

Je dédie ce travail à :

Mes chers **parents**, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

Mes chères sœurs : **Imene, Achouak** et **Ikram** pour leurs encouragements permanents, et leur soutien moral.

Mes chers frères : **Houssem, Abd e Raouf** et **Achraf** pour leur appui et leur encouragement.

Tous mes collègues et mes amies qui ont toujours été présents et fidèles, en particulier, ma chère amie : **Chenafi Kamilia** pour son aide précieuse et sa gentillesse, tout au long de mes études universitaire.

Toute la famille "**Menaceur**" et la famille "**Chaibi**".

REMERCIEMENTS

Avant tout, je remercie **Dieu** Le Tout Puissant qui m'a guidé tout au long de ma vie, qui m'a permis de m'instruire et d'arriver aussi loin dans mes études, qui m'a donné courage et patience pour passer tous les moments difficiles.

Je tiens à exprimer toute ma reconnaissance à mon encadreur **Dr. Sayah Abdallah** pour ses conseils bénéfiques, sa guidence, sa compéhension, et ses apports précieux tout au long de la réalisation de ce travail. Merci de votre disponibilité, de votre grande patience et de vos conseils forts judicieux.

Naturellement, je ne peux oublier **Almi Nassima** qui m'a apporté leur aide et qui a contribué à l'élaboration de ce travail .

J'adresse mes sincères remerciements à les membres de jury **Dr**. **Meraghni Djamel** et **Dr. Kheireddine Souraya** d'avoir accepté d'évaluer mon travail.

Mes remerciements sont adressés également à tous les professeurs et les membres de département de mathématique de l'université de Mohamed Khider Biskra.

Mes vifs remerciement à mes chers amis et collègues, ainsi que tous ceux qui m'ont aidé de près ou de loin.

Finalement, je tiens à remercier ma famille, surtout mes parents qui m'ont toujours soutenue et encouragé tout au long de la réalisation de ce travail.

Table des matières

Table des figures

Liste des tableaux

Introduction

L'estimation de la fonction des quantiles à partir d'un ensemble fini d'observations est un problème fondamental dans la littérature, cette estimation a trouvé des champs d'application par exemple en : climatologie, hydrologie, assurance et finance, ect....

Deux types d'approches d'estimation de la fonction des quantiles sont utilisées et largement discutées dans littérature : "l'approche paramétrique et l'approche non paramétrique". L'approche paramétrique a comme inconvénient principal est de nécessiter une connaissance préalable du phénomène aléatoire considéré et l'approche non paramétrique est une estimtion des quantiles directement à partir de l'information disponible sur l'ensemble d'observations.

L'estimation de la fonction des quantiles a été étudiée par : Nadaraya (1964), Yamato(1973), Azzalini (1981), Yang (1985), Parzen (1979), Falk (1984), Padgett (1986), Sheather et Marron (1990), Ralescu et Sun (1993), Harrell et Davis (1982) et Park(2006).

L'objectif de ce mémoire est d'utiliser l'estimation non paramétrique de la fonction des quantiles, dont plusieurs méthodes ont été proposées citons par exemple le quantile empirique, l'estimateur implicite et l'estimateur explicite (estimateur à noyau) qui est le but de mon mémoire.

Ce mémoire est composé de trois chapitres.

Le premier chapitre est consacré aux statistiques d'ordre. Tout d'abord, on donne les définitions de la statistique d'ordre, les distributions associées, et après nous étudions la convergence presque sûre des statistique d'ordre et ses moments.

Le second chapitre est consacré à quelques rappels sur la fonction de distribution, et certaines définitions des quantiles, puis nous avons présenté les deux différentes approches d'estimations des quantiles utilisées : "l'approche paramétrique et l'approche non paramétrique".

Dans le dernier chapitre, nous présentons les résultats des simulations (avec le logiciel de statistique R).

Chapitre 1

Les Statistiques d'ordre

Les statistiques d'ordre jouent un rôle de plus en plus important dans la statistique non paramétrique.

Dans ce chapitre, nous avons commencé, par donner les définitions des statistiques d'ordre, les distributions associées, puis nous étudions la convergence presque sûre des statistiques d'ordre et ses moments.

1.1 Définitions des statistiques d'ordre

Soit $(X_1, X_2, ..., X_n)$ une suite de variables aléatoires $(v.a)$ indépendantes et identiquement distribuées $(i.i.d)$ selon une distribution F tel que : $F(x) = P(X \leq x)$, pour $x \in \mathbb{R}$.

Définition 1.1.1 ***(Statistique d'ordre)***

On appelle statistique d'ordre notée par $(X_{1,n}, X_{2,n}, ..., X_{n,n})$ les variables aléatoires rangées par ordre croissant :

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

Remarque 1.1.1 *1.On a deux cas importants de la statistique d'ordre sont les statistiques du* ***minimum*** *et du* ***maximum (Les valeurs extrêmes),*** *telles que :*

$$m_n = X_{1,n} = \min(X_1, ..., X_n)$$
$$M_n = X_{n,n} = \max(X_1, ..., X_n)$$

où pour $k = 1, ..., n$ la $v.a$ $X_{k,n}$ appelée la $k^{ième}$ statistique d'ordre (**statistique de rang** k).

2.Les valeurs aléatoires de la statistique d'ordre $X_{1,n}, X_{2,n}, ..., X_{n,n}$ ne sont plus indépendantes et de distributions différentes.

Exemple 1.1.1 *Pour un échantillon* $(X_1, \ldots, X_4) = (0.1; 0.7; 0.2; 0.15)$

- Les statistiques d'ordre sont :$X_{1,4} = 0.1 \leq \ X_{2,4} \leq 0.15 \leq \ X_{3.4} = 0.2 \leq X_{4,4} = 0, 7$.
- Statistique du minimum est :$m_4 = X_{1,4} = 0.1$.
- Statistique du maximum est :$M_4 = X_{4,4} = 0.7$.

1.2 Distributions des Statistiques d'ordre

1.2.1 Densité Conjointe de n Statistiques d'ordre

Proposition 1.2.1 *Soit* $(X_1, ..., X_n)$ *une suite des variables aléatoires i.i.d de densité* f*. Alors la densité conjointe de n statistiques d'ordre* $(X_{1,n}, X_{2,n}, ..., X_{n,n})$ *est :*

$$f_{X_{1,n},...,X_{n,n}}(x_1, ...x_n) = n!\prod_{i=1}^{n} f(x_i), -\infty < x_1 < ... < x_n < +\infty. \tag{1.1}$$

1.2.2 Distribution de la k^{ieme} Statistique d'ordre

Théorème 1.2.1 *Soient* $X_1, ..., X_n$ *un échantillon d'une distribution* F *et* $X_{k,n}$ *la* $k^{\text{ième}}$ *statistique d'ordre associée. Alors , pour tout* k *appartenant à*$\{1, ..., n\}$ *la fonction de distribution* $F_{X_{k,n}}$ *da la variable* $X_{k,n}$ *est donnée par :*

$$F_{X_{k,n}}(x) = \sum_{r=k}^{n} C_n^r \left[F(x)\right]^r \left[1 - F(x)\right]^{n-r} \quad , x \in \mathbb{R}, \tag{1.2}$$

et sa densité est :

$$f_{X_{k,n}}(x) = \frac{n!}{(k-1)!(n-k)!} \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x) \ , x \in \mathbb{R}. \tag{1.3}$$

Preuve. On a

$$\begin{aligned}
F_{X_{k,n}}(x) &= P\left(X_{k,n} \leq x\right) \\
&= P\left\{\text{au moins } k \text{ des } X_r \text{ sont inférieur à } x\right\} \\
&= \sum_{r=k}^{n} P\left\{\text{exactement } r \text{ de } X_1, ..., X_n \text{ sont inférieur à } x\right\} \\
&= \sum_{r=k}^{n} C_n^r \left[F(x)\right]^r \left[1 - F(x)\right]^{n-r} \quad , x \in \mathbb{R},
\end{aligned}$$

où

$$C_n^r = \frac{n!}{r!(n-r)!}.$$

■

1.2.3 Distribution Conjointe de deux Statistiques d'ordre

Théorème 1.2.2 *Soient* $X_1, ..., X_n$ *un échantillon d'une distribution* F *et* $X_{i,n}$ *et* $X_{j,n}$ *les* $i^{\text{ième}}$ *et* $j^{\text{ième}}$ *statistiques d'ordre associées, avec* $1 \leq i \leq j \leq n$ *. Alors la fonction de* *distribution $F_{(X_{i,n},X_{j,n})}$ d'un couple de statistique d'ordre $(X_{i,n}, X_{j,n})$ est donnée par :*

$$F_{(X_{i,n},X_{j,n})}(x,y) = P(X_{i,n} \leq x, X_{j,n} \leq y). \tag{1.4}$$

On a deux cas :

- $\mathbf{1}^{er}$**Cas** : Si $x < y$

$$F_{(X_{i,n},X_{j,n})}(x,y) = \sum_{r=j}^{n}\sum_{s=i}^{r} \frac{n!}{s!(r-s)!(n-r)!} \left[F(x)\right]^{s} \left[F(y)-F(x)\right]^{r-s} \left[1-F(y)\right]^{n-r}, \tag{1.5}$$

$$-\infty < x < y < +\infty$$

- $\mathbf{2}^{ème}$**Cas** : Si $x \geq y$

$$F_{(X_{i,n},X_{j,n})}(x,y) = F_{X_{j,n}}(y). \tag{1.6}$$

La densité conjointe d'un couple $(X_{i,n}, X_{j,n})$ dans le **cas génerale** est :

$$f_{(X_{i,n},X_{j,n})}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \left[F(x)\right]^{(i-1)} \left[F(y)-F(y)\right]^{j-i-1} \left[1-F(y)\right]^{n-j} f(x)f(y), \tag{1.7}$$

$$-\infty < x < y < +\infty.$$

En particulier, pour $i = 1$ et $j = n$, on obtient la densité conjointe du minimum et du maximum

$$f_{(X_{1,n},X_{n,n})}(x,y) = n(n-1)\left[F(y)-F(x)\right]^{n-2} f(x)f(y), -\infty < x < y < +\infty.$$

Preuve. considérons le cas $x < y$:

$$
\begin{aligned}
F_{(X_{i,n},X_{j,n})}(x,y) &= P(X_{i,n} \leq x, X_{j,n} \leq y) \\
&= P\{\text{au moins } i \text{ de } X_1,...,X_n \text{ Sont inférieur à } x \\
&\text{et au moins } j \text{ de } X_1,...,X_n \text{ Sont inférieur à } y\} \\
&= \sum_{r=j}^{n}\sum_{s=i}^{r} P\{\text{exactement } s \text{ de } X_1,...,X_n \text{ Sont inférieur à } x \\
&\text{et exactement } r \text{ de } X_1,...,X_n \text{ Sont inférieur à } y\} \\
&= \sum_{r=j}^{n}\sum_{s=i}^{r} \frac{n!}{s!(r-s)!(n-r)!} \left[F(x)\right]^s \left[F(y)-F(x)\right]^{r-s} \left[1-F(y)\right]^{n-r}, \\
&-\infty < x < y < +\infty.
\end{aligned}
$$

De meme pour le cas $x > y$:

$$
\begin{aligned}
F_{(X_{i,n},X_{j,n})}(x,y) &= P(X_{i,n} \leq x, X_{j,n} \leq y) \\
&= P(X_{j,n} \leq y) \\
&= F_{X_{j,n}}(y).
\end{aligned}
$$

■

1.2.4 Distribution des Valeurs extrêmes

Soit $X_1,...,X_n$ .un échantillon de densité f et de fonction de distribution F.

• Distribution du minimum

La distribution du minimum $X_{1,n}$ donnée par :

$$
F_{X_{1,n}}(x) = 1 - \left[1 - F_X(x)\right]^n, \forall x \in \mathbb{R}, \tag{1.8}
$$

et sa densité est :

$$f_{X_{1,n}}(x) = nf_X(x)\left[1 - F_X(x)\right]^{n-1}, \forall x \in \mathbb{R}. \tag{1.9}$$

Preuve. On a

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

et on déduit que la densité est :

$$f_{X_{1,n}}(x) = \frac{dF_{X_{1,n}}(x)}{dx} = nf_X(x)\left[1 - F_X(x)\right]^{n-1}, \forall x \in \mathbb{R}.$$

■

• **Distribution du maximum**

La distribution du maximum $X_{n,n}$ est donnée par :

$$F_{X_{n,n}}(x) = \left[F_X(x)\right]^n, \forall x \in \mathbb{R} \tag{1.10}$$

et sa densité est :

$$f_{X_{n,n}}(x) = n\left[F(x)\right]^{n-1} f(x), \forall x \in \mathbb{R}. \tag{1.11}$$

Preuve. On a

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

et on déduit que

$$f_{X_{n,n}}(x) = \frac{dF_{X_{n,n}}(x)}{dx} = n\,[F(x)]^{n-1}\, f(x), \forall x \in \mathbb{R}.$$

■

1.3 Convergence presque sûre des statistiques d'ordre

Proposition 1.3.1 *Soit* $p = 1$ *(resp* $p = 0$*), et soit* $(k(n), n \geq 1)$ *une suite d'entiers telle que :*

$$1 \leq k(n) \leq n \quad \text{et} \quad \lim_{n\to\infty} \frac{k(n)}{n} = p. \tag{1.12}$$

Alors, $(X_{(k(n),n)}, n \geq 1)$ converge presque sûrement vers $x_F = \inf\{x, F(x) = 1\}$ (resp $\tilde{x}_F = \sup\{x, F(x) = 0\}$).

1.4 Moments des Statistiques d'ordre

Proposition 1.4.1 *Soit $X_{k,n}$ la $k^{ième}$ Statistique d'ordre associée à l'échantillon de taille n de densité f et de fonction de distribution F continue.Alors le $m^{ième}(m = 1, 2, ..)$ moment de la $k^{ième}(k = 1, ..., n)$ statistique d'ordre est :*

$$\begin{aligned}
\mu_{k,n}^{(m)} = EX_{k,n}^{m} &= \int_{-\infty}^{+\infty} x^m f_{X_{k,n}}(x)dx \\
&= \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{+\infty} x^m \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x)dx \\
&= \frac{1}{\mathbf{B}(k, n-k+1)} \int_{-\infty}^{+\infty} x^m \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x)dx, \qquad (1.13)
\end{aligned}$$

où

$$\mathbf{B}(\alpha, \beta) = \tfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{et} \quad \Gamma(\alpha) = \int_{0}^{+\infty} t^{\alpha-1} \exp(-t)dt, \forall\alpha > 0.$$

1.4.1 Existence des moments des Statistiques d'ordre

Théorème 1.4.1 *Soient $X_1, ..., X_n$ un échantillon d'une distribution F continue, et $X_{1,n}, ..., X_{n,n}$ les statistiques d'ordre associées. Soit m un entier strictement positif. Si X possède un moment d'ordre m. Alors, pour tout k appartenat à $\{1, ..., n\}$, la $k^{ième}$ statistique d'ordre $X_{k,n}$ possède un moment d'ordre m.*

Preuve. on a la distribution de $X_{k,n}$ est donée par :

$$F_{X_{k,n}}(t) = \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{t} \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x)dx.$$

■

Nous avons alors,

$$EX_{k,n}^{m} = \frac{n!}{(k-1)!(n-k)!}\int_{-\infty}^{+\infty} x^{m}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k} f(x)dx,$$

et donc,

$$E[\mid X_{k,n}^{m} \mid] \leq \frac{n!}{(k-1)!(n-k)!}\int_{-\infty}^{+\infty} \mid x \mid^{m} f(x)dx = \frac{n!}{(k-1)!(n-k)!}E[\mid X \mid^{m}].$$

Remarque 1.4.1 *Le théorème montre que l'existence du moment d'ordre* m *de* X *entraîne l'existence du moment d'ordre* m *de* $X_{k,n}$.

$$(EX^{m} \quad \textit{existe} \Longrightarrow \forall k \in \{1,...,n\},\ EX_{k,n}^{m} \quad \textit{existe}).$$

La réciproque n'est pas vraie

$$(\forall k \in \{1,...,n\},\ EX_{k,n}^{m} \ \text{ existe} \nRightarrow EX^{m} \ \text{ existe}).$$

Chapitre 2

Estimation de la fonction des quantiles

En statistiques et en théorie des probabilités, les quantiles sont les valeurs qui divisent un jeu de données en intervalles contenant le même nombre de données.

Les quantiles d'une variable aléatoire discrète (entière) ou continue (réelle), sont les valeurs que prend la variable pour des valeurs de probabilité. On les appelles encore fractiles, et ce sont les valeurs réciproques de la fonction de distribution de la loi de probabilité considérée.

2.1 Fonction de distribution empirique

Définition 2.1.1 *On appelle la fonction de distribution empirique associée à un échantillon $X_1, ..., X_n$ la fonction notée F_n définie par :*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(X_i \leq x)}(x) \tag{2.1}$$

$$= \frac{card\{X_i \leq x, i = 1, .., n\}}{n}$$
$$= \begin{cases} 0 & \text{Si} \quad x < X_{1,n} \\ \frac{i}{n} & \text{Si } X_{i,n} \leq x < X_{i+1,n} \\ 1 & \text{Si} \;\; x \geq X_{n,n} \end{cases}$$

où $\mathbf{1}_A(x)$ est la fonction de l'indicatrice :

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{Si } x \in \mathbb{R} \\ 0 & \text{Sinon} \end{cases}$$

2.1.1 Propriétés de la distribution empirique

En utilisant les propriétés de la distribution binomiale, nous obtenons les résultats suivants :

Corollaire 2.1.1 *La moyenne et la variance de F_n sont :*

$$E\left[F_n(x)\right] = F(x) \qquad et \qquad Var\left[F_n(x)\right] = \frac{F\left(x\right)\left[1 - F(x)\right]}{n}.$$

Donc, pour tout x,

•$F_n(x)$ est un estimateur **sans biais** de $F(x)$.

•La variance de $F_n\left(x\right)$ tend vers 0, quand n tend vers ∞.

Convergence de la fonction de distribution empirique

Théorème 2.1.1 *Soit* $(X_n, n \in \mathbb{N})$ *une suite des variables aléatoires réelles indépandantes et de même loi* F*, alors :*

$$\forall x \in \mathbb{R}, \quad \lim_{n \longrightarrow n} \sup_{x \in \mathbb{R}} \mid F_n(x) - F(x) \mid = 0 \quad p.s \tag{2.2}$$

Théorème 2.1.2 *Soit* $(X_1, X_2, ..., X_n)$ *une suite de variables aléatoires* $(i.i.d)$ *de même loi* F*, alors :*

$$\frac{\sqrt{n}\,(F_n(x) - F(x))}{\sqrt{F(x)\,[1 - F(x)]}} \xrightarrow{L} N(0,1) \qquad quand\ n \to \infty.$$

2.2 Inverse Généralisé

Définition 2.2.1 *On appelle inverse généralisé d'une fonction de distribution* F *définié de* $\mathbb{R}$ *dans* $[0.1]$ *la fonction notée* $F^{\leftarrow}$ *définie de* $[0.1]$ *dans* $\mathbb{R}$ *par :*

$$F^{\leftarrow}(p) = \inf \{x \in \mathbb{R} : F(x) \geq p\}, \quad 0 \leq p \leq 1. \tag{2.3}$$

L'inverse généralisé $F^{\leftarrow}$ coïncide avec l'inverse F^{-1} lorsque la fonction F est strictement croissante et continue.

2.3 Les Quantiles

2.3.1 La fonction des quantiles

Définition 2.3.1 *Si* F *est la fonction de distribution d'une variable aléatoire réelle* X*. On appelle fonction quantile de* X *notée* Q *l'inverse généralisé de* F *définie par :*

$$\forall p \in]0,1[\,, \quad Q(p) = F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}\,. \tag{2.4}$$

2.3.2 Quantile d'ordre p

Définition 2.3.2 *Soit X une variable aléatoire de fonction de distribution F continue et strictement croissante. Pour tout $p \in]0,1[$, on appelle quantile d'ordre p noté x_p la racine(unique) de l'équation :*

$$F(x_p) = p, \quad i.e. \;\; (x_p = F^{-1}(p)). \tag{2.5}$$

Remarque 2.3.1 *Pour certaines valeurs de p, il ya des noms spéciaux aux quantiles :*

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

2.3.3 Quantile extrême

Définition 2.3.3 *On appelle quantile extrême le quantile d'ordre $(1-p)$, défini par :*

$$x_{1-p} = F^{-1}(1-p) = \inf\{x \in \mathbb{R} : F(x) \geq 1-p\}\,, \tag{2.6}$$

où p est proche de zéro.

2.3.4 Fonction quantile de queue

Notons $\overline{F} = 1 - F(x)$ la fonction de queue (ou de survie).

Définition 2.3.4 *La fonction de quantile de queue est définie par :*

$$U(s) = F^{\leftarrow}(1 - \frac{1}{s}), \quad \textit{avec } 0 < s < \infty, \tag{2.7}$$

$F^{\leftarrow}$ étant l'inverse généralisé de F.

•La fonction quantile **empirique de queue** est définie par :

$$U_n(s) = F^{\leftarrow}(1 - \frac{1}{s}), \quad \text{avec } 0 < s < \infty. \tag{2.8}$$

2.4 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 applications.

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.

2.4.1 L'approche paramétrique

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

$$Q_X(p) = F_{\theta}^{-1}(p), \tag{2.9}$$

donc l'estimateur naturel est :

$$\hat{Q}_X(p) = F_{\hat{\theta}}^{-1}(p). \tag{2.10}$$

Cette méthode est pratique, car il existe plusieurs téchniques pour l'obtention $\hat{\theta}$ (méthodes de moments, maximum de vraisemblance ,...), mais le choix de $\mathcal{F}$ 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 :

$$Q_X(p) = \mu + \sigma\Phi^{-1}(p), \tag{2.11}$$

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

Définition 2.4.1 *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 :*

$$\hat{Q}_X(p) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(p), \tag{2.12}$$

où

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

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

$$Q_X(p) \sim E(X) + \sigma \hat{z}_p, \tag{2.13}$$

où

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

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 :

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

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

$$\hat{Q}_n(p) = \hat{\mu} + \sigma \hat{z}_p. \tag{2.14}$$

2.4.2 L'approche non-paramétrique

Dans la littérature, plusieurs méthodes ont été dédiées à l'estimation non paramétrique de la fonction des quantiles. Nous citons par exemple :

• **Quantile empirique**

Définition 2.4.3 *La fonction quantile empirique de l'échantillon $X_1, ..., X_n$ est donnée par :*

$$Q_n(p) = F_n^{-1}(p) = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}, \quad avec\ 0 < p < 1 \tag{2.15}$$

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

Définition 2.4.4 *Soit $(X_1, ..., X_n)$ un n échantillon issu d'une distribution F et $X_{1,n}, ..., X_{n,n}$ les statistiques 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.*

$$X_{([np]+1),n} = Q_n(p) = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}, \tag{2.16}$$

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 :

$$ABiais\{X_{k,n}\} = \frac{p_k q_k Q''(p_k)}{2(n+2)} + \frac{p_k q_k}{(n+2)^2}\left\{\frac{1}{3}(q_k - p_k)Q_k''' + \frac{1}{8}Q_k''''\right\}, \tag{2.17}$$

et

$$AVar\{X_{k,n}\} = \frac{p_k q_k}{(n+2)}Q_k'^2 + \frac{p_k q_k}{(n+2)^2}\left\{2(q_k - p_k)Q_k' Q_k'' + p_k q_k(Q_k' Q_k''' + \frac{1}{2}Q_k'')\right\}. \tag{2.18}$$

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

$$AMSE\{X_{k,n}\} = (ABiais\{X_{k,n}\})^2 + Avar\{X_{k,n}\}. \tag{2.19}$$

Remarque 2.4.1 *On peut représenter la fonction des quantiles empirique en fonction des statistiques d'ordre comme suit :*

$$Q_n(p) = X_{k,n} \quad avec\ \frac{k-1}{n} \leq p < \frac{k}{n}. \tag{2.20}$$

Convergence des quantiles empiriques

Proposition 2.4.1 *Soit $p \in]0,1[$ supposons que F est continue et qu'il existe une seule solution x_p à l'équation $F(x) = p$. Soit $(k(n), n \geq 1)$ une suite d'entiers telle que :*

$$1 \leq k(n) \leq n \quad et \quad \lim_{n\to\infty} \frac{k(n)}{n} = p.$$

Alors, la suite des quantiles empiriques $(X_{k(n),n})_{n\geq 1}$ converge presque sûrement vers x_p

$$X_{k(n),n} \overset{p.s}{\to} x_p \quad quand \quad n \to \infty, \tag{2.21}$$

où x_p est le quantile d'ordre p.

Preuve. Le rapport $\frac{k(n)}{n}$ converge vers p , nous posons $k(n) = [np] + 1$.

On a :

$$F_n(X_{[np]+1,n}) = \frac{[np]}{n},$$

■

$$\begin{aligned} F(X_{k(n),n}) - F(x_p) &= \big(F(X_{k(n),n}) - F_n(X_{k(n),n}) + F_n(X_{k(n),n}) - F(x_p)\big) \\ &= F(X_{k(n),n}) - F_n(X_{k(n),n}) + \frac{[np]}{n} - p, \end{aligned}$$

et on a :

$$d_k(F_n, F) = \sup \mid F(X) - F_n(X) \mid \xrightarrow{p} 0,$$

d'où

$$F(X_{k(n),n}) - F(x_p) \xrightarrow{p} 0,$$

et puisque F étant continue et strictement croissante on conclut en composant par F^{-1}

$$X_{k(n),n} \xrightarrow{p.s} x_p \quad \text{quand} \quad n \to \infty.$$

Proposition 2.4.2 *Soit $p \in]0,1[$ supposons que la loi de X possède une densité f continue en x_p et telle que $f(x_p) > 0$. On suppose de plus que $k(n) = np + o(\sqrt{n})$. On a la convergence en loi suivante :*

$$\sqrt{n}(X_{k(n),n} - x_p) \xrightarrow{L} N\left(0, \frac{p(1-p)}{[f(x_p)]^2}\right) \quad \textit{quand} \quad n \to \infty. \tag{2.22}$$

Nous donnons une introduction à l'estimation non paramétrique par la méthode du noyau.

La méthode du noyau

La méthode du noyau est la plus populaire parmi les multiples méthodes d'estimation non paramétriques de la densité. Cette popularité de l'estimateur à noyau peut s'expliquer par au moins trois raisons : la simplicité de sa forme, ses modes de convergence multiples et sa flexibilité qui s'interprète par la liberté de l'utilisateur dans le choix du noyau k et du paramètre de lissage h. L'estimateur à noyau a été proposé initialement par Rosenblatt [1956] et Parzen [1962]. Cet estimateur est une fonction de deux paramètres : Le noyau k et le paramètre de lissage h.

Soient $X_1, ..., X_n$ des variables aléatoires $(i.i.d)$ de fonction de distribution continue F et de densité f . Considérons la fonction de distribution empirique F_X :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(X_i \leq x)}.$$

Définition 2.4.5 ***(etimatuer de Parzen-Rosenblatt)*** *On appelle estimateur à noyau de la densité f ou estimatuer de Parzen-Rosenblatt, l'estimateur qui donnée par :*

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right), \tag{2.23}$$

où

- $k : \mathbb{R} \to \mathbb{R}^+$ est une fonction intégrable dite noyau telle que $\int\limits_{-\infty}^{+\infty} k(u)du = 1$ et vérifiant généralement les conditions suivantes :

$$\int\limits_{-\infty}^{+\infty} t^2 \mid k(t) \mid dt < \infty \quad , \quad \int\limits_{-\infty}^{+\infty} tk(t)dt = 0 \text{ et } \int\limits_{-\infty}^{+\infty} k^2(t)dt < \infty.$$

- Le paramètre $h = h_n > 0$ est une fenêtre ou (paramètre de lissage) qui satisfait $h = h_n \to 0$ quand $n \to \infty$.

Pour obtenir un estimateur non paramétrique de F, nous intégrons l'estimateur de la densité, cet estimateur de F est basé sur l'etimateur à noyau de type Nadaraya (1964), qui est défini par :

$$\begin{aligned} \hat{F}_n(x) &= \int\limits_{-\infty}^{x} \hat{f}_n(y)dy \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \end{aligned} \tag{2.24}$$

où la fonction K est définie par :

$$K(x) = \int\limits_{-\infty}^{x} k(t)dt.$$

Exemples des noyaux (Noyaux symétriques) Les noyaux les plus utilisés dans l'estimation sont donnés dans le tableau suivant :

Noyau	$k(t)$
Rectangulaire(Uniforme)	$\frac{1}{2}\mathbf{1}_{(\|t\|\leq 1)}$
Triangulaire	$(1- \mid t \mid)\mathbf{1}_{(\|t\|\leq 1)}$
Gaussien	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right), t\in\mathbb{R}$
Quartique (Biweight)	$\frac{15}{16}(1-t^2)^2\mathbf{1}_{(\|t\|\leq 1)}$
Epanechnikov(Parabolique)	$\frac{3}{4}(1-t^2)\mathbf{1}_{(\|t\|\leq 1)}$

TAB. 2.1 – Exemples des noyaux (Noyaux symétriques)

.

Revenant maintenant à la construction de l'estimateur de la fonction des quantiles :

Pour l'estimation du quantile non paramétrique, on a deux classes d'estimateurs ont été dérivées. Une classe dite implicite fondée sur des techniques d'inversion numérique et une classe explicite, qui est liée à statistique d'ordre.

• **L'estimateur implicite**

L'estimateur correspondant de la fonction de quantile $Q = F^{-1}$ est défini par :

$$\hat{Q}_n(p) = \inf\left\{x \in \mathbb{R} : \hat{F}_n(x) \geq p\right\}, 0 < p < 1.$$

Sous certaines hypothèses sur le noyau k, la densité f et le paramètre de lissage h, Nadaraya (1964) a montré que $\hat{Q}_n(p)$ a une distrubution asymptotiquement normale standard. La consistance presque sûre, a été obtenue par Yamato (1973).Les conditions nécessaires et suffisantes pour la normalité asymtotiquement de $\hat{Q}_n(p)$ ont été obtenues par Ralesco et Sun (1992). Azzalini (1981) a utilisé des arguments fondés sur des approximations du second ordre et a effectué quelques comparaisons numériques entre $\hat{Q}_n(p)$ et le quantile empirique échantillonné pour estimer le quantiles $95^{ième}$de la distribution Gamma (1).

Azzalini (1981) a considéré la propriété de second ordre de $\hat{F}_n$ sous les hypothèses suivantes :

(1) $h \to 0$ quand $n \to \infty$.

(2) Le noyau est à support fini $k(t) = 0$ si $|t| > t_0$ pour certains t_0 positif.

(3) La densité f est continue dans l'intervalle $(x - t_0 h, x + t_0 h)$.

(4) $f'(x)$ existe.

Il a souligné que le paramètre de lissage optimal pour $\hat{F}$ est de la forme :

$$h_{opt} = \left(\frac{u}{4nv}\right)^{\frac{1}{3}}, \tag{2.25}$$

où

$$u = f(x)\left(t_0 - \int_{-t_0}^{t_0} K^2(t)dt\right), \quad \text{et} \quad v = \left(\frac{1}{2}f'(x)\int_{-t_0}^{t_0} t^2 k(t)dt\right)^2.$$

De plus, Azzalini (1981) a suggéré, sans apporter de preuve que (2.25) est le choix asymptotiquement optimal de h pour $\hat{Q}_n(p)$.

Théorème 2.4.1 *Supposons que f' est bornée et continue dans un voisinage de x_p avec $f'(x_p) \neq 0$,et la fonction du noyau k est une densité bornée, continue et symétrique par rapport à zero, qui satisfait :*

$$\int_{-\infty}^{+\infty} tk(t)dt = 0, \quad et \quad \mu_2(k) = \int_{-\infty}^{+\infty} t^2 k(t)dt < \infty.$$

Alors, pour tout $p \in]0, 1[$, l'erreur quadratique moyenne de $\hat{Q}_n(p)$ est :

$$MSE(\hat{Q}_n(p)) = \frac{p(1-p)}{nf^2(x_p)} + \frac{h^4 (f'(x_p))^2}{4f^2(x_p)} - \frac{h}{nf(x_p)}\phi(k) + o\left(\frac{h}{n} + h^4\right), \tag{2.26}$$

où

$$\phi(k) = 2\int_{-\infty}^{+\infty} yk(y)K(y)dy.$$

Corollaire 2.4.1 *L'erreur quadratique moyenne asymptotique de $\hat{Q}_n(p)$ est :*

$$AMSE(\hat{Q}_n(p)) = \frac{p(1-p)}{nf^2(x_p)} + \frac{h^4\left(f'(x_p)\right)^2}{4f^2(x_p)} - \frac{h}{nf(x_p)}\phi\left(k\right).$$

Le paramètre de lissage optimale pour $AMSE(\hat{Q}_n(p)$ est donnée par :

$$\hat{h}_{opt} = \left(\frac{f(x_p)\phi(k)}{n(f'(x_p))^2\mu_2^2(k)}\right)^{\frac{1}{3}}, \tag{2.27}$$

et l'erreur quadratique moyenne asymptotique associée pour $\hat{h}_{opt}$ est :

$$AMSE_{opt}(\hat{Q}_n(p)) = n^{-1}\left[\frac{p(1-p)}{f^2(x_p)} - \frac{3}{4}\left(\frac{(\phi(k))^4}{nf^2(x_p)(f'(x_p))^2\mu_2^2(k)}\right)^{\frac{1}{3}}\right].$$

• **L'estimateur explicite**

Un autre estimateur de la fonction des quantiles est fourni par Yang (1985) et également retracé à Parzen (1979) est donné par :

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

La normalité asymptotique et l'erreur quadratique moyenne de $\tilde{Q}_n(p)$ ont été fournies par Yang (1985), et Faulk (1984) a montré que le comportement asymptotique de $\tilde{Q}_n(p)$ est meilleure que celui du quantile empirique $Q_n(p)$. Padgett (1986) a généralisé la définition de $\tilde{Q}_n(p)$ aux données censurées à droite.

Dans l'étude de $\tilde{Q}_n(p)$ Yang (1985) et Padgett (1986) ont examiné plusieurs fonctions noyau dont les fonctions triangulaires k_Y donées par :

$$k_Y = \begin{cases} 1- \mid u \mid & \text{si} \quad \mid u \leq 1 \mid \\ 0 & \text{si} \quad \text{non} \end{cases}$$

Sachs et Visaker (1981) sont trouvé que les fonctions du noyau k_Y ont des propriétés optimales dans l'estimation non paramétriques de la fonction de densité.

Parzen (1979), Padgett (1986), Sheather et Marron (1990), et Ralescu et Sun (1993) ont considéré des noyaux gaussiens. Mais tous ces estimateurs ont un grand biais lorsque p est proche de 1. Afin de couriger ce biais, Harrell et Davis (1982) ou Park (2006) suggèrent d'utuliser le noyau asymétrique, à savoir le noyau de type Beta qui est donée :

$$HD_n(p) = \frac{\Gamma(n+1)}{\Gamma((n+1)p)\Gamma((n+1)(1-p))} \int_0^1 F_n^{-1}(y) y^{(n+1)p-1} (1-p)^{(n+1)(1-p)-1} dy,$$

où $F_n^{-1}(x)$ est l'iverse de la fonction de distribution empirique qui est définie par :

$$F_n^{-1}(x) = \begin{cases} X_{i,n} & \text{si} \quad \frac{(i-1)}{n} \leq y \leq \frac{i}{n} \\ X_{n,n} & \text{si} \quad \frac{n-1}{n} \leq y \leq 1 \end{cases}$$

S'appuyant sur Falk (1984), Sheater et Marron (1990) ont donné (MSE) de $\tilde{Q}_n(p)$ comme suit :

Théorème 2.4.2 *Supposons que $Q''(p)$ est continue au voisinage de p et k et une densité à support compact, symétrique par rapport à zéro. Alors :*

$\star$ Si F n'est pas symétrique ou F est symétrique mais $p \neq \frac{1}{2}$, alors :

$$MSE(\tilde{Q}_n(p)) = \frac{p(1-p)}{n} \left(Q'(p)\right)^2 + \frac{h^4}{4} \left(Q''(p)\right)^2 \mu_2^2(k) - \frac{h}{n} \left(Q'(p)\right)^2 \phi(k) + o\left(\frac{h}{n} + h^4\right).$$

$\star$ Si F est symétrique et $p = \frac{1}{2}$, alors :

$$MSE(\tilde{Q}_n(p) = n^{-1} \left(Q'\left(1/2\right)\right)^2 \left[0.25 - 0.5h\phi(k) + (nh)^{-1} R(k)\right] + o\left(n^{-1}h + (nh)^{-2}\right),$$

où

$$R(k) = \int_{-\infty}^{+\infty} k^2(x)dx.$$

Corollaire 2.4.2 *Si F n'est pas symétrique ou F symétrique mais $p \neq \frac{1}{2}$, l'erreur quadratique moyenne asymptotique de $\tilde{Q}_n(p)$ est :*

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

Le paramètre de lissage optimale pour $AMSE(\tilde{Q}_n(p))$ est donnée par :

$$\tilde{h}_{opt} = \left(\frac{\left(Q'(p)\right)^2 \phi(k)}{n(Q''(p))^2 \mu_2^2(k)} \right)^{\frac{1}{3}}, \tag{2.28}$$

et l'erreur quadratique moyenne asymptotique associée pour $\tilde{h}_{opt}$ est :

$$\begin{aligned} AMSE_{opt}(\tilde{Q}_n(p)) &= n^{-1} \left[p(1-p) \left(Q'(p)\right)^2 + \frac{3}{4} \left(\frac{\left(Q'(p)\right)^8 (\phi(k))^4}{n \left(Q''(p)\right)^2 \mu_2^2(k)} \right)^{\frac{1}{3}} \right] \\ &= n^{-1} p(1-p) \left(Q'(p)\right)^2 + o(n^{-4/3}). \end{aligned}$$

Si F est symétrique et $p = \frac{1}{2}$, alors :

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

Choix du paramètre de lissage h

Nous intéressons au choix du paramètre de lissage h de $\tilde{Q}_n(p)$, pour tout p, à part si F est symétrique et $p = 1/2$.

Définition 2.4.6 *Un noyau est dit d'ordre m pour quelques $m \geq 2$ si :*

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

où $\mu_m \neq 0$.

Nous avons vu dans le corollaire 2.4.2 que pour un choix donné de k, la valeur asymptotiquement optimale de h dépend les première et seconde dérivées de la fonction des quantiles $Q(p)$. Ainsi, les estimations de $Q'(p)$ et $Q''(p)$ sont nécessaires pour le choix de h. Si la première et la seconde dérivée de k existent, alors nous pouvons estimer ces quantités par la première et la seconde dérivée de $\tilde{Q}_n(p)$, ce qui donne l'estimateur :

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

et

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

où k est un noyau d'ordre m.

L'estimation résultante de paramètre de lissage asymptotiquement optimale est donnée par :

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

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

Théorème 2.4.3 *Supposons que $Q^{(m+2)}$ est continue au voisinage de p, et que k est un noyau d'ordre m à support compact, symétrique par rapport à zéro. Alors, Le paramètre* *de lissage asymptotiquement optimale pour $\tilde{Q}'(p)$ est donnée par :*

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

et Le paramètre de lissage asymptotiquement optimale pour $\tilde{Q}''(p)$ est alors donnée par :

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

Chapitre 3

Simulation

Enfin nous terminons ce mémoire par une simulation en choisissant la distribution de **Burr** de paramétres $(1,3,1)$ et les estimateurs non paramétriques (**implicite** et **explicite),** ainsi que les noyaux "**Epanechnikov** et **Beweight"** qui sont montionnés dans le tableau 2.1, et deux échantillons de taille respective $n = 100$ et $n = 200$, et comme paramètre de lissage la fenêtre optimale de l'erreur quadratique moyenne qui est de la forme suivante :

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

Les deux tableaux ci dessous, représentent les estimateurs non paramétriques (**implicite** et **explicite)** pour la fonction des quantiles $Q\left(p\right)$ où p $(p = 0.15\ ;\ 0.2;\ 0.25;\ 0.3;\ 0.35)$ de la distribution **Burr** de paramétres $(1,3,1)$**,** en prenant des échantillons de tailles $n = 100$ et $n = 200$

	p	0.15	0.2	0.25	0.3	0.35
	$Q(p)$	0.52942	0.75	1.0001	1.2858	1.6154
$n=100$	$\tilde{Q}_n(p)$	0.6549686	0.8265662	1.0558014	1.2574023	1.6057907
$n=100$	$\hat{Q}_n(p)$	0.6666667	0.8585859	1.0000000	1.0000000	1.0000000
$n=200$	$\tilde{Q}_n(p)$	0.5710112	0.7464317	0.9875611	1.1783480	1.4577423
$n=200$	$\hat{Q}_n(p)$	0.5829146	0.8140704	0.9899497	1.0000000	1.0000000

TAB. 3.1 – le quantile d'ordre p de la distribution Burr, noyau Epanechnikov

	p	0.15	0.2	0.25	0.3	0.35
	$Q(p)$	0.52942	0.75	1.0001	1.2858	1.6154
$n=100$	$\tilde{Q}_n(p)$	0.5569641	0.7081520	0.9187345	1.0612944	1.3188551
$n=100$	$\hat{Q}_n(p)$	0.5678392	0.7777778	0.9090909	1.0000000	1.0000000
$n=200$	$\tilde{Q}_n(p)$	0.5644873	0.7359766	0.9450178	1.1748786	1.5261341
$n=200$	$\hat{Q}_n(p)$	0.5376884	0.7738693	0.9749246	1.0000000	1.0000000

TAB. 3.2 – le quantile d'ordre p de la distribution Burr, noyau Beweight

Les figures suivantes représentent les estimateurs non paramétriques (**implicite** et **explicite)** pour la fonction des quantiles $Q(p)$ où $0 \leq p \leq 1$ de la distribution **Burr** de paramétres $(1,3,1)$ **,** en prenant des échantillons de tailles $n = 100$ et $n = 200$ et ainsi que les noyaux **Epanechnikov** et **Beweight.**

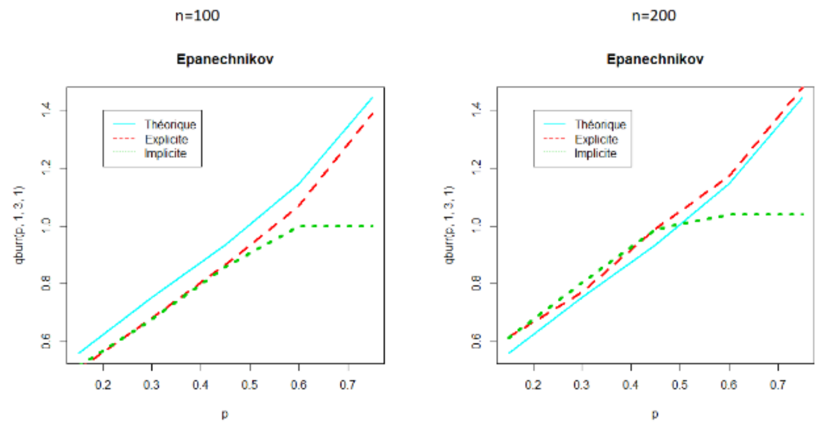

FIG. 3.1 – la fonction de quantile d'ordre p ou $0 \leq p \leq 1$, noyau Epanechnikov

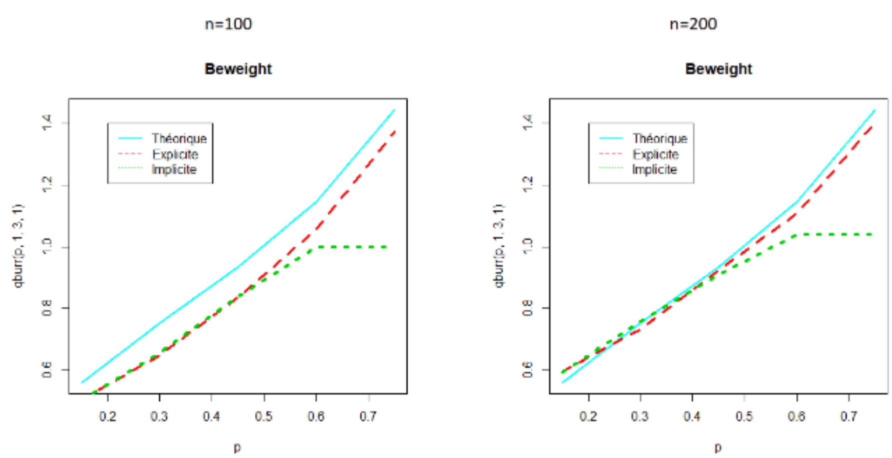

FIG. 3.2 – la fonction de quantile d'ordre p ou $0 \leq p \leq 1$, noyau Beweight

Grâce à ces résultats, on peut conclure que ces estiamteurs non paramétriques sont de bon estimateurs de la fonction des quantiles, ainsi que ces résultats nous permettent à comparer les deux différents estimateurs dont on remarque que l'estimateur à noyau est toujours meilleur que l'estimateur implicite surtout quand p croit.

Conclusion

Le travail présenté dans ce mémoire est l'estimation de la fonction des quantiles par la méthode du noyau.

Dans un premier temps, nous avons introduit la notion de la statistique d'ordre qui jouent un rôle dans l'estimation des quantiles.

Dans un second temps, nous avons donné quelques définitions nécessaires pour l'estimation des quantiles, Puis deux différentes approches ont été présentées pour cette estimation " l'approche paramétrique et l'approche non paramétrique ", nous avons présenté une étude plus détaillée sur l'approche non paramétrique et en particulier la méthode du noyau. Cette méthode nécessite le choix du noyau k et du paramètre de lissage h pour obtenir la meilleure estimation.

Enfin, pour confirmer notre étude nous avons fait des simulations (en utilisant le logiciel R) qui a validé les notions introduites précédemment.

Bibliographie

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

[2] Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (1992). A first course in order statistics. Society for Industrial and Applied Mathematics.

[3] BENAMEUR, S. (2018). Modeling of rare events for risk management (Doctoral dissertation, University Mohamed khider Biskra).

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

[5] David, H. A., Nagaraja, H. N. (2003). Order statistics, Third Edition , New York : Jhon Wiley.

[6] Epanechnikov, V. A. (1969). Nonparametric Estimation of a Multivariate Probability Density.Theory of Probability and its Applications. 14 (1) : 153-158.

[7] Falk, M. (1984). Relative Deficiency of Kernel Type Estimators of Quantiles. The Annals of Statistics. 12, 261-268.

[8] Harrell, F. E. and Davis, C. E. (1982). A New Distribution-Free Quantile Estimator. Biometrika. 69, 635-640.

[9] Jones, M. C. (1992). Estimating densities, quantile, quantile densities and density quantiles. Ann. Inst. Statist. Math., 44, 721-727.

[10] Kaigh, W. D. and Lachenbruch, P. A. (1982). A Generalized Quantile Estimator,Communications in Statistics. Part A - Theory and Methods, 11 , 2217-2238.

[11] Nadaraya, E. A. (1964). Some new estimates for distribution function. Theory of Probab. Appl. 9, 497-500.

[12] Padgett, W. J. (1986). A Kernel-Type Estimator of a Quantile Function from Right-Censored Data, Journal of the American Statistical Association. 81, 215-222.

[13] Park, C. (2006). Smooth nonparametric estimation of a quantile function under right censoring using beta kernels. Technical Report (TR 2006-01-CP), Department of Mathematical Sciences, Clemson University.

[14] Parzen, E. (1979). Nonparametric Statistical Data Modelling. Journal of the American Statistical Association. 74, 105-131.

[15] Ralescu, S. S. and Sun, S. (1993). Necessary and sufficient conditions for theasymptotic normality of perturbed sample quantiles. J. Statist. Plann. Inference. 35, 55-64.

[16] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math.Statist. 27, 832-837.

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

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

[19] Sheater, S. J. and Marron, J. S. (1990). Kernel quatile estimtors. Journal of the American Statistical Association. 85, 410-416.

[20] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function, Bull. Math. Statist. 15, 69-78.

[21] Yang, S. S. (1985). A Smooth Nonparametric Estimator of a Quantile Function. Journal of the American Statistical Association. 80, 1004-1011.

[22] Zelterman, D. (1990). Nonparametric quantile estimation. J. Statist. Plann. and inference. 26, 339-352.

Annexe A : Logiciel *R*

3.1 Qu'est-ce-que le langage R ?

• Le langage **R** est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• **R** a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

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

$i.i.d$	:	**i**ndépendantes et **i**dentiquement **d**istribuées.
MSE	:	**M**ean **S**quare **E**rror (erreur quadratique moyenne).
$AMSE$	:	**A**symptotic **M**ean **S**quare **E**rror (erreur quadratique moyenne asymptotique).
F	:	Fonction de distribution.
F_n	:	Fonction de distribution empirique.
F^{-1}	:	Inverse de la fonction de distribution.
$F^{\leftarrow}$	:	Inverse généralisé de la fonction de distribution.
Q	:	fonction de quantile.
Q_n	:	fonction de quantile empirique.
x_F	:	point terminale.
$E[X]$	:	**E**spérance mathématique ou moyenne du v.a X .

$Var[X]$	:	**V**ariance de X .
$\mathbf{1}_A$	:	La fonction de l'indécatrice sur A .
Γ	:	Fonction de Gamma .
$\mathbf{B}$	:	Fonction de Beta.
$\mathbb{R}$	:	Ensemble des réels.
$\mathbb{N}$	:	Ensemble des entiers naturels.
Θ	:	Espace des paramétères.
$\xrightarrow{ps}$	:	convergence presque sûre.
$\xrightarrow{L}$	:	convegence en loi .
HD	:	Estimateur de **H**arrel-**D**avis.

Résumé :

Ce mémoire présente l’étude de la méthode d’estimation à noyau de la fonction des quantiles. Nous avons donné un aperçu sur l’estimation paramétrique et non paramétrique des quantiles et principalement la méthode du noyau. Cette méthode est basée sur l’utilisation d’un noyau k et d’un paramètre de lissage h.

Enfin, nous avons fait des simulations (en utilisant le logiciel **R)** et on a prouvé d’après ces résultats de simulation que ces estimateurs sont de bon estimateurs de la fonction des quantiles.

Mots Clés : Fonction des quantiles, estimation paramétrique, estimation non paramétrique, noyau, paramètre de lissage.

ملخص:

تقدم هذه المذكرة دراسة طريقة تقدير النواة للدالة الكمية. قدمنا نظرة عامة على التقدير الوسيطي واللاوسيطي للكميات وبشكل أساسي طريقة النواة. تعتمد هذه الطريقة على استخدام النواة k و وسيط المسح h.

أخيرًا ، أجرينا عمليات محاكاة (باستخدام برنامج R) وثبت من نتائج المحاكاة هذه أن هذه المقدرات هي تقديرات جيدة للوظيفة الكمية.

الكلمات المفتاحية : الدالة الكمية ، التقدير الوسيطي، التقدير اللاوسيطي، نواة،وسيط المسح.

Abstract:

This work presents the study of the kernel estimation method of the quantile function. We gave an overview on the parametric and non-parametric estimation of quantiles and mainly the kernel method. This method is based on the use of a kernel **k** and a smoothing parameter **h**.

Finally, we performed simulations (using R software) and it was proved from these simulation results that these estimators are good estimators of the quantile function.

Keywords: Quantile function, parametric estimation, nonparametric estimation, kernel, smoothing parameter.