

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE.
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE RECHERCHE SIENTIFIQUE.

Université Mouloud Mammeri, Tizi-Ouzou
Faculté des Sciences
Département de Mathématiques

Mémoire de Master en mathématiques appliquées
Option: *Processus Aléatoires et Statistique de la Décision*

Thème

Estimation d'une fonction de densité par la méthode des noyaux et application à la VaR

Présenté par
FERHAT Lamia

Devant le jury d'examen composé de :

M. Boudiba Mohand Arezki	Maître de conférence A	UMMTO	Président
M. Berkoun Youcef	Maître de conférence A	UMMTO	Rapporteur
M. Hamadouche Djamel	Professeur	UMMTO	Examineur
M. Mehiri Mohamed	Chargé de recherches	UMMTO	Examineur

Soutenu le 07 / 10 / 2012

Remerciements

*Je tiens à exprimer toute ma reconnaissance à **Mr Berkoun Youcef** pour l'honneur qu'il m'a fait en assurant la direction et le suivi scientifique et technique du présent mémoire. Je le remercie pour sa grande contribution à l'aboutissement de ce travail, et pour sa disponibilité.*

*Je remercie vivement **Mr Boudiba Mohand Arezki** pour l'honneur qu'il me fait en acceptant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adressent également à **Mr Hamadouche Djamel et Mr Mehiri Mohamed** pour avoir accepté d'examiner ce travail.*

Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Enfin, je tiens à remercier ma famille et particulièrement mes parents pour leur affection et leur soutien tout au long de mes études.

Table des matières

Introduction Générale	7
1 Estimation non paramétrique d'une densité	9
1.1 Estimation par la méthode de l'histogramme	9
1.2 Propriétés de l'estimateur par histogramme	10
1.2.1 Étude du biais	10
1.2.2 Étude de la variance	11
1.2.3 L'erreur quadratique moyenne (MSE)	11
1.2.4 L'erreur quadratique moyenne intégrée (MISE)	12
1.3 Estimation par la méthode des noyaux	13
1.4 Noyaux	14
1.4.1 Exemples de noyaux	15
1.4.2 Noyaux d'ordre r	15
1.5 Estimateur à noyau	16
1.6 Propriétés de l'estimateur à noyau	17
1.6.1 Etude du biais	17
1.6.2 Etude de la variance	18
1.6.3 Erreur quadratique moyenne (Mean-Squared Error:MSE)	20
1.6.4 Erreur quadratique moyenne intégrée (Mean-Integrated Squared Error:MISE)	21
1.6.5 Choix théorique optimal du paramètre de lissage	24
1.6.6 Le choix optimal pratique du paramètre de lissage h	26
1.6.7 Simulation des données	33
1.6.8 Le choix optimal du noyau	36
1.7 Estimateur à noyau dans le cas multivarié	38
2 La valeur de risque (La VaR)	41
2.1 Les mesures de risque	41
2.2 Caractéristiques d'une mesure de risque	41
2.2.1 Mesure de risque cohérente	42
2.2.2 Mesure de risque monétaire	42
2.3 Mesures de risque usuelles	42
2.3.1 La variance	42
2.4 Value-at-Risk	42
2.4.1 Présentation théorique de la méthode VaR	43
2.4.2 Présentation mathématique de la méthode VaR	43
2.4.3 Propriétés de la VaR	45

3	Estimation d'un quantile par la méthode des noyaux	49
3.1	Les estimateurs d'un quantile	49
3.2	Inverse d'estimateur à noyau d'une fonction de répartition	50
3.3	L'estimateur à noyau d'un quantile	50
3.4	Propriétés asymptotiques de $Q_n(p)$ et les estimateurs relatifs	51
3.5	Sélection du paramètre de lissage optimal	53
3.6	La normalité asymptotique de l'estimateur à noyau d'un quantile	53
3.7	Représentation de Bahadur	54
	Conclusion	57
	Bibliographie	61

Table des figures

1.1	Différents histogrammes associés à un même ensemble de données avec différents paramètres de lissages	13
1.2	Les courbes des noyaux les plus communs	15
1.3	Estimateur à noyau basé sur 7 observations ($h=4$)	17
1.4	Le "trad-off" biais-variance en fonction de h	20
1.5	Illustration d'un phénomène de sous-lissage lors de l'estimation d'une densité	25
1.6	Illustration d'un phénomène de sur-lissage lors de l'estimation d'une densité	26
1.7	Illustration d'une estimation idéale.	26
1.8	L'expression de h_{NR} pour quelques noyaux.	29
1.9	Représentation d'un estimateur à noyau en utilisant h_{NR}	29
1.10	Choix du h par la méthode de validation croisé.	31
1.11	Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon de loi normale centrée réduite, $n = 100$ et $h_{PI} = 0.338$	34
1.12	Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon de loi normale centrée réduite, $n = 100$ et $h_{CV} = 0.42$	35
1.13	Comparaison des lissages par l'estimateur à noyau continu d'Epanchnikov en faisant varier la fenêtre h	35
1.14	Quelques noyaux et leurs efficacité.	37
1.15	Estimation de la densité avec différents noyaux.	38
1.16	Densité d'un noyau d'Epanechnikov dans le cas multivarié.	39
2.1	Value-at-Risk de la somme de deux v.a. de Pareto.	47

Introduction Générale

L'estimation d'une densité de probabilité sous-jacente à un ensemble fini d'observations est un problème fondamental en statistique qui a fait l'objet d'une vaste littérature. On retrouve cette problématique dans de nombreux domaines des sciences et techniques tels que le traitement du signal et des images, la mécanique, la robotique, etc.

La modélisation de phénomènes aléatoires est traditionnellement réalisée par une distribution de probabilité qui est généralement inconnue. On doit donc l'identifier à partir d'un nombre fini d'observations, identification qui consiste généralement à estimer la densité de probabilité en tout point de l'espace considéré. La tâche de l'estimation de densité revient à produire, à partir d'un ensemble de données, un modèle probabiliste.

On trouve deux types d'approches d'estimation de la densité de probabilité : estimation paramétrique et non paramétrique. L'approche paramétrique a comme inconvénient principal la connaissance au préalable de la loi du phénomène étudié. L'approche non paramétrique estime la densité à partir de l'information disponible. On dit souvent que dans cette approche les données parlent d'elles mêmes. L'avantage principal de l'estimation non paramétrique de la densité de probabilité est de ne pas nécessiter d'hypothèses a priori sur l'appartenance de cette densité à une famille de lois connues.

Si la qualité de l'estimation paramétrique est fortement liée à la validité de l'hypothèse faite sur la loi de probabilité, celle de l'approche non paramétrique dépend du nombre d'observations et de certains paramètres (noyau, bande lissage, etc.).

Nous nous intéressons ici plutôt aux approches non paramétriques : approche basée sur l'histogramme et approche basée sur le noyau. L'estimation non paramétrique de la densité de probabilité nous permettra de donner un estimateur d'un quantile par la méthode des noyaux (estimateur appelé communément kernel quantile estimator).

L'estimation d'un quantile est d'un grand intérêt quand on n'assume aucune forme paramétrique de la densité sous-jacente. Une des applications principales en théorie du risque est l'estimation de la value at risk.

Ce travail s'articule autour de 3 chapitres.

Le premier chapitre définit en premier lieu la méthode d'estimation par histogramme, puis la méthode d'estimation par noyau, ainsi que leurs propriétés asymptotiques. Le principe de base de la méthode des noyaux s'apparente relativement bien à la notion d'histogramme couramment utilisé pour l'analyse exploratoire d'un échantillon. L'histogramme donne une idée de la forme de la distribution empirique d'un échantillon en calculant la proportion d'observations se trouvant dans chacun des intervalles de largeur h . Le choix de la largeur h des fenêtres de l'histogramme est déterminant. Le concept de fenêtre de l'histogramme est aussi présent dans la méthode des

noyaux. On utilise alors le terme du "paramètre de lissage" pour désigner la fenêtre h . Comme dans le cas de l'histogramme, l'estimation de la fonction de densité par la méthode des noyaux est principalement conditionnée par le paramètre de lissage. L'importance du paramètre de lissage dans l'estimation par la méthode des noyaux a eu pour effet de concentrer des recherches sur des techniques de calcul du paramètre de lissage conduisant ainsi à une variété importante de méthodes; on a essayé d'identifier quelques méthodes de sélection du paramètre de lissage les plus prometteuses pour l'estimation de la densité avec quelques applications sous R.

Le deuxième chapitre définit la notion de mesure de risque et ses propriétés; parmi les mesures de risque usuelles on s'intéressera plus particulièrement à la VaR qui fera l'objet de notre travail.

Dans le troisième chapitre, on s'intéresse à l'estimation du quantile par la méthode des noyaux, ainsi qu'à ses propriétés asymptotiques.

Chapitre 1

Estimation non paramétrique d'une densité

L'approche non paramétrique prend son sens lorsqu'on ne possède aucune information précise sur la forme et la classe de la vraie densité. Dans cette approche, ce sont les observations qui vont nous permettre de déterminer l'estimation de la densité f . Cette approche sera utilisée puisque l'on ne peut supposer aucune forme de fonction pré spécifiée pour f .

1.1 Estimation par la méthode de l'histogramme

L'estimateur non paramétrique d'une densité le plus utilisé est l'histogramme. Ce dernier est habituellement formé par la division de la droite réelle en intervalles de même taille.

L'histogramme est une fonction étagée où les hauteurs correspondent à la proportion de l'échantillon contenue dans l'intervalle divisé par la largeur de celui-ci. Supposons que h correspond à la largeur des intervalles, (h est appelé le paramètre de lissage) et que n correspond à la taille de l'échantillon.

Soit l'échantillon X_1, \dots, X_n issu d'une variable aléatoire X de densité f

- On sélectionne un origine x_0 et on divise la droite réelle en intervalles de même taille :

$$B_j = [x_0 - (j-1)h, x_0 + (j-1)h] \quad j \in \mathbf{Z}$$

- Soit n_j nombre d'observations appartenant à B_j

- Soit $f_j = \frac{n_j}{n}$ et $f_j = \int_{B_j} f(u) du$

L'estimateur de f par la méthode de l'histogramme est donné par

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n \sum_j I_{(X_i \in B_j)} I_{(x \in B_j)}$$

Si m_j est le centre de B_j , il est clair que l'histogramme assigne le même estimateur $f_n(m_j)$ pour tout $x \in [m_j - \frac{h}{2}, m_j + \frac{h}{2}]$

La probabilité qu'une observation de X appartienne à l'intervalle $[m_j - \frac{h}{2}, m_j + \frac{h}{2}]$ est donnée par :

$$P(X \in [m_j - \frac{h}{2}, m_j + \frac{h}{2}]) = \int_{m_j - \frac{h}{2}}^{m_j + \frac{h}{2}} f(u) du$$

L'estimation de la densité par l'histogramme au point x est donnée par:

$$f_n(x) = \frac{1}{nh} \# \{X_i \in [m_j - \frac{h}{2}, m_j + \frac{h}{2}]\}$$

Avant de construire l'histogramme, on doit choisir le positionnement des intervalles ainsi que le paramètre de lissage h . Chacun de ces choix peut avoir un effet significatif sur l'histogramme résultant, ce qui constitue une difficulté.

Une autre difficulté de la méthode de l'histogramme est que l'on estime toutes les densités par des fonctions étagées ce qui n'est pas toujours le cas. De plus, cet estimateur de la densité est une fonction discontinue, donc l'application de certaines opérations sur l'estimé, comme par exemple une dérivée ou une intégration, devient impossible ou très difficile à effectuer.

1.2 Propriétés de l'estimateur par histogramme

On étudiera quelques propriétés statistiques de l'estimation par la méthode de l'histogramme. On commence par la présentation du concept de l'erreur quadratique moyenne qui est définie comme suit:

$$MSE(f_n(x)) = E[f_n - f(x)]^2$$

qui s'écrit aussi sous la forme suivante:

$$MSE(f_n) = Var(f_n(x)) + Biases^2(f_n(x))$$

où

$$Var(f_n) = E[(f_n(x))^2] - (E[f_n(x)])^2$$

et

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

1.2.1 Étude du biais

Supposons que l'on dispose d'un échantillon d'observations $X_1 \dots, X_n$, provenant d'une distribution univariée continue et possédant pour fonction de densité la fonction f que l'on désire estimer. Pour $x \in B_j$ fixe, on a:

$$\begin{aligned} E(f_n(x)) &= \frac{1}{nh} \sum_{i=1}^n E[I_{(X_i \in B_j)}] \\ &= \frac{1}{n} E[I_{(X_1 \in B_j)}] \end{aligned}$$

On note que $I_{(X_1 \in B_j)} \sim B(p)$ tel que $p = \int_{(j-1)h}^{jh} f(u)du$

$$I_{(X_1 \in B_j)} = \begin{cases} 1, & \text{avec la probabilit } p \\ 0, & \text{avec la probabilit } 1 - p \end{cases}$$

On a $E[I_{(X_1 \in B_j)}] = \int_{(j-1)h}^{jh} f(u) du$
ce qui veut dire que f_n est biaisé et son biais est:

$$\begin{aligned} \text{Biais} &= \frac{1}{h} \int_{(j-1)h}^{jh} f(u) du - f(x) \\ &= \frac{1}{h} \int_{(j-1)h}^{jh} f(u) - f(x) du \end{aligned}$$

En appliquant le développement de Taylor pour $f(u) - f(x)$ au point x on a:

$$\text{Biais} = f'(x) \left(h(j - \frac{1}{2}) - x \right) + o(h) \quad (1.1)$$

1.2.2 Étude de la variance

Calculons la variance de l'estimateur f_n

$$\text{Var}(f_n(x)) = \text{Var} \left(\frac{1}{nh} \sum_{i=1}^n I(X_i \in B_j) \right)$$

$$\begin{aligned} \text{Var}(f_n(x)) &= \frac{1}{n^2 h^2} \sum_{i=1}^n \text{Var}(I(X_i \in B_j)) \\ &= \frac{1}{nh^2} \text{Var}(I(X_i \in B_j)) \\ &= \frac{1}{nh^2} \int_{(j-1)h}^{jh} f(u) du \left(1 - \int_{(j-1)h}^{jh} f(u) du \right) \end{aligned}$$

$$\text{Var}(f_n(x)) = \frac{1}{nh} f(x) + o\left(\frac{1}{nh}\right) \quad (1.2)$$

On remarque que la variance diminue si h augmente et inversement pour le biais.

1.2.3 L'erreur quadratique moyenne (MSE)

En utilisant (1.1) et (1.2) on trouve:

$$\text{MSE}(f_n(x)) = f'(x)^2 \left(h(j - \frac{1}{2}) - x \right)^2 + \frac{1}{nh} f(x) + o(h) + o\left(\frac{1}{nh}\right) \quad (1.3)$$

On conclut que le MSE converge si $h \rightarrow 0$ et $nh \rightarrow \infty$

1.2.4 L'erreur quadratique moyenne intégrée (MISE)

$$MISE = \int_{-\infty}^{+\infty} MSE(f_n) dx$$

En utilisant (1.3), on peut écrire :

$$MISE(f_n) = \int_{\mathbb{R}} \frac{1}{nh} + \int_{\mathbb{R}} \sum_j I(x \in B_j) f'(x) \left(h(j - \frac{1}{2}) - x \right) \quad (1.4)$$

$$= \frac{1}{nh} \sum_j f'(x)^2 \left(h(j - \frac{1}{2}) - x \right)^2 dx \quad (1.5)$$

$$= \frac{1}{nh} + \sum_j \int_{B_j} f'(x)^2 dx \int_{B_j} \left(h(j - \frac{1}{2}) - x \right)^2 dx \quad (1.6)$$

$$= \frac{1}{nh} + \frac{h^2}{12} \int_{\mathbb{R}} f'(x)^2 dx \quad (1.7)$$

En posant $C(f) = \int f'(x)^2 dx$, on a :

$$MISE(f_n) = \frac{1}{nh} + \frac{h^2 C(f)}{12}$$

Le paramètre de lissage optimal est celui qui minimise le MISE et qui est donné par :

$$\frac{\partial MISE(f_n)}{\partial h} = \frac{-1}{nh^2} + \frac{1}{6} C(f) = 0$$

Par conséquent :

$$h_{opt} = \left(\frac{6}{C(f)} \right)^{1/3} n^{-1/3} \quad (1.8)$$

Exemple 1. On dispose des données de 709 modèles d'avion du 20^e siècle. On désire estimer la densité de l'envergure de ces avions par la méthode de l'histogramme en utilisant différents paramètres de lissage :

Remarque :

Dans ce travail on utilise le langage R pour la simulation des données. Accéder aux données `data(aircarf)` avec la commande `provide.data(aircarf)` nécessite l'installation du package (`sm`).

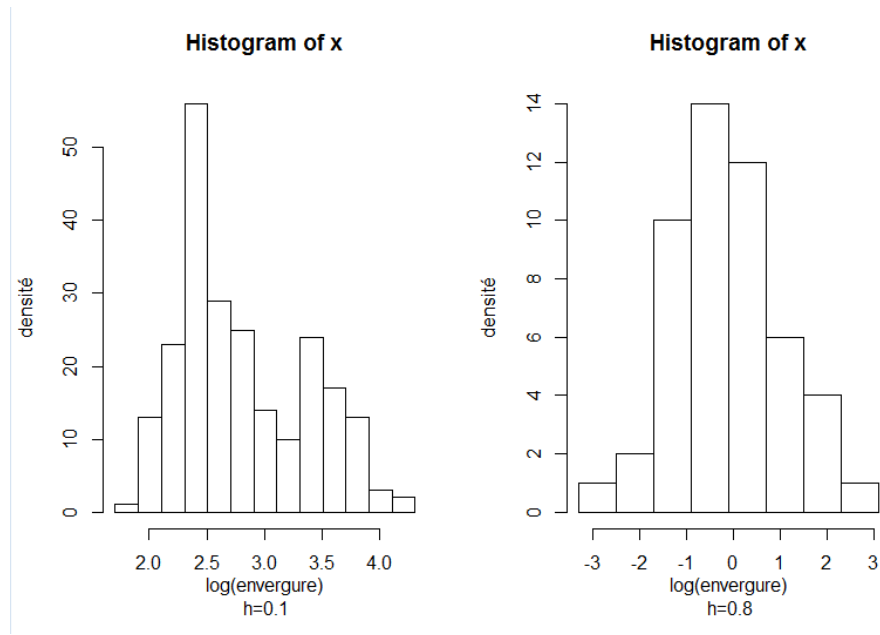


FIG. 1.1 – Différents histogrammes associés à un même ensemble de données avec différents paramètres de lissages

La figure (1.1) nous montre deux histogrammes basés sur le même ensemble de données et avec deux paramètres de lissage petit et grand. Un paramètre de lissage trop petit conduit à un histogramme plus découpé, tandis que l'autre donne un histogramme plus lissé comme le montre la figure (1.1)

A l'exception du choix du paramètre de lissage, ces difficultés peuvent être évitées si l'on utilise la méthode d'estimation de la densité par le noyau. C'est la raison pour laquelle nous utiliserons cet estimateur plutôt que l'histogramme dans ce travail. Une étude approfondie de celui-ci est présentée dans la section suivante.

1.3 Estimation par la méthode des noyaux

Le concept de noyau a d'abord été introduit par ROSENBLATT(1956), mais c'est CACOULOS(1966) qui a été le premier à utiliser le terme "noyau" pour désigner la fonction que l'on utilise dans les méthodes non paramétriques. En hydrologie, c'est YAKOTZ(1983) et FELUCH(1983) qui ont introduit indépendamment la méthode des noyaux lors d'une conférence de l'AGU à l'automne 1983.

Dans la méthode des noyaux, une fonction $K(x)$ est associée à chaque observation de l'échantillon. La seule véritable restriction concernant le noyau K est que son intégration sur tout le domaine de définition de x doit être égale à un. On rencontre parfois d'autres restrictions théoriques qui sont appliquées à K , comme la symétrie ou la positivité sur tout le domaine de définition du noyau (ADAMOWSKI, 1989). Toutefois, ces restrictions sont surtout introduites afin de simplifier les développements théoriques. L'estimation non paramétrique de la fonction de densité peut se voir comme le cumul des fonctions K de chaque observation sur tout le domaine:

Supposons que nous observons n variables aléatoires *i.i.d* X_1, \dots, X_n de densité f . L'objectif de notre étude est la construction d'un estimateur de f en un point fixe x .

Notons $F(x) = \mathbf{P}(X_1 \leq x)$ la fonction de répartition de la loi de X_1 .

La densité est la dérivée de la fonction de répartition, ce qui permet d'écrire pour tout x :

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

Considérons la fonction de répartition empirique:

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

La loi des grands nombres permet d'affirmer que F_n est un estimateur de F , c'est-à-dire

$$F_n(x) \xrightarrow{p} F(x) \tag{1.9}$$

De plus, le théorème de Glivenko-Cantelli nous donne:

$$\sup_{x \in \mathbf{R}} |F_n(x) - F(x)| \xrightarrow{a.s} 0 \tag{1.10}$$

Il est même possible d'obtenir des intervalles de confiance et de tester l'adéquation des données à différentes lois. Néanmoins, il n'est pas évident d'utiliser F_n pour estimer f .

Une des premières idées intuitives est de considérer pour $h > 0$ fixé "petit":

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{\{-h \leq X_i - x \leq h\}}.$$

On a alors:

$$E[f_n(x)] = \frac{1}{2h} (E[F_n(x+h)] - E[F_n(x-h)]) = \frac{1}{2} (F(x+h) - F(x-h))$$

$E[f_n(x)]$ tend vers $f(x)$ quand $h \rightarrow 0$. Il faut donc faire dépendre h de la taille de l'échantillon, et le faire tendre vers 0 quand $n \rightarrow \infty$, de sorte que $f_n(x)$ soit un estimateur asymptotiquement sans biais de $f(x)$.

L'estimateur f_n reste une fonction en escalier. Pour obtenir quelque chose de plus lisse, on peut remarquer que:

$$f_n(x) = \frac{1}{2nh_n} \sum_{i=1}^n \mathbf{1}_{]x-h_n, x+h_n]}(X_i) = \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \mathbf{1}_{\{x-h_n < X_i \leq x+h_n\}} \tag{1.11}$$

$$= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \mathbf{1}_{[-1,1]} \left(\frac{x - X_i}{h_n} \right) = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x - X_i}{h_n} \right) \tag{1.12}$$

Cet estimateur appelé estimateur de Rosenblatt, est le premier exemple d'estimateur à noyau construit à l'aide du noyau $K(u) = \frac{1}{2} \mathbf{1}_{\{-1 < u \leq 1\}}$.

1.4 Noyaux

Définissons maintenant plus généralement la notion d'estimateur à noyau:

Définition 1. Soit $K : \mathbf{R} \rightarrow \mathbf{R}$, on dit que K est un noyau si et seulement si:

$$\int K(u) du = 1$$

- K est dit positif si $K(u) \geq 0 \forall u$.
- K est dit symétrique si $K(u) = K(-u) \forall u$

1.4.1 Exemples de noyaux

Voici quelques exemples de noyaux les plus communément utilisés:

- le **noyau rectangulaire**: $K(u) = \frac{1}{2}\mathbf{1}_{[-1,+1]}(u)$. C'est celui qui donne l'estimateur de type histogramme appelé **noyau de Rosenblatt**.
- le **noyau triangulaire**: $K(u) = (1 - |u|)\mathbf{1}_{[-1,+1]}(u)$
- le **noyau d'Epanechnikov**: $\frac{3}{4}(1 - u^2)\mathbf{1}_{[-1,+1]}(u)$
- le **noyau de Tukey ou biweight**: $K(u) = \frac{15}{16}(1 - u^2)^2\mathbf{1}_{[-1,+1]}(u)$
- le **noyau gaussien**: $K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-u^2/2}, u \in \mathbb{R}$

Les deux premiers ont l'avantage d'être simples, le noyau triangulaire étant continu partout et conduisant à une estimation f_n continue. Le troisième doit sa notoriété à une propriété d'optimalité théorique mais sans grand intérêt pratique. Le quatrième est, à notre sens, le plus intéressant car donnant une estimation dérivable partout, tout en étant simple à mettre en oeuvre. En fait il s'agit du noyau le plus simple parmi les noyaux de forme polynomiale dérivables partout. Ainsi il assure le lissage local de la fonction f_n . Ce noyau est d'une forme très proche du noyau Gaussien et il est donc préférable. Notons que plus la valeur de h est élevée plus on élargit la fenêtre, ce qui donne un effet de lissage globale de f_n plus important.

Voici quelques courbes de noyaux usuels présentées ci-dessous :

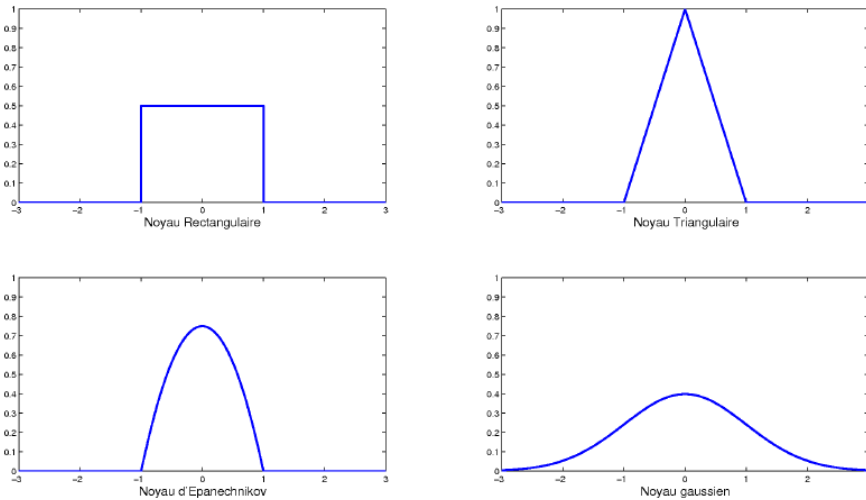


FIG. 1.2 – Les courbes des noyaux les plus communs

1.4.2 Noyaux d'ordre r

Définition 2. Soit $r \geq 1$ un entier. On dit qu'un noyau K est d'ordre r si:

$$\forall j = 1, \dots, r, \int u^j K(u) du = 0 \text{ et } \int u^{r+1} K(u) du \neq 0.$$

1.5 Estimateur à noyau

L'estimateur à noyau est probablement l'estimateur le plus utilisé et certainement le plus étudié mathématiquement, car il possède des propriétés qui le rendent fort intéressant.

Définition 3. Un estimateur à noyau noté f_n de la fonction f est défini par:

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \quad (1.13)$$

où $\{h_n\}_{n \geq 1}$ est une suite de réels positifs appelés **paramètres de lissage** ou **largeur de la fenêtre**, qui tend vers 0 quand n tend vers l'infini.

Comme nous allons le voir par la suite, si le noyau K est une fonction de densité alors l'estimateur à noyau f_n est lui aussi une fonction de densité. De plus, ce dernier possède les propriétés de continuité et de différentiabilité. De sorte que si, par exemple, K est la densité normale alors f_n possède des dérivées de tout ordre.

Propriété 1. Un estimateur à noyau est une densité

Démonstration:

$$\begin{aligned} \int_{-\infty}^{+\infty} f_n(x) dx &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x - X_i}{h_n}\right) dx \\ &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) h_n du \quad (\text{changement de variable } u = \left(\frac{x - X_i}{h_n}\right)) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) du = \frac{1}{n} n = 1 \end{aligned}$$

Pour mieux saisir l'intuition derrière l'estimateur à noyau, nous avons construit cet estimateur à partir de l'équation (1.13) en utilisant un ensemble de données constitué seulement de 7 observations. Le noyau K a été choisi comme étant la densité d'une loi normale de moyenne 0 et de variance 1 et le paramètre de lissage h égale à 4. On centre d'abord un noyau individuel sur chacune des 7 observations et la valeur de l'estimateur à noyau $f(z)$ au point z est simplement la somme des ordonnées de chacun des 7 noyaux individuels à ce point x comme représenté à la figure (1.3). Dans une région où l'on a plusieurs observations, la vraie densité a une valeur relativement grande et l'estimateur de la densité, par la méthode du noyau, nous donne effectivement une valeur relativement grande ce qui est observé dans la figure (1.3)

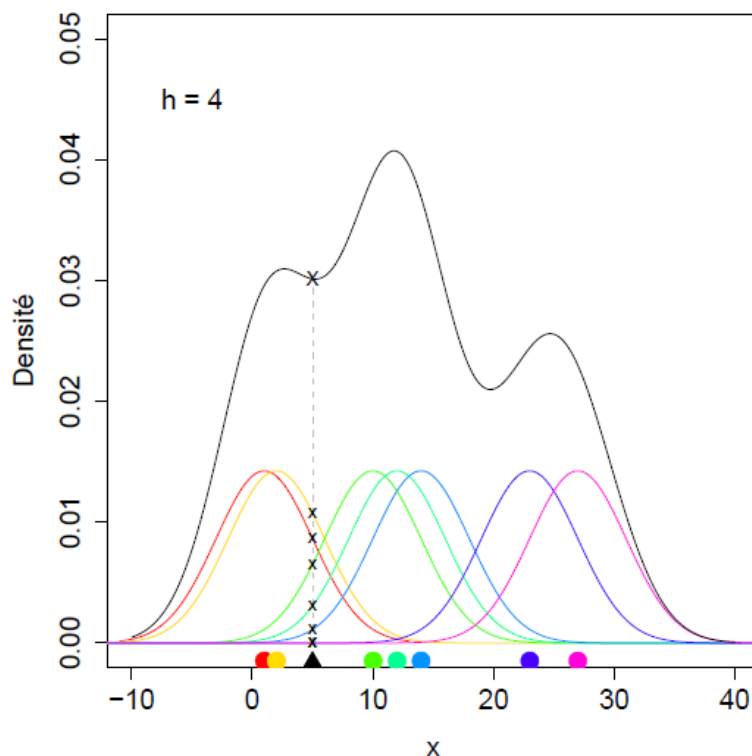


FIG. 1.3 – Estimateur à noyau basé sur 7 observations ($h=4$)

Par exemple si $x = 5$ on a $\hat{f}(x) = 0.03$ qui est égale à la somme des densités des 7 noyaux gaussiens au même point $x = 5$

1.6 Propriétés de l'estimateur à noyau

Nous allons maintenant donner quelques propriétés statistiques élémentaires de l'estimateur de la densité à noyau ainsi que différentes méthodes pour choisir le paramètre de lissage.

1.6.1 Etude du biais

Supposons que l'on dispose d'un échantillon d'observations $X_1 \dots, X_n$, issu d'une v.a X possédant pour fonction de densité la fonction f que l'on désire estimer. On suppose que f_n est l'estimateur à noyau obtenu en utilisant le noyau K et le paramètre de lissage h et f_n défini par l'équation (1.13). Supposons que:

$$K(u) \geq 0, \int K(u)du = 1, \int K(u)udu = 0, \int u^2K(u)du < \infty$$

et en supposant que la densité de probabilité f admet les deux premières dérivées (continues) nécessaire.

$$\begin{aligned}
E[f_n(x)] &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} E \left[K \left(\frac{x - X_i}{h} \right) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{x - t}{h} \right) f(t) dt \\
&= \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{x - t}{h} \right) f(t) dt
\end{aligned}$$

La transformation $z = \frac{x-t}{h}$, i.e. $t = -hz + x, |dz| = \frac{1}{h}$

$$E[f_n(x)] = \int_{-\infty}^{+\infty} K(z) f(x - hz) dz$$

Un développement de Taylor de $f(x - hz)$ nous donne:

$$f(x - hz) = f(x) - hz f'(x) + \frac{1}{2} (hz)^2 f''(x) + o(h^2)$$

$$\begin{aligned}
E[f_n(x)] &= \int_{-\infty}^{+\infty} K(z) f(x) dz - \int_{-\infty}^{+\infty} K(z) hz f'(x) dz + \int_{-\infty}^{+\infty} K(z) \frac{(hz)^2}{2} f''(z) dz + o(h^2) \\
&= f(x) \int_{-\infty}^{+\infty} K(z) dz - h f'(x) \int_{-\infty}^{+\infty} z K(z) dz + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} z^2 K(z) dz + o(h^2) \\
&= f(x) + \frac{h^2}{2} k_2 f''(x) + o(h^2)
\end{aligned}$$

$$Biais(f_n(x)) \approx \frac{h^2}{2} k_2 f''(x) + o(h^2) \tag{1.14}$$

Le biais dépend de:

- h : paramètre de lissage.
- k_2 : la variance du noyau.
- $f''(x)$ la seconde dérivée de la fonction de densité au point x .

1.6.2 Etude de la variance

La variance de $f_n(x)$ est donnée par:

$$\begin{aligned}
Var(f_n(x)) &= Var \left(\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) \right) \\
&= \frac{1}{n^2 h^2} \sum_{i=1}^n Var \left(K \left(\frac{x - X_i}{h} \right) \right)
\end{aligned}$$

car les $X_i, i=1, \dots, n$, sont i.i.d

$$\begin{aligned} \text{Var} \left(K \left(\frac{x - X_i}{h} \right) \right) &= E \left[K \left(\frac{x - X_i}{h} \right)^2 \right] - \left(E \left[K \left(\frac{x - X_i}{h} \right) \right] \right)^2 \\ &= \int K \left(\frac{x - X_i}{h} \right)^2 f(t) dt - \left(\int K \left(\frac{x - t}{h} \right) f(t) dt \right)^2 \end{aligned}$$

$$\begin{aligned} \text{Var}(f_n(x)) &= \frac{1}{n} \int \frac{1}{h^2} K \left(\frac{x - t}{h} \right)^2 f(t) dt - \frac{1}{n} \left(\int K \left(\frac{x - t}{h} \right) f(t) dt \right)^2 \\ &= \frac{1}{n} \int \frac{1}{h^2} K \left(\frac{x - t}{h} \right)^2 f(t) dt - \frac{1}{n} (f(x) + \text{Biais}(f_n(x)))^2 \end{aligned}$$

En effectuant le changement de variable suivant $z = \frac{x-t}{h}$, on obtient:

$$\text{Var}(f_n(x)) = \frac{1}{nh} \int K(x)^2 f(x - hz) dz - \frac{1}{n} (f(x) + o(h^2))^2$$

Et en effectuant un développement limité à l'ordre 2, il vient:

$$\text{Var}(f_n(x)) = \frac{1}{nh} \int K(z)^2 (f(x) - hz f'(x) + o(h)) dz - \frac{1}{n} (f(x) + o(h^2))^2$$

$$\text{Var}(f_n(x)) = \frac{1}{nh} f(x) \int K^2(z) dz + o\left(\frac{1}{nh}\right)$$

d'où:

$$\text{Var}(f_n(x)) \approx \frac{1}{nh} f(x) \int K^2(z) dz$$

•Discussion du comportement du biais et de la variance:

- * Le biais décroît si h diminue mais la variance augmente.
- * La variance diminue si h augmente mais le biais augmente.
- * Pour que la variance tende vers zéro, il faut que $nh \rightarrow \infty$.
- * Plus la courbure de la densité est haute en x , plus le biais est grand.
- * La variance est plus grande pour des valeurs plus grandes de la densité.

La figure suivante nous permet de mieux voir le comportement du biais et de la variance.

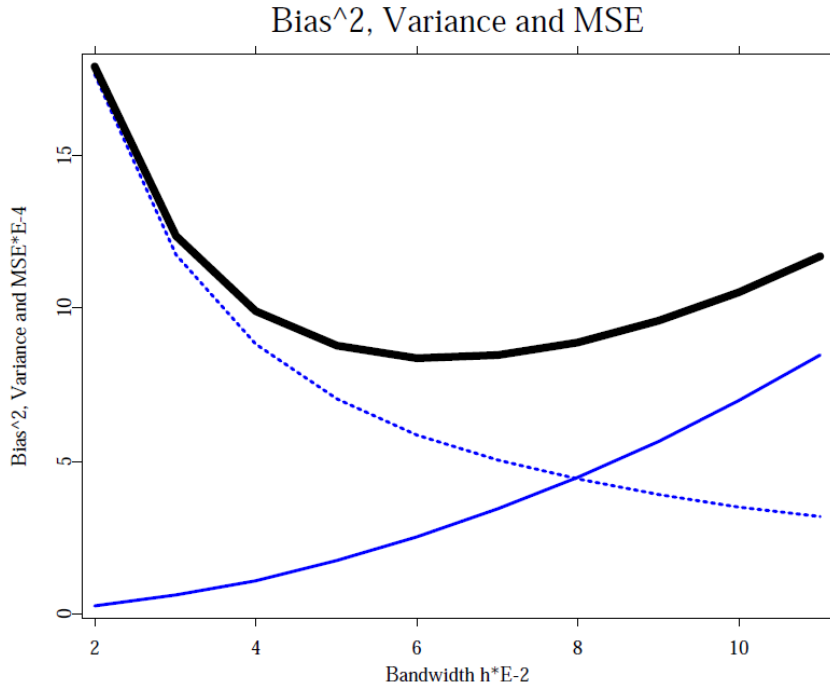


FIG. 1.4 – Le "trad-off" biais-variance en fonction de h .

La variance est représentée par la courbe en pointillé et le biais par la courbe fine, la courbe en gras représente le MSE.

1.6.3 Erreur quadratique moyenne (Mean-Squared Error:MSE)

L'analyse de la performance de l'estimateur à noyau exige la spécification d'un critère d'erreur approprié afin de mesurer l'erreur d'estimation aussi bien qu'en un point que sur l'ensemble des points. Nous étudierons dans un premier temps la proximité de notre estimateur de la densité f_n de la vraie densité f . L'estimateur f_n dépend des données, du noyau K et du paramètre de lissage h . Cette dépendance n'est généralement pas exprimée explicitement. Pour chaque x , $f_n(x)$ peut être considérée comme une variable aléatoire. Lorsque nous considérons l'estimation en un point, une mesure naturelle de la dispersion est l'erreur quadratique moyenne aussi notée MSE qui est:

$$MSE(f_n(x)) = E[(f_n(x) - f(x))^2]$$

Par des propriétés élémentaires de l'espérance et de la variance, on peut montrer que le MSE peut s'écrire.

$$\begin{aligned} MSE(f_n(x)) &= E[f_n(x) - E[f_n(x)]] + (E[f_n(x)] - f(x))^2 \\ &= (E[f_n(x)] - f(x))^2 + Var[f_n(x)] \end{aligned}$$

En remplaçant le biais et la variance par leurs valeurs respectives, on trouve:

$$MSE = \frac{1}{4}h^4k_2^2f''(x)^2 + \frac{1}{nh}f(x)j_2 + o\left(h^4 + \frac{1}{nh}\right)$$

Où $k_2 = \int zK(z)dz$ et $j_2 = \int K(z)^2dz$.

Asymptotiquement, on a:

$$AMSE = \frac{1}{4}h^4k_2^2f''(x)^2 + \frac{1}{nh}f(x)j_2 \tag{1.15}$$

1.6.4 Erreur quadratique moyenne intégrée (Mean-Integrated Squared Error: MISE)

Cependant, Il peut être intéressant d'avoir une mesure globale de la précision de f_n comme estimateur de f au lieu d'avoir une mesure de précision en un point donné. Cette mesure globale qu'on note MISE est définie par:

$$MISE(f_n(x)) = E\left[\int \{f_n(x) - f(x)\}^2 dx\right].$$

On peut montrer que:

$$\begin{aligned} MISE(f_n(x)) &= \int E[(f_n(x) - f(x))^2] dx \\ &= \int MSE(f_n(x)) dx \\ &= \int (E[f_n(x)] - f(x))^2 dx + \int Var[f_n(x)] dx \end{aligned}$$

$$MISE(f_n(x)) = \frac{1}{4}k_2^2\beta(f) + \frac{1}{nh}j_2 + o\left(h^4 + \frac{1}{nh}\right),$$

Où: $\beta(f) = \int f''(x)^2 dx$.

Asymptotiquement, on a:

$$AMISE(f_n(x)) = \frac{1}{4}h^4k_2^2\beta(f) + \frac{1}{nh}j_2 \tag{1.16}$$

Si l'on compare les deux composantes du *MISE*: on observe l'un des problèmes fondamentaux dans l'estimation de densité. C'est-à-dire que si l'on désire éliminer le biais, on choisit h petit alors que l'intégrale de la variance devient grande. Par contre si on choisit h grand de façon à réduire l'intégrale de la variance, on introduit systématiquement un biais dans l'estimation. Donc, le choix du paramètre de lissage h doit être un compromis entre le biais et la variance. Nous verrons un peu plus loin quelques méthodes permettant de choisir le paramètre de lissage de façon optimale. La plupart de ces méthodes ont pour critère la minimisation d'une certaine mesure de dispersion comme le *MSE*, ou le *MISE*. Nous verrons par la suite que le choix du noyau n'a pas vraiment d'impact sur l'estimateur de la densité.

Théorème 1. *Li and Racine (2007)*

Si on suppose que $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, alors

$$\hat{f}_n(x) \rightarrow_p f(x)$$

Démonstration

D'après l'inégalité de Chebychev on a:

$$\begin{aligned} P(|f_n(x) - f(x)| > \epsilon) &\leq \frac{E(f_n(x) - f(x))^2}{\epsilon^2} \\ &= \frac{1}{\epsilon^2}h^4k_2^2f''(x)^2 + \frac{1}{nh\epsilon^2}f(x)j_2 \\ &\rightarrow 0 \end{aligned}$$

où la deuxième ligne est donnée par la formule (1.15)

Lemme 1. Lyapounov *Supposons que pour tout n, w_{n1}, \dots, w_{nn} sont des v.a indépendantes, tel que $E[w_{ni}] = 0$ et $\sigma_{ni}^2 = E[w_{ni}^2] < \infty$.*

Soit $s_n^2 = \sum_{i=1}^n \sigma_{ni}^2$. Pour un certain $\delta > 0$, si la condition suivante est satisfaite:

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} E[|w_{ni}|^{2+\delta}] = 0 \quad (1.17)$$

alors :

$$\sum_{i=1}^n w_{ni}/s_n \rightarrow_d N(0,1) \quad (1.18)$$

La condition (1.17) est appelée la condition de **Lyapounov**.

Lemme 2. *La condition de Lyapounov est satisfaite quand w_{n1}, \dots, w_{nn} sont i.i.d, $\sigma_n^2 = E[w_{ni}^2] > 0$ et $\lim_{n \rightarrow \infty} E[|w_{ni}|^{2+\delta}]/n^{\delta/2} = 0$ pour un certain $\delta > 0$.*

Démonstration:

Si les v.a sont i.i.d, on a

$$\begin{aligned} s_n^2 &= n\sigma_n^2, \text{ et} \\ s_n^{2+\delta} &= (n^{1/2}\sigma_n)^{2+\delta} = n^{1+\delta/2}\sigma_n^{2+\delta} \end{aligned}$$

De plus:

$$\begin{aligned} \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} E[|w_{ni}|^{2+\delta}] &= \sigma_n^{-2-\delta} n^{-1-\delta/2} \sum_{i=1}^n E[|w_{ni}|^{2+\delta}] \\ &= \sigma_n^{-2-\delta} n^{-\delta/2} E[|w_{ni}|^{2+\delta}] \end{aligned}$$

La condition de Lyapounov est satisfaite si $n^{-\delta/2} E[|w_{ni}|^{2+\delta}] \rightarrow 0$. \square

Supposons que $\lim_{n \rightarrow \infty} \sigma_n^2$ existe dans le cas de données i.i.d, le résultat de Lyapounov(CLT) peut être indiqué comme suit:

Corollaire 1. *Supposons que pour tout n, w_{n1}, \dots, w_{nn} sont i.i.d, avec $E[w_{ni}] = 0$, $\lim_{n \rightarrow \infty} E[w_{ni}^2] > 0$ et est finie, et $\lim_{n \rightarrow \infty} E[|w_{ni}|^{2+\delta}]/n^{\delta/2} = 0$ pour un certains $\delta > 0$, alors*

$$\frac{1}{n^{1/2}} \sum_{i=1}^n w_{ni} \sim N(0, \lim_{n \rightarrow \infty} E[w_{ni}^2]) \quad (1.19)$$

On démontre maintenant la normalité asymptotique de l'estimateur de densité à noyau.

Théorème 2. Pagan and Ullah (1999)

supposons que $nh \rightarrow \infty$ et $1/(nh)^{1/2}h^2 \rightarrow 0$ et $f(x) > 0$, on a

$$(nh)^{1/2}(f_n(x) - f(x)) \rightarrow_d N(0, f(x) \int K^2(u)du)$$

Démonstration:

$$\begin{aligned}
(nh)^{1/2} (f_n(x) - f(x)) &= \frac{1}{n^{1/2}} \sum_{i=1}^n \left(\frac{1}{h^{1/2}} K \left(\frac{X_i - x}{h} \right) - \frac{1}{n^{1/2}} E \left[K \left(\frac{X_i - x}{h} \right) \right] \right) \\
&+ (nh)^{1/2} \left(\frac{1}{h} E \left[K \left(\frac{X_i - x}{h} \right) \right] - f(x) \right)
\end{aligned}$$

D'après ce qu'on a vu précédemment on :

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

On définit:

$$w_{ni} = \frac{1}{h^{1/2}} K \left(\frac{X_i - x}{h} \right) - \frac{1}{h^{1/2}} E \left[K \left(\frac{X_i - x}{h} \right) \right]$$

$$(nh)^{1/2} (f_n(x) - f(x)) = \frac{1}{n^{1/2}} \sum_{i=1}^n w_{ni} + O_p \left((nh)^{1/2} h^2 \right) \quad (1.20)$$

$$= \frac{1}{n^{1/2}} \sum_{i=1}^n w_{ni} + o_p(1) \quad (1.21)$$

car on par hypothèse $(nh)^{1/2} h^2 \rightarrow 0$.

On vérifie maintenant les conditions du corollaire 1

Par définition de w_{ni} , $E[w_{ni}] = 0$ et

$$E[w_{ni}^2] = \frac{1}{h} E \left[K^2 \left(\frac{X_i - x}{h} \right) \right] - \frac{1}{h} \left(E \left[K \left(\frac{X_i - x}{h} \right) \right] \right)^2 \quad (1.22)$$

or

$$E \left[K \left(\frac{X_i - x}{h} \right) \right] = h \int K(u) f(x + uh) du \quad (1.23)$$

$$= O(h), \quad (1.24)$$

et

$$\frac{1}{h} E \left[K^2 \left(\frac{X_i - x}{h} \right) \right] = h \int K^2(u) f(x + uh) du \quad (1.25)$$

$$\rightarrow f(x) \int K^2(u) du \quad (1.26)$$

les résultats (1.24) et (1.26) impliquent que

$$\lim_{n \rightarrow \infty} E[w_{ni}^2] = f(x) \int K^2(u) du.$$

Pour finir nous montrons que $E[|w_{ni}^{2+\delta}|]/n^\delta \rightarrow 0$. On utilise l'inégalité suivante (Davison, 1994, théorème 9.2 en page 140) afin de développer $E[|w_{ni}^{2+\delta}|]$. Pour $r > 0$,

$$E \left[\left| \sum_{i=1}^m X_i \right|^r \right] \leq c_r \sum_{i=1}^m E[|X_i|^r] \quad (1.27)$$

où $c_r = 1$ quand $r \leq 1$, et $c_r = m^{r-1}$ quand $r \geq 1$. D'après l'inégalité (1.27),

$$E[|w_{ni}^{2+\delta}|] \leq 2^{1+\delta} \left(\frac{1}{h^{1+\delta/2}} E \left[\left| K \left(\frac{X_i - x}{h} \right) \right|^{2+\delta} \right] + \frac{1}{h^{1+\delta/2}} \left| E \left[K \left(\frac{X_i - x}{h} \right) \right] \right|^{2+\delta} \right)$$

par (1.24)

$$\frac{1}{h^{1+\delta/2}} \left| E \left[K \left(\frac{X_i - x}{h} \right) \right] \right|^{2+\delta} = O(h^{1+\delta/2})$$

et

$$\begin{aligned} \frac{1}{h^{1+\delta/2}} E \left[\left| K \left(\frac{X_i - x}{h} \right) \right|^{2+\delta} \right] &= \frac{1}{h^{1+\delta/2}} \int \left| K \left(\frac{u - x}{h} \right) \right|^{2+\delta} f(u) du \\ &= \frac{1}{h^{\delta/2}} \int |K(u)|^{2+\delta} f(x + uh) du \\ &= O\left(\frac{1}{h^{\delta/2}}\right) \end{aligned}$$

Par conséquent,

$$E[|w_{ni}^{2+\delta}|]/n^{\delta/2} = O\left(\left(\frac{1}{nh^{1/2}}\right)^\delta\right)$$

D'après les hypothèses du théorème 2, $nh \rightarrow \infty$ et $1/(nh^{1/2}) \rightarrow 0$, on a le résultat suivant $(nh)^{1/2}(f_n(x) - f(x)) \rightarrow^d N(0, f(x) \int K^2(u) du)$ \square

1.6.5 Choix théorique optimal du paramètre de lissage

Pour le paramètre de lissage on fait la distinction entre

- h paramètre de lissage constant (ou global).
- $h(x)$ paramètre de lissage variable (local).

Ces choix différents du paramètre de lissage résultent des estimateurs à noyau suivants:

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

$$f_{n,L}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(x)} K \left(\frac{x - X_i}{h(x)} \right)$$

Le choix $h(x)$ implique qu'un noyau différent est utilisé en chaque point.

Nous allons décrire des choix théoriques optimaux des paramètres de lissage h et $h(x)$.

Un critère approprié pour sélectionner un paramètre de lissage constant h est la *MISE*. Le paramètre de lissage optimal est la valeur de h qui minimise la *MISE*:

$$\frac{dMISE(f_n)}{dh} = h^3 k_2^2 \beta(f) - \frac{1}{nh^2} j_2.$$

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5} \quad (1.28)$$

où $\gamma(K) = j_2 k_2^{-2}$

Un critère approprié pour sélectionner un paramètre de lissage variable (local) $h(x)$ est la mesure de performance locale MSE .

$$\frac{dMSE}{dh} = h^3 k_2^2 f''(x)^2 - \frac{1}{nh^2} f(x) j_2.$$

d'où

$$h_{opt}(x) = \left(\frac{1}{n} \frac{f(x)\gamma(K)}{\beta(f)} \right)^{1/5} \quad (1.29)$$

Remarque: Les choix h_{opt} et $h_{opt}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités inconnues f et f'' .

Exemple 2. On utilise les mêmes données de l'exemple 1 et on fait varier le paramètre de lissage pour estimer la densité par la méthode du noyau, et on obtient les résultats représentés dans les trois figures suivantes:

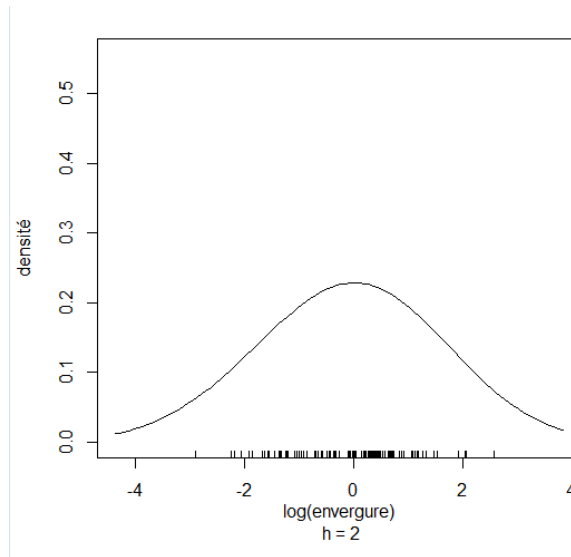


FIG. 1.5 – Illustration d'un phénomène de sous-lissage lors de l'estimation d'une densité .

Le choix d'une valeur de h trop grande conduit à une courbe trop lisse. La courbe estimée ne traduit pas suffisamment les variations de la vraie distribution. (fig 1.5)

Par contre, en choisissant un paramètre de lissage très petit, l'allure de la distribution change. Il s'agit d'une distribution surestimée (fig 1.6)

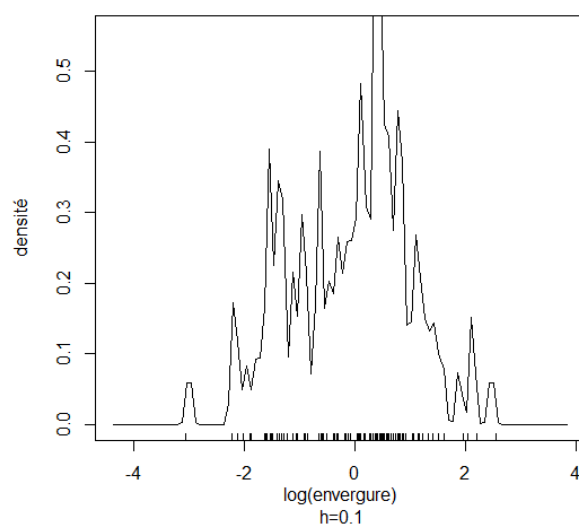


FIG. 1.6 – Illustration d'un phénomène de sur-lissage lors de l'estimation d'une densité .

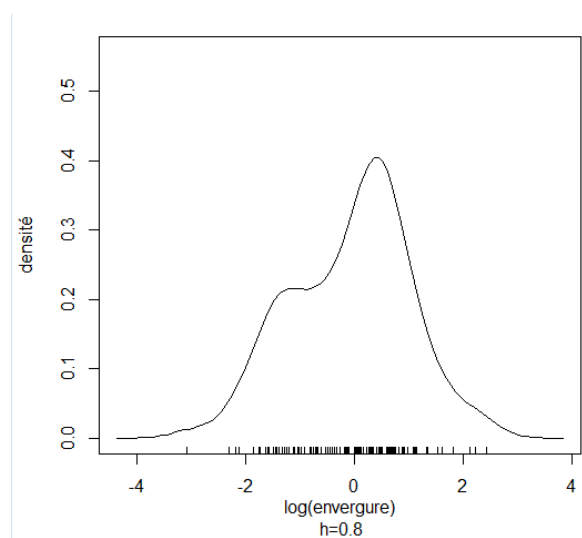


FIG. 1.7 – Illustration d'une estimation idéale.

L'estimation de la densité nécessite également le choix adéquat de la fenêtre h , et pour cette valeur idéale du paramètre de lissage, nous obtenons une allure qui suit parfaitement la vraie distribution. (fig 1.7)

Les courbes obtenues dans cette exemple illustrent à quel point les formes estimées sont différentes en fonction de l'ordre de grandeur du paramètre de lissage. La principale difficulté repose sur le choix optimal de la fenêtre h .

1.6.6 Le choix optimal pratique du paramètre de lissage h

Cas d'une distribution normale

Rappelons l'expression pour le paramètre de lissage optimal constant:

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5}$$

Supposons que f appartient à une famille de distributions normales $\mathcal{N}(\mu; \sigma^2)$, de moyenne μ et variance σ^2 inconnues, alors

$$f(x) = \frac{1}{\sigma} \varphi \left(\frac{x - \mu}{\sigma} \right)$$

avec $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, la densité de probabilité normale réduite.

$$f''(x) = \frac{1}{\sigma^3} \varphi'' \left(\frac{x - \mu}{\sigma} \right).$$

La quantité inconnue $\beta(f)$ s'écrit alors

$$\beta(f) = \int (f''(x))^2 dx = \frac{1}{\sigma^6} \int \left\{ \varphi'' \left(\frac{x - \mu}{\sigma} \right) \right\}^2 dx = \frac{1}{\sigma^5} \int \{\varphi''(v)\}^2 dv$$

$$\text{et } \varphi(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$$

$$\text{d'où } \varphi'(v) = -\frac{v}{\sqrt{2\pi}} e^{-v^2/2}$$

$$\text{et } \varphi''(v) = \frac{1}{\sqrt{2\pi}} (v^2 - 1) e^{-v^2/2}$$

$$\begin{aligned} \beta(f) &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ \int_{-\infty}^{+\infty} v^4 e^{-v^2} dv - 2 \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \end{aligned}$$

Posons $\mu = \sqrt{2}v \rightarrow d\mu = \sqrt{2}dv$

$$\begin{aligned} &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} \frac{u^2}{2} e^{-u^2/2} \frac{du}{\sqrt{2}} + \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-u^2/2} du \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{4} \sqrt{\pi} + \sqrt{\pi} \cdot 1 \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \frac{3}{4} \sqrt{\pi} = \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}} \end{aligned}$$

L'expression du paramètre de lissage optimal asymptotique devient

$$h_{AMISE} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \sigma n^{-1/5}$$

Le paramètre de lissage du type "normale référence" est défini par

$$\hat{h}_{NR} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \hat{\sigma} n^{-1/5}$$

Où $\hat{\sigma}$ est un estimateur de σ .

Quelques choix possibles pour $\hat{\sigma}$ sont donnés ci dessous:

- L'estimateur sans biais de la variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- L'écart interquartile empirique standardisé:

$$\frac{\text{Ecart interquartile empirique}}{\Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4})} \equiv \frac{R}{\Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4})} \simeq \frac{R}{1.349}$$

où $\Phi(\cdot)$ est la fonction de répartition de la loi normale réduite.

Remarquons que $\Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4})$ est l'écart interquartile d'une variable aléatoire normale réduite. La motivation pour la standardisation utilisant cette quantité est simple:

Si $X \sim N(\mu, \sigma^2)$, alors $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ et

$$P\{\Phi^{-1}(\frac{1}{4}) \leq Z \leq \Phi^{-1}(\frac{3}{4})\} = 0.50$$

$$\Leftrightarrow P\{\Phi^{-1}(\frac{1}{4}) \leq \frac{X-\mu}{\sigma} \leq \Phi^{-1}(\frac{3}{4})\} = 0.50$$

$$P\{\sigma\Phi^{-1}(\frac{1}{4}) + \mu \leq X \leq \Phi^{-1}(\frac{3}{4}) + \mu\} = 0.50$$

L'écart interquartile de X est

$$F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4}) = \sigma[\Phi^{-1}(\frac{3}{4}) - \Phi^{-1}(\frac{1}{4})]$$

ce qui justifie l'estimateur proposé.

On propose d'utiliser le minimum entre S et $R/1.349$, c'est-à-dire d'utiliser le paramètre de lissage suivant:

$$\hat{h}_{NR} = \left\{ \frac{8\sqrt{\pi}\gamma(K)}{3} \right\}^{1/5} \min(S, \frac{R}{1.349})n^{-1/5}.$$

Pour quelques noyaux, on donne les expressions de \hat{h}_{NR}

noyau K	paramètre de lissage pratique \hat{h}_{NR}
densité normale réduite $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$\hat{h}_{NR} = 1.06 \min(S, \frac{R}{1.349})n^{-1/5}$
noyau "Epanechnikov" $\frac{3}{4}(1-x^2)I\{ x \leq 1\}$	$\hat{h}_{NR} = 2.34 \min(S, \frac{R}{1.349})n^{-1/5}$
noyau "biweight" $\frac{15}{16}(1-x^2)^2I\{ x \leq 1\}$	$\hat{h}_{NR} = 2.78 \min(S, \frac{R}{1.349})n^{-1/5}$

FIG. 1.8 – L'expression de h_{NR} pour quelques noyaux.

Exemple 3. on a pris comme échantillon 100 réalisations d'une loi normale $N(0,1)$. Le premier graphe de la figure(1.8) en rouge représente la densité de l'estimateur à noyau de f en utilisant le h_{NR} , le graphe bleu est la densité d'estimateur à noyau de f avec un paramètre de lissage h choisi par défaut.

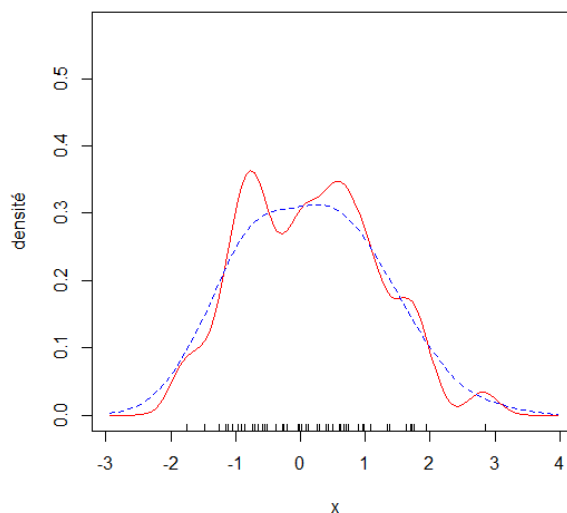


FIG. 1.9 – Représentation d'un estimateur à noyau en utilisant h_{NR} .

La méthode des moindres carrés avec validation croisée (MC-VC)

La méthode de validation croisée (en anglais: cross-validation) du type moindres carrés a été introduite par Rudemo (1982) et Bowman(1984). Cette méthode permet d'obtenir un paramètre de lissage simple et attrayant. La méthode est motivée par la décomposition suivante de l'erreur quadratique moyenne intégrée $MISE\{f_n(\cdot)\}$ de l'estimateur à noyau:

$$\begin{aligned}
MISE\{f_n\} &= E \left[\int f_n(x) - f(x)^2 \right] \\
&= E \left[\int \hat{f}^2(x) dx \right] + 2E \left[\int f_n(x) f(x) dx \right] + \int f^2(x) dx
\end{aligned}$$

Remarquons que le terme $\int f^2(x) dx$ ne dépend pas de h , et donc minimiser $MISE\{f_n(\cdot)\}$ par rapport à h est équivalent à minimiser

$$MISE\{\hat{f}\} - \int f^2(x) dx = E \left[\int f_n^2(x) dx + 2 \int f_n(x) f(x) \right] dx$$

L'expression à droite de cette équation est inconnue car elle dépend de la densité f . Un estimateur pour $\int f_n(x) f(x) dx$ est donné par

$$\frac{1}{n} \sum_{i=1}^n f_{n,-i}(X_i) \tag{1.30}$$

où

$$f_{n,-i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_h(x_i - X_j) \tag{1.31}$$

est l'estimateur à noyau basé sur l'échantillon "réduit" $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, où l'observation X_i à été supprimée. On appelle cet estimateur le "*leave - one - out estimator*".

Le terme "validation croisé" vient du fait qu'une partie de l'échantillon est utilisée pour obtenir l'information sur une autre partie: les observation $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ sont utilisées pour avoir une idée de $f(X_i)$.

L'estimateur (1.30) est un estimateur sans biais de $E\{\int f_n(x) f(x) dx\}$. En effet,

$$\begin{aligned}
E\left\{\frac{1}{n} \sum_{i=1}^n f_{n,-i}(X_i)\right\} &= \frac{1}{n} \sum_{i=1}^n E\{f_{n,-i}(X_i)\} \\
E\{f_{n,-i}(X_i)\} &= \frac{1}{n-1} \sum_{j=1, j \neq i} E\{K_n(X_i - X_j)\} \\
&= E\{K_h(X_1 - X_2)\} \\
&= \int \int K_h(x - y) f(x) f(y) dx dy \\
&= \int \left\{ \int K_h(x - y) f(y) dy \right\} f(x) dx \\
&= \int E\{f_n(x)\} f(x) dx \\
&= E\left\{ \int f_n(x) f(x) dx \right\}
\end{aligned}$$

Un estimateur sans biais pour la quantité

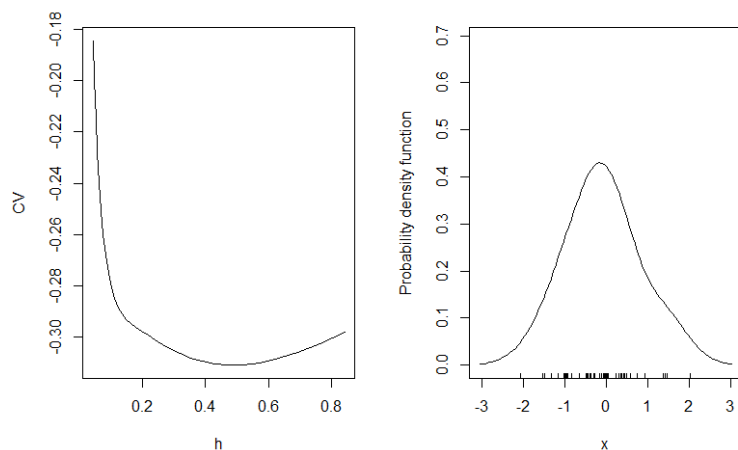


FIG. 1.10 – Choix du h par la méthode de validation croisé.

$$MISE\{f_n\} - \int f^2(x)dx = E \left[\int f_n^2(x)ds \right] + 2 \int f_n(x)f(x) dx$$

est donné par

$$LSCV(h) = \int f_n^2(x)dx - \frac{2}{n} \sum_{i=1}^n f_{n,-i}(X_i)$$

Cette quantité est appelée la quantité de validation croisée (LSCV),

LSCV veut dire en anglais "Least-square cross-validation"

Le paramètre de lissage du type "validation croisé" est la valeur de h qui minimise cette quantité, c'est-à-dire

$$\hat{h}_{LSCV} = \operatorname{argmin}_h LSCV(h)$$

La méthode du maximum de vraisemblance avec validation croisée (MV-VC)

Les paramètres d'un modèle statistique peuvent être estimés par la méthode du maximum de vraisemblance. Dans cette méthode, on retient les paramètres qui maximisent la fonction de vraisemblance ou, plus souvent le logarithme de la fonction de vraisemblance. Par exemple, pour l'estimation d'un paramètre inconnu d'une densité, l'estimateur du maximum de vraisemblance serait:

$$\hat{\theta} = \arg \max_{\theta} l(\theta|x) \quad (1.32)$$

où θ est un vecteur de paramètres, x est un vecteur d'observations, et

$$l(\theta|x) = \log \prod_{i=1}^n f(x_i|\theta) = \sum_{i=1}^n \log f(x_i|\theta) \quad (1.33)$$

Dans le contexte non paramétrique, la forme de f dépend des observations et l'équation (1.31) n'est pas directement applicable. HABBEMA et al.(1974) et DUIN(1976) ont donc suggéré de considérer la méthode de validation croisée pour estimer la fonction de vraisemblance. Supposons que l'on dispose d'une nouvelle donnée y indépendante des données déjà disponibles x . En considérant x fixe, la fonction de log-vraisemblance de h se note de la façon suivante:

$$l(h|y) = \log[\hat{f}(y|h,x)] \quad (1.34)$$

où \hat{f} est une densité de noyau basée sur les observations x . Comme on ne dispose généralement pas d'un échantillon supplémentaire, la technique de validation croisée est nécessaire. On remplace donc (1.32) par:

$$l(h|x_i) = \log[\hat{f}_{-i}(x_i|h)] \quad (1.35)$$

Ceci peut être fait pour chacune des observations. En prenant la moyenne et en utilisant (1.30), on peut obtenir après quelques manipulations mathématiques l'expression suivante pour la log-vraisemblance avec validation croisée:

$$MVVC_f = \frac{1}{n} \sum_{i=1}^n \log \left[\sum_{j \neq i} K \left(\frac{x_i - x_j}{h} \right) \right] + \log \left[\frac{1}{h(n-1)} \right] \quad (1.36)$$

L'estimation du paramètre de lissage optimal par la méthode du maximum de vraisemblance avec validation croisée est obtenue en maximisant cette expression.

La méthode d'ADAMOSWSKI

La méthode d'ADAMOWSKI (1985) s'appuie sur la fonction de distribution pour estimer le paramètre de lissage. Elle consiste à minimiser la différence entre l'estimation de la fonction de distribution par la méthode des noyaux et une estimation de la fonction de distribution empirique. On doit d'abord classer les observations par ordre croissant et calculer la valeur de la probabilité empirique p_j associée à chacune des observations. Le paramètre de lissage optimal est celui qui minimise l'erreur quadratique entre l'estimation non paramétrique et les p_j :

$$AC_F(h) = \sum_{j=1}^n [F_{n,K}(x_j) - p_j]^2 \quad (1.37)$$

Il existe plusieurs formules de probabilité empirique. Les formules de probabilités empiriques sont souvent utilisées pour visualiser la distribution des observations d'un échantillon. Elles permettent aussi de détecter la présence de valeurs singulières. Ces formules peuvent s'exprimer sous la forme générale suivante:

$$p_j = \frac{j - \alpha}{n + 1 + 2\alpha} \text{ avec } 0 \leq \alpha \leq 1 \quad (1.38)$$

où j est la j^{ime} valeur de l'échantillon rangé par ordre croissant et α est une constante qui dépend du type de distribution. ADAMOWSKI(1981) propose de considérer un coefficient $\alpha = 0.25$ dans l'expression (1.37) pour le calcul des probabilités empiriques. Ainsi, l'estimation du paramètre de lissage optimal par la méthode d'ADAMOWSKI est obtenue en minimisant l'expression (1.36)

La méthode Plug-in

Afin de faire disparaître quelques effets indésirables du biais, Sheather et Jones[1991] mettent en place une procédure de type plug-in en trois étapes. Ils choisissent d'estimer $R(f'') = \int_{-\infty}^{+\infty} (f''(x))^2 dx$ par:

$$S(a) = \frac{1}{n(n-1)} a^5 \sum_{i=1}^n \sum_{j=1}^n K^{(4)} \left(\frac{x_i - x_j}{a} \right), \quad (1.39)$$

où $K^{(4)}$ désigne la dérivée quatrième du noyau suffisamment lisse K et où a est un nouveau paramètre de lissage appelé paramètre pilote.

Pour un noyau K gaussien, le paramètre de lissage h_{SJ} est alors la solution du critère $SJ(h)$ défini par:

$$SJ(\hat{h}) = \left(\frac{1}{2\sqrt{\pi}} \right)^{\frac{1}{5}} S(\hat{\alpha}(\hat{h})) (f^{(2)})^{-1/5} n^{-1/5} - \hat{h} \quad (1.40)$$

avec

$$\hat{\alpha}(h) = 1.357 \left(\frac{S(a)}{T(b)} \right)^{1/7} h^{5/7}$$

$$T(b) = \frac{1}{n(n-1)b^7} \sum_{i=1}^n \sum_{j=1}^n K^{(6)} \left(\frac{x_i - x_j}{b} \right)$$

$$a = 0.920 \hat{\lambda} n^{-1/7}$$

$$b = 0.912 \hat{\lambda}^{-1/9}$$

où $\hat{\lambda}$ représente l'estimateur d'une mesure d'échelle (par exemple l'écart interquartile).

1.6.7 Simulation des données

Dans cette partie, nous illustrons certains estimateurs à noyaux continus symétriques à savoir le noyau d'Epanechnikov, le noyau Gaussien, le noyau Biweight et le noyau triangulaire. Nous simulons un échantillon de taille $n = 100$ de la loi normale centrée et réduite. Pour chaque noyau fixé, la fenêtre optimale est choisie par les méthodes de validation croisée par moindres carrés et de Plug-in.

Nous obtenons les résultats suivants:

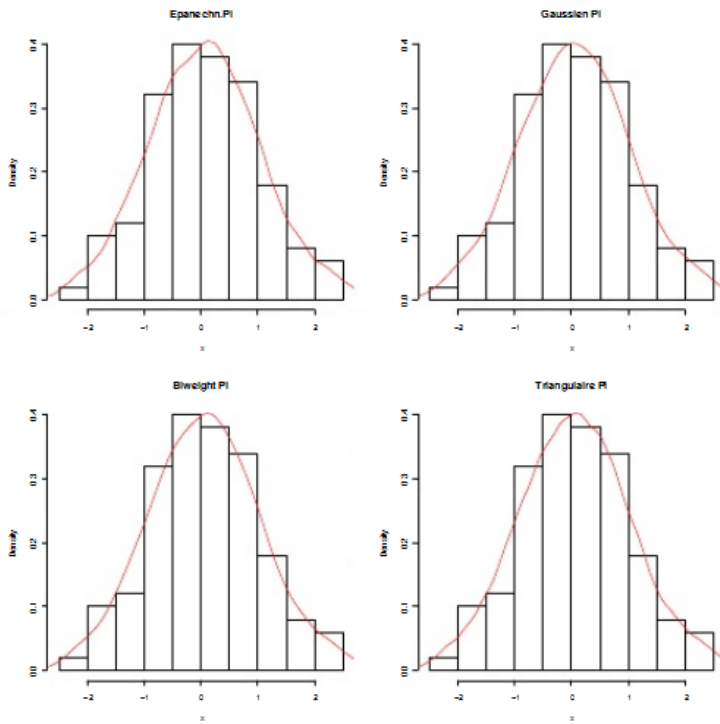


FIG. 1.11 – Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon de loi normale centrée réduite, $n = 100$ et $h_{PI} = 0.338$.

Nous obtenons dans la figure (1.11) pratiquement des estimations similaires pour chaque noyau continu utilisé; ceci s'explique par le fait que les noyaux continus symétriques possèdent tous des efficacités proches l'une de l'autre. Pour chaque noyau continu symétrique fixé, la figure (1.12) présente la fenêtre optimale $h_{cv} = 0.1636$. Pour cette valeur de h , les estimations des différentes densités sont pratiquement similaires.

Nous comparons différentes estimations en faisant varier la valeur de la fenêtre pour le même noyau continu. Nous choisissons le noyau optimal d'Epanechnikov.

Les simulations effectuées dans la figure (1.13) mettent en lumière le fait que les performances pratiques des estimateurs à noyaux continus symétriques considérés dépendent fortement du choix de la fenêtre h . Par conséquent, ce choix est plus crucial que le choix du noyau. Les valeurs de h sont celles choisies par Plug-In ($h_{PI} = 0.338$), validation croisée par vraisemblance ($h_{cv} = 0.429$) et deux autres valeurs arbitraires tel que $h = 0.5$ et $h = 1$.

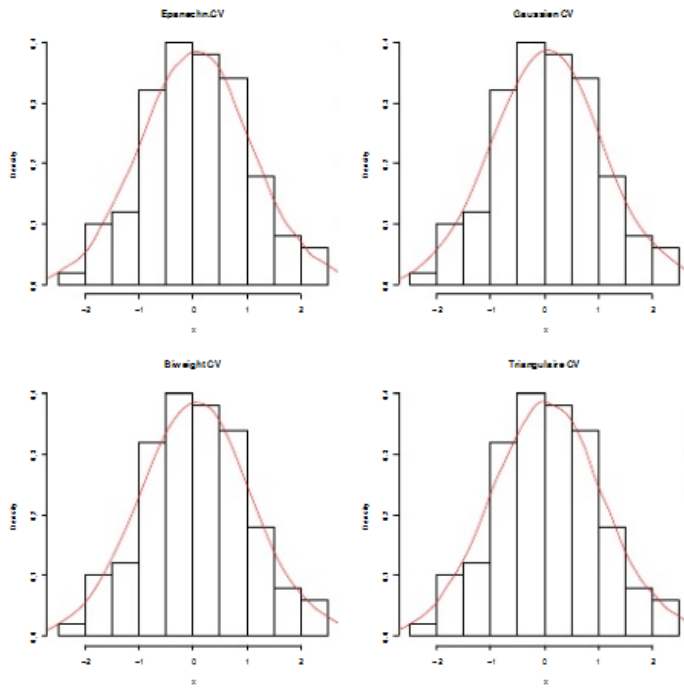


FIG. 1.12 – Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon de loi normale centrée réduite, $n = 100$ et $h_{CV} = 0.42$.

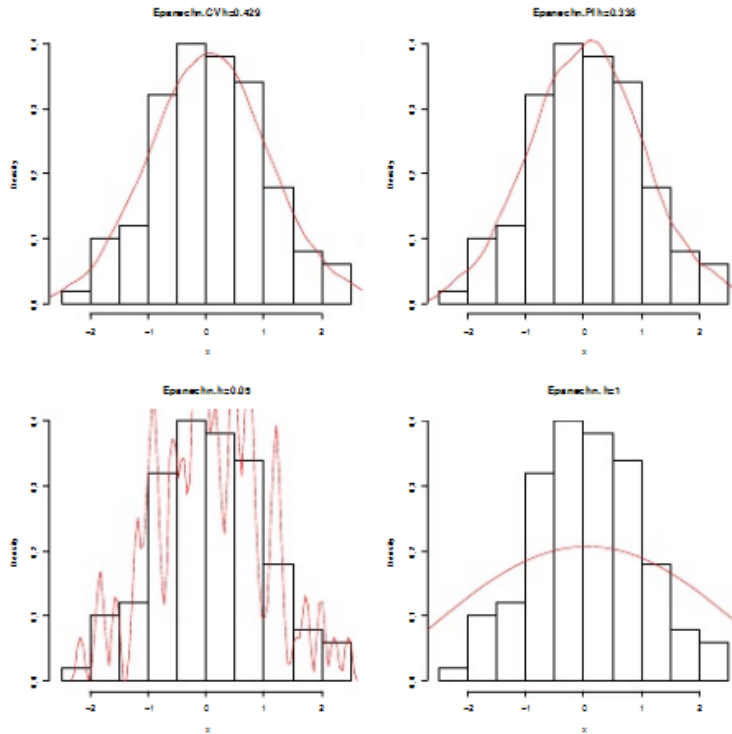


FIG. 1.13 – Comparaison des lissages par l'estimateur à noyau continu d'Epanchnikov en faisant varier la fenêtre h .

1.6.8 Le choix optimal du noyau

Supposons que l'on a choisi le paramètre de lissage h de telle sorte que le MISE soit minimum.

Comme on l'a montré avant, ce paramètre de lissage optimal h_{opt} est égal à

$$h_{opt} = \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)} \right)^{1/5}$$

On remarque que le h_{opt} dépend de la densité f qui est inconnue. Néanmoins, nous pouvons tirer des conclusions. On remarque que h_{opt} tend vers zéro mais de façon très lente quand n augmente.

En substituant h_{opt} dans la formule (1.16), on montre alors que si h est choisi de façon optimale la valeur appropriée pour le MISE sera:

$$MISE_{opt}(f_n) = \frac{5}{4} \left(\frac{\beta(f)j_2^4 k_2^2}{n^4} \right)^{1/5}$$

Pour minimiser le MISE, il faut choisir le noyau K qui minimise la valeur $C(K)$ tel que:

$$C(K) = (j_2^2 k_2^2)^{1/5}$$

Hodges et Lehmann (1956) ont montré que ce problème de minimisation est résolu en choisissant $K(z)$ égale à

$$K(z) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}z^2) & \text{si } |z| \leq \sqrt{5} \\ 0 & \text{sinon} \end{cases}$$

On le note K_{EP} parce que le noyau est appelé le noyau d'Epanechnikov. Nous pouvons donc considérer l'efficacité d'un noyau K (notée $eff(K)$) quelconque en le comparant à K_{EP} puisque ce dernier minimise le MISE si h est choisi de façon optimale. donc

$$eff(K) = \left(\frac{MISE_{opt}(\hat{f}) \text{ using } K_{EP}}{MISE_{opt}(\hat{f}) \text{ using } K} \right)^{5/4}$$

Le tableau suivant donne quelques noyaux et leurs efficacités respectives.

	Kernel K	eff(K)
Epanechnikov	$\frac{3}{4}(1 - u^2)I(u \leq 1)$	1.000
Quartic (biweight)	$\frac{15}{16}(1 - u^2)^2I(u \leq 1)$	0.994
Triweight	$\frac{35}{32}(1 - u^2)^3I(u \leq 1)$	0.987
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$	0.951
Uniform	$\frac{1}{2}I(u \leq 1)$	0.930
Triangle	$(1 - u)I(u \leq 1)$	0.986

FIG. 1.14 – Quelques noyaux et leurs efficacité.

Exemple 4. *Nous considérons les données `data(faithful)` qui se trouvent dans le package `datasets` et on représente dans la figure suivante les différentes densités des quatre noyaux différents en utilisant le paramètre de lissage optimale.*

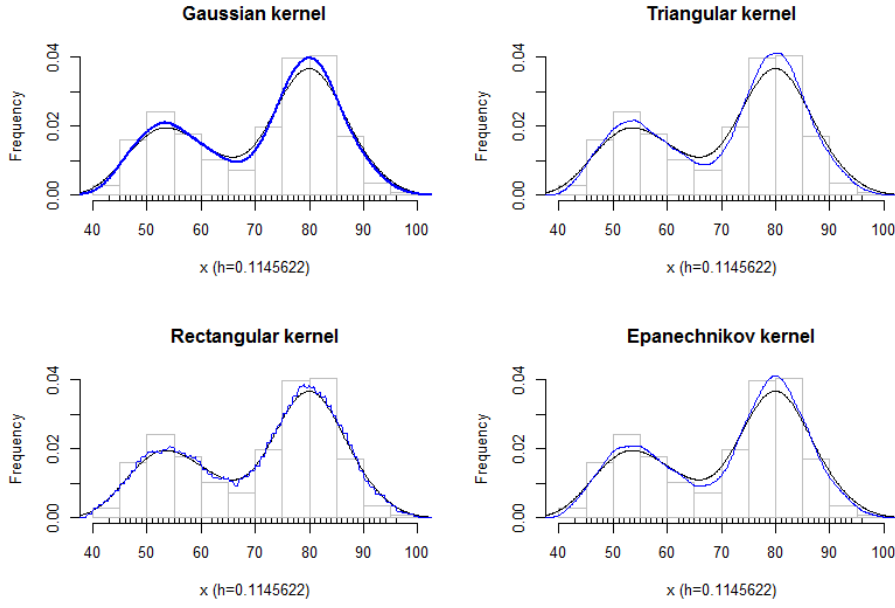


FIG. 1.15 – Estimation de la densité avec différents noyaux.

Remarque:

D'après la figure, on voit que les densités estimées par les quatre noyaux sont pratiquement les mêmes sauf le noyau rectangulaire qui donne un estimateur moins lisse que les autres, ce qui veut dire que le choix du noyau n'a pas impact sur l'estimateur à noyau.

1.7 Estimateur à noyau dans le cas multivarié

On suppose que X_i est un q -vecteur aléatoire et on veut estimer la densité $f(x) = f(x_1, \dots, x_q)$. L'estimateur à noyau multivarié est donné par:

$$f_n(x) = \frac{1}{n|H|} \sum_{i=1}^n K(H^{-1}(X_i - x)) \quad (1.41)$$

Où $K(u)$ est la fonction du noyau multivarié qui dépend du vecteur $H = (h_1, \dots, h_q)'$ et $|H| = h_1 h_2 \dots h_q$. avec

$$\int K(u)(du) = \int K(u) du_1 \dots du_q = 1$$

Typiquement, $K(u)$ est:

$$K(u) = k(u_1)k(u_2) \dots K(u_q).$$

Quand $K(u)$ est le produit des noyaux alors la densité marginale de $f_n(x)$ est égale a la densité de l'estimateur à noyau dans le cas univarié avec k la fonction du noyau et h_j le paramètre de lissage.

On peut donc montrer que quand $K(u)$ prend la forme d'un produit, le biais de l'estimateur est:

$$Biais(\hat{f}(x)) = \frac{\kappa_\nu(k)}{\nu!} \sum_{j=1}^q \frac{\delta^\nu}{\delta x_j^\nu} f(x) h_j^\nu + o(h_1^\nu + \dots + h_q^\nu)$$

et la variance est donnée par:

$$\text{Var}(\hat{f}(x)) = \frac{f(x)R(k)^q}{n|H|} + o\left(\frac{1}{n}\right)$$

Où $R(k) = \int k^q(u)du$

Par conséquent le AMISE est :

$$\text{AMISE}(\hat{f}(x)) = \frac{\kappa_\nu^2(k)}{(\nu!)^2} \int \left(\sum_{j=1}^q \frac{\delta^\nu}{\delta x_j^\nu} f(x) h_j^\nu \right)^2 (dx) + \frac{R(k)^q}{n|H|}$$

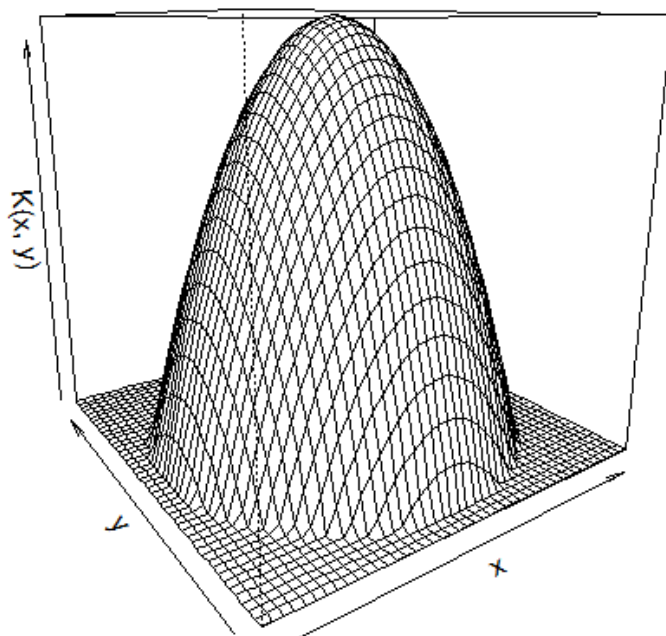


FIG. 1.16 – Densité d'un noyau d'Epanechnikov dans le cas multivarié.

Chapitre 2

La valeur de risque (La VaR)

2.1 Les mesures de risque

Cette section résume en quelques lignes les outils mathématiques usuels d'une mesure de risque. Une attention particulière est portée aux caractéristiques particulières que l'on peut attendre des mesures de risque pour pouvoir être utilisées dans la pratique.

Nous reprenons ici la définition d'une mesure de risque telle qu'elle est donnée dans Denuit et Charpentier (2004)

Définition 4. *Mesure de risque*

On appelle mesure de risque toute application ρ associant un risque X un réel $\rho(X) \in \overline{\mathbb{R}}$

Après avoir défini ce qu'est une mesure de risque, nous rappelons les principales propriétés qu'elles doivent vérifier pour être jugées satisfaisantes.

2.2 Caractéristiques d'une mesure de risque

La définition d'une mesure de risque est très générale puisque toute fonctionnelle réelle positive d'une variable aléatoire peut être considérée comme étant une mesure de risque. Aussi, en pratique, on exige pour de telles mesures qu'elles disposent de propriétés mathématiques dont la transcription conceptuelle permette de les juger. En pratique, on exige fréquemment qu'une mesure de risque ρ qui est définie comme suit:

$$\rho : \mathcal{V} \rightarrow \overline{\mathbb{R}}$$

avec \mathcal{V} l'ensemble des variables aléatoires à valeurs réelles.

cette mesure de risque possède une partie des caractéristique suivantes:

- Invariance en loi:

$$\forall X, Y \in \mathcal{V}, \quad X \stackrel{\ell}{=} Y : \rho(X) = \rho(Y)$$

- Invariance par translation:

$$\forall X \in \mathcal{V}, \forall \alpha \in \mathbb{R}, \quad \rho(X + \alpha) = \rho(X) - \alpha$$

- Sous-additivité:

$$\forall X, Y \in \mathcal{V}, \quad \rho(X + Y) \leq \rho(X) + \rho(Y)$$

- Homogénéité :

$$\forall \lambda \geq 0, \forall X \in \mathcal{V}, \quad \rho(\lambda X) = \lambda \rho(X)$$

- Monotonie :

$$\forall X, Y \in \mathcal{V} \quad X \leq Y : \rho(Y) \leq \rho(X)$$

- Pertinence

$$\forall X \in \mathcal{V}, \quad x \leq 0 \text{ et } X \neq 0 : \rho(X) > 0$$

2.2.1 Mesure de risque cohérente

L'association des quatre dernières propriétés de la mesure de risque à donné naissance au concept de cohérence d'une mesure de risque, voir Artzner et al.(1999).

Définition 5. *Mesure cohérente*

Une mesure de risque invariante par translation, sous-additive, homogène et monotone est dite cohérente.

2.2.2 Mesure de risque monétaire

Définition 6. *Une mesure invariante par translation et monotone est dite monétaire*

2.3 Mesures de risque usuelles

L'objectif de ce paragraphe est de présenter les mesures de risque les plus usuelles, et on s'attardera particulièrement sur la Value-at-Risk .

2.3.1 La variance

Soit $X \in \mathcal{V}$,

$$Var(X) = \sigma_X^2 = E[X - E[X]]^2 \tag{2.1}$$

La variance n'est pas une mesure du risque cohérente. Car elle ne vérifie pas la propriété d'invariance par translation ni celle de la monotonie,

$$Var(X + a) = Var(X) \neq Var(X) - a$$

2.4 Value-at-Risk

Historique:

La notion de Value-at-Risk (VaR) est apparue pour la première fois dans le secteur de l'assurance. A la fin des années 1980, la banque Bankers Trust fut l'une des premières institutions à utiliser cette notion sur les marchés financiers aux Etats-Unis, mais c'est principalement la banque JP Morgan qui dans les années 90 a popularisée ce concept notamment grâce à son système Risk Metrics (pour un historique complet de la notion de Value-at-Risk et de sa diffusion se

reporter au livre de Dowd, 2005). La Value-at-Risk est ensuite devenue, en moins d'une dizaine d'années, une mesure de référence du risque sur les marchés financiers, consacrée notamment par la réglementation prudentielle définie dans le cadre des accords de Bâle II.

2.4.1 Présentation théorique de la méthode VaR

Définition: qu'est ce que la VaR?

Pour introduire l'idée de base, prenons un exemple. Vous avez investi une partie de vos économies dans un portefeuille d'actions. Votre conseiller vient de vous informer que la valeur de votre portefeuille a encore baissé le mois dernier et que celui-ci vaut maintenant 5000 DT. Après avoir écouté ses explications sur les raisons de cette mauvaise performance, vous désirez sûrement avoir une idée de la perte maximale que le portefeuille pourrait enregistrer à la fin du mois.

La réponse la plus correcte serait que vous pourriez perdre toutes vos économies. Or cette réponse n'est pas satisfaisante parce qu'elle ne vous apporte rien de nouveau et, surtout, parce que le scénario de perte totale a trop peu de chances de se produire. Il serait plus réaliste et plus professionnel de dire: "En l'absence d'événements exceptionnels, il y a 95% de chances que le portefeuille reparte à la hausse ou qu'il perde 1000 DT ou moins d'ici la fin du mois." C'est le genre de réponse que la méthode VaR permet de donner.

De façon générale, la Value-at-Risk est définie comme la perte maximale potentielle qui ne devrait être atteinte qu'avec une probabilité donnée sur un horizon temporel donné (Engle et Manganelli, 2001). La Value at Risk est donc la pire perte attendue sur un horizon de temps donné pour un niveau de confiance donné. Cette définition très simple constitue l'un des principaux attraits de la Value-at-Risk: il est en effet très facile de communiquer sur la VaR et de proposer ainsi une mesure homogène et générale (quelque soit la nature de l'actif, la composition du portefeuille etc.) de l'exposition au risque.

2.4.2 Présentation mathématique de la méthode VaR

Avant de donner une présentation mathématique de la méthode VaR on va d'abord donner quelques rappels sur les fonctions de répartition et les quantiles

1.4.2.1 Rappels sur les fonctions de répartition et les quantiles

Définition 7. Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction $F : \mathbb{R} \rightarrow [0,1]$ définie par

$$\forall x \in \mathbb{R}, F(x) = \mathbf{P}(X \leq x).$$

La proposition suivante regroupe des propriétés classiques de la fonction de répartition..

Proposition 1. La fonction de répartition F d'une variable aléatoire réelle X est croissante, continue à droite et vérifie

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F(x) = 1.$$

Ensuite, la fonction de répartition caractérise la loi: si X et Y sont deux variables aléatoires réelles qui ont même fonction de répartition, alors $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Enfin, une suite $(X_n, n \in \mathbb{N}^*)$ de variables aléatoires réelles converge en loi vers X si et seulement si, pour tout point de continuité $x \in \mathbb{R}$ de la fonction de répartition F de X , on a

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$$

où F_n désigne la fonction de répartition de X_n .

Soit $x \in \mathbb{R}$ et $(X_n, n \in \mathbb{N}^*)$ une suite de variable aléatoires réelles indépendantes et de même loi.

D'après la loi forte des grands nombres, on a p.s:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{I}_{\{X_k \leq x\}} = \mathbb{P}(X_1 \leq x)$$

Le théorème de Glivenko-Cantelli assure en fait que cette convergence est uniforme en x .

Définition 8. • Soit X une variable aléatoire réelle de fonction de répartition F . Pour $p \in]0,1[$, on appelle quantile ou fractile d'ordre p de X le nombre

$$\begin{aligned} x_p &= \inf\{x \in \mathbb{R}, F(x) \geq p\} \\ &= \sup\{x \in \mathbb{R}, F(x) \leq p\} \end{aligned}$$

• L'application continue à gauche $p \in]0,1[\rightarrow x_p \in \mathbb{R}$ s'appelle l'inverse généralisé de F . On la note F^{-1} :

$$\forall p \in]0,1[, F^{-1}(p) = x_p.$$

1.4.2.2 La VaR

La VaR acronyme désignant la value-at Risk, est l'un des derniers nés de mesures de risque et l'instrument le plus utilisé de nos jours en pratique pour contrôler le risque; c'est une mesure très simple à appréhender puisqu'elle définit le risque par une valeur numérique unique.

Définition 9. La Value-at-Risk n'est rien d'autre qu'un fractile de la distribution de perte et profit associée à la détention d'un actif ou d'un portefeuille d'actifs sur une période donnée. La mesure de Value-at-Risk ne fait que refléter l'information contenue dans la queue gauche (associée aux pertes) de la distribution des rendements d'un actif. Si l'on considère un taux de couverture de α (ou de façon équivalente un niveau de confiance de $(1 - \alpha)$) la Value-at-Risk correspond tout simplement au fractile de niveau α de la distribution de perte et profit valable sur la période de détention de l'actif:

$$\mathbf{P}[X \leq VaR_\alpha] = \alpha \Leftrightarrow F(VaR_\alpha) = \alpha$$

d'où

$$VaR_\alpha(\alpha) = F^{-1}(\alpha)$$

où $F(\cdot)$ désigne la fonction de répartition associée à la distribution de perte et profit.

1.4.2.3 Les paramètres de la VaR

La VaR d'un portefeuille est caractérisée par les deux paramètres suivants:

• **Le niveau de confiance choisi** (ou seuil de probabilité) s'élevant à 95% ou 99% par exemple et qui est la probabilité que les pertes éventuelles du portefeuille ou de l'actif ne dépassent pas la Value-at-Risk.

• **L'horizon temporel choisi:** Ce paramètre est très important car plus l'horizon est long, plus les pertes peuvent être importantes. Par exemple, pour une distribution normale des rendements, il faut multiplier la Value-at-Risk à un jour par \sqrt{t} pour avoir la Value-at-Risk sur t jours.

1.4.2.4 Interprétation mathématique

Selon Esch, Kieffer et Lopez (1997) et Jorion (2000), la VaR de l'actif en considération, pour une durée t et un niveau de probabilité α , se définit comme le montant de la perte étendue de sorte que ce montant, pendant la période $[0,t]$, ne devrait pas être plus important que la VaR avec une probabilité de $(1 - \alpha)$. Autrement dit:

$$\mathbf{P}[X_t > VaR_\alpha] = 1 - \alpha \Leftrightarrow \mathbf{P}[X_t < VaR_\alpha] = \alpha$$

Où X_t est la perte sur le titre à l'instant t .

En centrant et réduisant l'expression, nous obtenons:

$$\mathbf{P} \left[\frac{X_t - E(X_t)}{\sigma(X_t)} \leq \frac{VaR_\alpha - E(X_t)}{\sigma(x_t)} \right] = \alpha$$

Nous pouvons donc définir:

$$\frac{VaR_\alpha - E(X_t)}{\sigma(X_t)} = Z_\alpha$$

D'où nous pouvons calculer la VaR comme étant:

$$VaR_\alpha = E(X_t) + Z_\alpha \sigma(X_t)$$

2.4.3 Propriétés de la VaR

Commençons par noter que la VaR est stable par transformation croissante: quelque soit le niveau de probabilité $\alpha \in [0,1]$ et la fonction g croissante et continue, on a

$$VaR_\alpha(g(X)) = g(VaR_\alpha(X))$$

En fait plus généralement, on a

Lemme 3. *Pour tout $\alpha \in [0,1]$, si g est une fonction strictement croissante continue à gauche,*

$$VaR_\alpha(g(X)) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(\alpha)) = g(VaR_\alpha(X))$$

et si g est une fonction strictement décroissante, continue à droite et F_X bijective, on a

$$VaR_\alpha(g(X)) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(1 - \alpha)) = g(VaR_{(1-\alpha)}(X))$$

Démonstration: Voir Charpentier Arthur (2010)

- Artzner et al (1997) ont montré que la VaR n'est pas une mesure de risque cohérente car elle ne vérifie pas la condition de sous-additivité.

Ce résultat peut se démontrer à l'aide d'un contre-exemple.

Rappel:

On dit que X suit une loi de Pareto de paramètres α et k , on écrit $X \rightarrow Par(\alpha, k)$ si sa fonction de répartition

$$P[X \leq x] = 1 - \left(\frac{x}{k}\right)^{-\alpha} \quad x \geq k, \alpha, k > 0$$

et sa fonction de densité

$$f_X(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}} \quad x \geq k, \alpha, k > 0.$$

Considérons deux variables aléatoires X, Y indépendantes suivant une même loi de Pareto.
 $X \rightarrow Par(1,1)$ et $Y \rightarrow Par(1,1)$,

$$P(X \leq t) = P(Y \leq t) = 1 - \frac{1}{t}, t > 0$$

Nous avons alors

$$\begin{aligned} \alpha &= P[X \leq VaR_\alpha(X)] \\ &= 1 - \frac{1}{VaR_\alpha(X)} \\ &= \frac{VaR_\alpha(X) - 1}{VaR_\alpha(X)} \end{aligned}$$

$$\begin{aligned} : \alpha(VaR_\alpha(X)) &= VaR_\alpha(X) - 1 \\ : VaR_\alpha(X)(1 - \alpha) &= 1 \\ : VaR_\alpha(X) &= \frac{1}{1 - \alpha} \end{aligned}$$

Par la suite

$$VaR_\alpha(X) = VaR_\alpha(Y) = \frac{1}{1 - \alpha}$$

De plus on peut vérifier

$$P[X + Y \leq t] = 1 - \frac{2}{2+t} + 2 \frac{\ln(1+t)}{(2+t)^2}, t > 0$$

Vu que

$$P[X + Y \leq 2VaR_\alpha(X)] = \alpha - \frac{(1 - \alpha)^2}{2} \ln\left(\frac{1 + \alpha}{1 - \alpha}\right) < \alpha$$

L'inégalité suivante

$$VaR_\alpha(X) + VaR_\alpha(Y) \leq VaR_\alpha(X + Y)$$

est vraie quel que soit α , la VaR ne peut pas être donc sous-additive dans ce cas.
 La figure suivante illustre ce résultat:

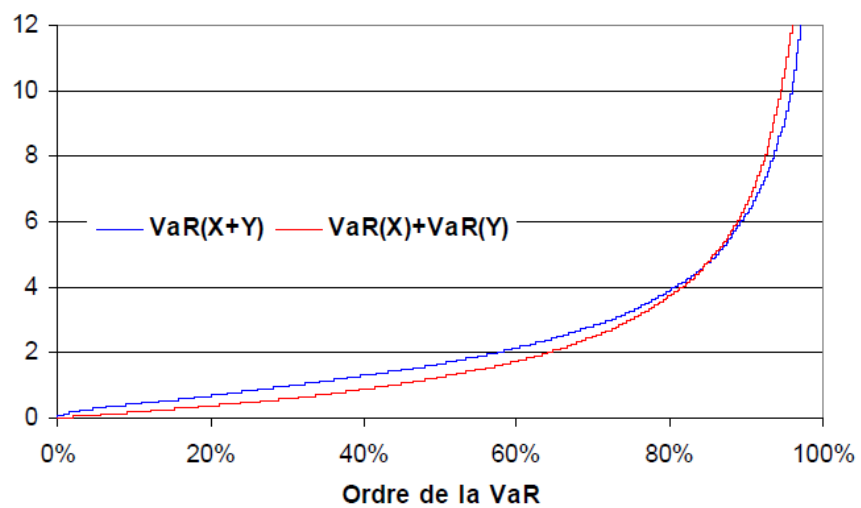


FIG. 2.1 – Value-at-Risk de la somme de deux v.a. de Pareto.

Chapitre 3

Estimation d'un quantile par la méthode des noyaux

3.1 Les estimateurs d'un quantile

Soit X_1, X_2, \dots, X_n un échantillon d'une v.a X de fonction de répartition F et de densité f . $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$ la statistique d'ordre correspondante. La fonction quantile Q est définie comme suit:

$$Q(p) = \inf\{x : F(x) \geq p\} \text{ avec } 0 < p < 1$$

On note pour tout $0 < p < 1$, le quantile d'ordre p de F par ξ_p , $\xi_p = Q(p)$.

Comme on l'a vu auparavant l'estimateur non paramétrique d'une fonction de distribution est la fonction empirique $F_n(x)$, qui est définie par:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_A(X_i)$$

où $I_A = 1$ si $x \in A$ et 0 ailleurs, où $A =]-\infty, x]$. En conséquence, l'estimateur non paramétrique de ξ_p est le quantile empirique

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

où $[np]$ désigne la partie entière de np .

Soit $p_r = r/(n+1)$ et $q_r = 1 - p_r$. Si on utilise la statistique d'ordre $X_{(r)}$ pour estimer le quantile d'ordre p_r , le biais asymptotique de $X_{(r)}$ est

$$\text{Biais}\{X_{(r)}\} = \frac{p_r q_r Q''(p_p)}{2(n+2)} + \frac{p_r q_r}{(n+2)^2} \left\{ \frac{1}{3}(q_r - p_r) Q_r''' + \frac{1}{8} Q_r'''' \right\} \quad (3.1)$$

et la variance asymptotique est

$$\text{Var}\{X_{(r)}\} = \frac{p_r q_r}{(n+2)} Q_r'^2 + \frac{p_r q_r}{(n+2)^2} \left\{ 2(q_r - p_r) Q_r' Q_r'' + p_r q_r (Q_r' Q_r''' + \frac{1}{2} Q_r'') \right\} \quad (3.2)$$

3.2 Inverse d'estimateur à noyau d'une fonction de répartition

L'estimateur à noyau d'un quantile le plus populaire est donné par Nadaraya (1964). Il est défini comme suit:

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

où

$$K_h = \int_{-\infty}^x \frac{1}{h} k\left(\frac{t}{h}\right) dt,$$

k est la fonction noyau, l'estimateur du quantile $Q = F^{-1}$ est défini par

$$Q_n(p) = \inf\{x : F_n(x) \geq p\}, \quad 0 < p < 1. \quad (3.3)$$

Supposons que:

- (1) f est différentiable de dérivé f' ;
- (2) f' est continue au voisinage de ξ_p et $f'(\xi_p) \neq 0$;
- (3) $\int_{-\infty}^{+\infty} xk(x)dx = 0$ et $\int_{-\infty}^{+\infty} x^2k(x)dx < \infty$.

Théorème 3. (Ming-Yen Cheng et Shan Sun [106]) *Sous les hypothèses (1) et (3), l'erreur quadratique moyenne asymptotique de $\hat{Q}(p)$ est*

$$AMSE\{Q_n(p)\} = \frac{p(1-p)}{nf(\xi_p)^2} + \frac{h^4}{4} \frac{f'(\xi_p)^2}{f(\xi_p)^2} \mu_2(k)^2 - \frac{h}{n} \frac{1}{f(\xi_p)} \psi(k)$$

et asymptotiquement le paramètre de lissage optimal de $Q_n(p)$ est donné par

$$h_{opt,1} = \left[\frac{f(\xi_p)\psi(k)}{n\{f'(\xi_p)\}^2\mu_2(k)^2} \right]^{\frac{1}{3}} \quad (3.4)$$

Où $\mu_2(k) = \int_{-\infty}^{+\infty} t^2k(t)dt$ et $\psi(k) = 2 \int yk(y)K(y)dy..$

3.3 L'estimateur à noyau d'un quantile

Un type d'estimateur à noyau d'un quantile est donné par Yang(1985) et également par Parzen(1979). Il est défini par:

$$Q_n(p) = \sum_{i=1}^n X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} K\left(\frac{p-x}{h}\right) dx \quad (3.5)$$

il est clair que quand $\frac{i}{n}$ est près de p , $\tilde{Q}_n(p)$ met plus de poids sur la statistique d'ordre $X_{(i)}$. Dans la pratique, l'approximation suivante de $Q_n(p)$ est souvent utilisée

$$Q_{n,1}(p) = \sum_{i=1}^n [n^{-1} K\left(\frac{i/n - p}{h}\right)] X_{(i)} \quad (3.6)$$

Cet estimateur est une adaptation de la régression lissée et étudié par Priestley et Chao (1972). Yang(1985) a montré que $Q_n(p)$ et $Q_{n,1}(p)$ sont asymptotiquement équivalents en moyenne quadratique. Si toutes les observations X_i sont multipliées par (-1), alors en général $Q_{n,1}(p)(-X_{(1)}, -X_{(2)}, \dots, -X_{(n)}) \neq -Q_{n,1}(1-p)(X_{(1)}, X_{(2)}, \dots, X_{(n)})$, ceci est dû au fait que le

poids de $X_{(n-i+1)}$ de $Q_n(p)$ diffère du poids $X_{(i)}$ de $Q_{n,1}(1-p)$. Ce problème peut être surmonté en remplaçant i/n dans la définition de $Q_{n,1}(p)$ par $(i-1/2)$ ou $i/(n+1)$. On obtient les estimateurs suivants:

$$Q_{n,2}(p) = \sum_{i=1}^n \left[n^{-1} K \left(\frac{\frac{i-1/2}{n} - p}{h} \right) \right] X_{(i)} \quad (3.7)$$

$$Q_{n,3}(p) = \sum_{i=1}^n \left[n^{-1} K \left(\frac{\frac{i}{n+1} - p}{h} \right) \right] X_{(i)} \quad (3.8)$$

La somme des poids de ces trois derniers estimateurs n'est pas en général égale à 1 du fait que si une constante c est ajoutée aux X_i alors

$$Q_{n,i}(p)(X_{(1)} + c, X_{(2)} + c, \dots, X_n + c) \neq Q_{n,i}(p)(X_{(1)}, X_{(2)}, \dots, X_n) + c \text{ pour } i = 1, 2, 3 \quad (3.9)$$

Ceci peut être réglé en divisant les poids par leur somme.

$Q_{n,2}(p)$ devient

$$Q_{n,4}(p) = \sum_{i=1}^n \left[n^{-1} K \left(\frac{\frac{i-1/2}{n} - p}{h} \right) \right] X_{(i)} / \sum_{i=1}^n \left[n^{-1} K \left(\frac{\frac{i-1/2}{n} - p}{h} \right) \right] \quad (3.10)$$

Cette estimateur est une adaptation de la régression lissée proposé par Nadaraya (1964) et Watson (1964).

3.4 Propriétés asymptotiques de $Q_n(p)$ et les estimateurs relatifs

Nous commençons cette section en notant que les résultats asymptotiques donnés dans cette section au sujet des estimateurs du quantile à noyau décrivent seulement la situation quand $p \in [0,1]$. Le théorème 1 nous donne l'expression de l'erreur quadratique moyenne asymptotique de $Q_n(p)$.

Théorème 4. (Sheather and Marron, 1990) *En supposant que Q'' est continue au voisinage de p et K est la densité d'un noyau, symétrique autour de 0. Pour tout $p \in (0,1)$, à l'exception de $p = 0.5$ quand F est symétrique*

$$MSE(Q_n(p)) = n^{-1} p(1-p) [Q'(p)]^2 - 2n^{-1} h [Q'(p)]^2 \int_{-\infty}^{+\infty} u K(u) K^{(-1)}(u) du \quad (3.11)$$

$$+ 1/4 h^4 [Q''(p)]^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + o(n^{-1}h) + o(h^4) \quad (3.12)$$

Quand F est symétrique, pour $p = 0.5$, on a:

$$MSE(Q_n(0.5)) = n^{-1} [Q'(1/2)]^2 - \{h \int_{-\infty}^{+\infty} u K(u) K^{(1)}(u) du \quad (3.13)$$

$$+ n^{-1} h^{-1} \int_{-\infty}^{+\infty} K^2(u) \} + o(n^{-1}h) + (n^{-2}h^{-2}) \quad (3.14)$$

Notons que pour un choix raisonnable de h le terme dominant dans le MSE est la variance asymptotique du quantile.

Preuve:

Premièrement on considère $p \in (0,1)$ avec $p \neq 0.5$ et F symétrique. Q'' est continue au voisinage de p . En utilisant la formule (1.1) donnée par David (1981), on a:

$$\begin{aligned} \text{Biais}(Q_n(p)) &= \sum_{i=1}^n \left[\int_{i-1/n}^{i/n} K\left(\frac{t-p}{h}\right) dt \right] \left\{ Q\left(\frac{i}{n+1}\right) - Q(p) \right\} + O(n^{-1}) \\ &= \int_0^1 K\left(\frac{t-p}{h}\right) \{Q(t) - Q(p)\} dt + O(n^{-1}) \\ &= 1/2h^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right] Q''(p) + o(h^2) + o(n^{-1}) \end{aligned}$$

Falk (1984,p263) a montré que:

$$\text{Var}(Q_n(p)) = n^{-1}p(1-p)[Q'(p)]^2 - n^{-1}h[Q'(p)]^2 \int_{-\infty}^{+\infty} uK(u)K^{(-1)}(u)du + o(n^{-1}h) \quad (3.15)$$

or

$$\text{MSE}(Q_n(p)) = \text{Var}(Q_n(p)) + \text{Biais}^2(Q_n(p))$$

d'où

$$\begin{aligned} \text{MSE}(Q_n(p)) &= n^{-1}p(1-p)[Q'(p)]^2 - 2n^{-1}h[Q'(p)]^2 \int_{-\infty}^{+\infty} uK(u)K^{(1)}(u)du \\ &= 1/4h^4[Q''(p)]^2 \left[\int_{-\infty}^{+\infty} u^2 K(u)du \right]^2 + o(n^{-1}h) + o(h^4) \end{aligned}$$

Corollaire 2. *Supposons que les conditions du théorème 1 sont satisfaites. Pour tout p , avec $p \neq 0.5$ et F symétrique, le paramètre de lissage optimal asymptotique est donné par*

$$h_{opt} = \alpha(K) \cdot \beta(Q) \cdot n^{-1/2} \quad (3.16)$$

où

$$\alpha(K) = \left[\frac{2 \int_{-\infty}^{+\infty} uK(u)K^{(-1)}(u)du}{\left\{ \int_{-\infty}^{+\infty} u^2 K(u)du \right\}} \right]^{1/3} \quad (3.17)$$

et

$$\beta(Q) = \left[\frac{Q'(p)}{Q''(p)} \right]^{2/3} \quad (3.18)$$

Avec $h = h_{opt}$

$$\text{MSE}(Q_n(p)) = n^{-1}p(1-p)[Q'(p)]^2 + O(n^{-4/3}) \quad (3.19)$$

Nous présentons un théorème qui établit quelques équivalences asymptotiques entre les différentes formes d'un estimateur d'un quantile par la méthode des noyaux.

On dit que deux estimateurs à noyau d'un quantile $Q_{n,i}(p)$ et $Q_{n,j}(p)$ sont asymptotiquement équivalents si pour une valeur raisonnable de h ,

$$E[(Q_{n,i}(p) - Q_{n,j}(p))^2] = o(n^{-4/3})$$

Théorème 5. *Simon J. Sheather, J.S. Marron [1990] Soit K la densité du noyau*

(i) *Pour $hn^{2/3} \rightarrow \infty$, $Q_n(p)$ et $Q_{n,2}(p)$ sont équivalents,*

(ii) Pour $hn^{2/3} \rightarrow \infty, Q_{n,2}(p)$ et $Q_{n,1}(p)$ sont équivalents,

(iii) Pour $hn^{5/6} \rightarrow \infty, Q_{n,1}(p)$ et $Q_{n,3}(p)$ sont équivalents,

(iv) Pour $hn^{5/6} \rightarrow \infty, Q_n(p)$ et $Q_{n,4}(p)$ sont équivalents,

Preuve

Voir Simon J. Sheather, J.S. Marron [1990]

3.5 Sélection du paramètre de lissage optimal

Dans cette section nous proposons un choix optimal de h pour $Q_n(p)$, pour tout $p \neq 0.5$ quand F est symétrique.

Nous voyons du (corollaire 1) que pour un choix de K la valeur asymptotiquement optimale de h dépend des première et deuxième dérivées du quantile. Les estimateurs de $Q'(p)$ et $Q''(p)$ sont nécessaires pour le choix du h .

$$\hat{Q}'_m(p) = \sum_{i=1}^n \left[\int_{i-1/n}^{i/n} a^{-2} K'_*(a^{-1}(t-p)) dt \right] X_{(i)}$$

et

$$\hat{Q}''_m(p) = \sum_{i=1}^n \left[\int_{i-1/n}^{i/n} b^{-3} K'_*(b^{-1}(t-p)) dt \right] X_{(i)}$$

où K_* est le noyau d'ordre m , symétrique autour de zéro.

L'estimation du quantile nous donne asymptotiquement le paramètre de lissage optimal suivant:

$$\hat{h}_{opt} = \alpha(K) \cdot \hat{\beta} \cdot n^{-1/3} \tag{3.20}$$

où $\hat{\beta} = \left[\frac{\hat{Q}'_m(p)}{\hat{Q}''_m(p)} \right]^{2/3}$ et $\alpha(K)$ est donné par (3.17)

3.6 La normalité asymptotique de l'estimateur à noyau dun quantile

Théorème 6. (Michael Falk 1985) Si F^{-1} a une dérivée seconde bornée autour de $q \in (0,1)$ et $(F^{-1})' > 0$, k a un support borné, $\int k(x)dx = 1$ et $\alpha_n \rightarrow_{n \rightarrow \infty} 0$ on a:

$$\frac{n^{1/2}(Q_n(F_n) - Q_n(F))}{\sigma_n} \rightarrow \mathcal{N}(0,1)$$

avec

$$\sigma_n^2 := \int_0^1 \left\{ \int_{-c}^c k(x)[q - \alpha_n x - 1_{(0,q-\alpha_n x)}(y)](F^{-1})'(q - \alpha_n x) dx \right\}^2 dy$$

et

$$\sigma_n^2 \rightarrow_{n \in \mathbb{N}} ((F^{-1})'(q))^2 q(1-q)$$

Idée de la démonstration

La représentation de Bahadur qu'on verra par la suite nous sera utile pour la démonstration de ce théorème car elle nous permettra d'écrire l'estimateur à noyau d'un quantile sous forme d'une somme de variables aléatoires *i.i.d.*

3.7 Représentation de Bahadur

Soit X_1, X_2, \dots, X_n des variables aléatoire *i.i.d* de fonction de distribution commune $F(X)$ et ξ_p le quantile de F . On note par $F_n^{-1}(p)$ l'estimateur de Parzen de ξ_p . Dans ce travail on établit la représentation de Bahadur pour $F_n^{-1}(p)$.

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

Bahadur (1966) a montré que si $F(x)$ est deux fois différentiable en ξ_p et $F'(\xi_p) = f(\xi_p) > 0$,

$$\hat{\xi}_{pn} - \xi_p = \frac{1}{n} \sum_{i=1}^n Y_i + R_n \quad (3.21)$$

où

$$Y_i = \frac{p - I_{(X_i \leq \xi_p)}}{f(\xi_p)} \quad i = 1, 2, \dots \quad (3.22)$$

et avec la probabilité 1

$$R_n = O(n^{-3/4}(\log n)^{1/2}(\log_2 n)^{1/4}), \quad n \rightarrow \infty \quad (3.23)$$

Parzen(1980) a introduit l'estimateur à noyau du quantile suivant:

$$\hat{F}_n^{-1}(p) = \frac{1}{h} \int_0^1 F_n^{-1}(x) K\left(\frac{x-p}{h}\right) \quad (3.24)$$

La représentation de Bahadur nous permet d'écrire:

$$\hat{F}_n^{-1}(p) - \xi_p = \frac{1}{n} \sum_{i=1}^n Z_i^{(n)} + R_n(p, h) \quad (3.25)$$

où pour tout n , $Z_1^{(n)}, Z_2^{(n)}, \dots, Z_n^{(n)}$ sont des variables aléatoires *i.i.d*. Avec un choix convenable du paramètre de lissage h . Le taux de convergence de $R_n(p, h)$ est $O((\log_2 n/n)^{2m/(2m+1)})$ si $F^{(m)}(x)$ est continue autour de ξ_p

On suppose que:

$$K(x) \in \mathcal{C}^3(-\infty, +\infty) \text{ avec le support compact dans } [-1, 1] \quad (3.26)$$

où $\mathcal{C}^3(-\infty, +\infty) = \{f : f^{(3)} \text{ est continue dans } (-\infty, +\infty)\}$

et pour un certain entier $m \geq 2$

$$\int_{-1}^1 K(x) dx = 1; \int_{-1}^1 x^j K(x) dx = 0, j = 1, \dots, m-1; \int_{-1}^1 x^m K(x) dx = \alpha_m \neq 0. \quad (3.27)$$

Rappel sur les conditions de Hölder:

En analyse, la continuité höldérienne ou la condition de Hölder est une condition suffisante (mais non nécessaire) pour qu'une application définie entre deux espaces métriques soit continue. La définition s'applique en particulier pour les fonctions d'une variable réelle.

Si (X, d_X) , et (Y, d_Y) , sont deux espaces métriques, une fonction $f : X \rightarrow Y$ est dite höldérienne s'il existe une constante $C > 0$, telle que:

$$\forall (x, y) \in X^2, d_Y(f(x), f(y)) \leq C d_X(x, y)^a$$

La continuité höldérienne d'une fonction dépend donc d'un paramètre réel strictement positif $a \in]0,1]$, et prend en compte toutes les variations de la valeur de la fonction sur son ensemble de définition.

- Lorsque $a = 1$, l'application est lipschitzienne.
- Si $0 < a \leq 1$ est fixé, l'ensemble des fonctions réelles höldériennes est un espace vectoriel.

Théorème 7. (Xiaojing Xiang 1994) on suppose que (3.26) est satisfaite, que $K^{(3)}$ satisfait la condition de Hölder avec un certain $\alpha > 0$, et que (3.27) est satisfaite avec $m \geq 2$. $F^{(m)}(x)$ existe et continue au voisinage de ξ_p , $F'(\xi_p) = f(\xi_p) > 0$, et $Q^{(m)}(p) \neq 0$.

Si $n \rightarrow \infty$, $h_n \searrow 0$ de telle manière que

(i) $(\log_2 n/n)^{-2/5-\beta} h_n \rightarrow \infty$ pour certain $\beta > 0$;

(ii) $\log h_n^{-1}/\log_2 n \rightarrow \infty$;

(iii) $2n \log h_n^{-1} = o(\log_2 n)$;

(iv) $\lim_{\epsilon \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \sup_n \left| \frac{h_{n'}}{h_n} - 1 \right| = 0$

avec $|n' - n| \leq \epsilon n$,

pour n grand,

$$\hat{F}_n^{-1}(p) - \xi_p = \frac{1}{n} \sum_{i=1}^n Z_i^{(n)} + R_n(p, h) \text{ a.s} \quad (3.28)$$

où

$$Z_i^{(n)} = \int_{-1}^1 Q'(p + h_n t) K(t) [p + h_n t - I_{(X_i \leq Q(p+h_n t))}] dt \quad (3.29)$$

étant des variables aléatoire i.i.d et

$$R_n(p, h_n) = C_1 h_n^m + C_2 \frac{\log_2 n}{n} h_n^{-1/2} I_n + o\left(h_n^m + \left(\frac{\log_2 n}{n}\right) h_n^{-1/2}\right) \text{ a.s} \quad (3.30)$$

où

$\overline{\lim}_{n \rightarrow \infty} \pm I_n = 1$ avec la probabilité 1 et

$$C_1 = \frac{1}{m!} \alpha_m Q^{(m)}(p), \quad C_2 = 2Q'(p)[p(1-p)]^{1/2} \left(\int_{-1}^1 K^2(x) dx \right)^{1/2}$$

En appliquant le théorème central limite a $\frac{1}{n} \sum_{i=1}^n Z_i$ on obtient la normalité asymptotique de l'estimateur du quantile.

On peut faire exactement la même chose pour l'estimation de la VaR, du moment que la VaR n'est qu'un quantile.

Conclusion

A travers ce travail, nous avons essayé de mettre l'accent sur la notion d'estimation non paramétrique et plus particulièrement sur l'estimation par la méthode des noyaux, car elle est relativement simple d'utilisation puisqu'il ne s'agit que de sélectionner la fonction noyau et de calculer le paramètre de lissage qui détermine le degré d'influence des observations pour l'estimation.

On a vu que le choix du noyau n'a pas une grande influence pour cette estimation, par contre le choix du paramètre de lissage a un impact important sur le biais.

Il serait intéressant de reprendre ce travail avec des données censurées ou avec des données manquantes et de les appliquer en séries chronologiques surtout pour les séries financières. De même, l'extension à des variables dépendantes, offrirait de nouvelles perspectives. De plus l'étude de la robustesse par rapport à la dépendance de l'estimateur à noyau d'un quantile pourrait être envisagée.

Bibliographie

- [1]Adamowski K. *Nonparametric kernel estimation of flood frequencies*.1985
- [2]Adamowski K. *Ploting formula for flood frequency*. 1981.
- [3]A. A. Borovkov., *Mathematical statistics. Gordon and Breach Science Publishers, Amsterdam*.1998.
- [4]Artzner, A., F. Delbaen, J-M. Eber et D. Heath, *Thinking coherently, Risk magazine, 10, November, 68-71*.1997.
- [5]Arthur Charpentier, *Mesures de risque*. 2010.
- [6]Bahadur, R. R, *A note on quantiles in large samples, Ann. Math. Stat., 37, 577-580*. (1966).
- [7]Berkoun Youcef. *statistiques inferentielle. Cours de 1ère année Master, U.M.M.T.O Tizi ouzou*, 2010.
- [8][Bowman. A. W, *An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71. 353-360*. 1984.
- [9]Branko Miladinovic. *Kernel density estimation of reliability with applications to extreme value distribution*. 2008
- [10]B. M. Falk. *Relative deficiency of kernel type estimators of quantile, The Annals of Statistics, vol.12* 1984.
- [11]Bowman, A. W, *An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71(2):353-360*.1984.
- [12]Cacoullos.T, *Estimation of multivariate density. Anals of the institue of statistical Mthematics 18, 178.189*. 1966.
- [13]D.Faucher, P.E.Rasmussen, B.Bobée, *Estimation non paramétrique des quantiles de crue par la méthode des noyaux*.2002
- [14]David, H.A, *Order Statistics, 2nd Edition, New York: John Wiley*. 1981.
- [15]Daniel Herlemont, *Value at Risk - étude de cas*. 17 juin 2011.
- [16]Davidson.J, *Stochastic Limit Theory. Oxford University Press, New York*. 1994.
- [17]Dowd Kevin, *Measuring Market Risk*. 2005.
- [18]Denuit, M. & Charpentier, A, *Mathématiques de l'assurance non-vie:Principes fondamentaux de théorie du risque*. 2004.
- [19]Duin R.P.W. *On the chose of smoothing parameters for Przen estimators of probability density fonction*. 1976.
- [20]E. Parzen *Nonparametric statistical data modeling, American Statistical Association, vol.74, pp.105-121*. 1979.
- [21]Esch, Kieffer, Lopez, *Value at Risk, vers un Risk Management moderne* 1997.
- [22]Hansen, B.E, *Uniform Convergence Rates for Kernel Estimation with Dependent Data*. 2008.
- [23] H. Follmer and A. Schied. *Stochastic finance. An introduction in discrete time, volume 27 de de Gruyter Studies in Mathematics*.2004.

- [24] Habbema J.D.F, *A stepwise discrimination program using density estimation*.1974.
- [25] Hodges, J. et Lehmann. E, *The efficiency of some nonparametric competitors of the t-test.*, *Annals Mathematical statistzcs* 27(1): 324-335. (1956).
- [26] Jorion, "Risk: Measuring the Risk in Value-At-Risk", *Financial Analysts Journal* 1996.
- [27] Kiefer. J, *On Bahadurs representation of sample quantiles*, *Ann. Math. Stat.*, 38, 1323-1342. 1967.
- [28] Li, Q., and J. S. Racine, *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton, New Jersey. 2007.
- [29] Le site "www.wikipedia.com".
- [30] Le site "http://maikolsolis.wordpress.com".
- [31] Le site web, *http : //193.49.79.89/esa_prof/index.php*
- [32] Manganelli , Simone and Engle, Robert F, *Value at Risk Models in Finance*. 2001.
- [33] Michel Lejeune, *Statistique la théorie et ses applications*. 2010.
- [34] Newey, W. K, *Kernel Estimation of Partial Means and a General Variance Estimator*. 1994.
- [35] N. Forcadel. *Mathématique du risque*. 2010.
- [36] Nadaraya, E.A, *On Estimating Regression*", *Theory of Probability and Its Applications*, 9, 141-142. 1964.
- [37] Pagan, A., and A. Ullah, *Nonparametric Econometrics, Themes in Modern Econometrics*. Cambridge University Press, New York. 1999.
- [38] Priestley, M.B. and Chao, M.T, "Nonparametric Function Fitting," *Journal of the Royal Statistical Society, Series B*, 34, 385-392. 1972.
- [39] P. Artzner, F.Delbaen, J-M. Eber et D. Heath, 203-228, *Coherent measures of risk*, *Mathematical Finance*, vol 9. Juillet 1999.
- [40] Rudemo, M. *Empirical Choice of Histograms and Kernel Density Estimators*. *Scandinavian Journal of Statistics*, 9:65-78. 1982.
- [41] Rosenblatt, *Remarks on some non-parametric estimate of density function* 1956
- [42] S.Benseghir. *Calcul de la VaR selon l'approche historique et la théorie des valeurs extrêmes sur un font alternatif de Dexia Asset Management*, 2006.
- [43] S. Sheather and J. Marron, *Kernel quantile estimators*, *Journal of the American Statistical Association*, vol.85, pp.410-416. 1990.
- [44] S.J Sheather and M.C. Jones, *A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation*, *J. Roy. Statist. Soc.*, B53, 683-690. 1991.
- [45] Sonia Jean, *Reconnaissance d'objets astronomiques HI par déconvolution et par estimation de densité*. 1997.
- [46] Silverman, B, *Density Estimation*. 1986.
- [47] Tarn Duong, *Kernel density estimation for bivariate data*. 2012
- [48] Walter Zucchini. *Applied smoothing techniques. Part 1: Kernel density estimation*. Octobre 2003.
- [49] Watson, O.S, "Smooth Regression Analysis", *Sankhya Series A*, 26, 359-372. 1964.
- [50] Xiaojing Xiang, *Bahadur Representation of Kernel Quantile Estimators*, *Source: Scandinavian Journal of Statistics*, Vol. 21, No. 2, pp. 169-178. Jun 1994.
- [51] Xiang. X, *Asymptotic theory for linear functions of ordered observations*. Ph.D. dissertation, University of Chicago. 1982.
- [52] Yang, S. S, *A smooth nonparametric estimator of a quantile function*. *J. Amer. Statist. Assoc.* 80, 1004-1011. 1985.

- [53] Yakowitz.S.J et Feluch, *Some "model-free" techniques for flood frequency analysis* 1983.
- [54] Zhi Ouyang. *Univariate Kernel Density Estimation*. August 2005
- [55] Zelterman. D, *Smooth Nonparametric Estimation of the Quantile Function* .1988.