

STATISTIQUE ET ANALYSE DES DONNÉES

WALTER SCHLEE

Estimation non paramétrique du α -quantile conditionnel et ses dérivées partielles

Statistique et analyse des données, tome 7, n° 1 (1982), p. 32-47

http://www.numdam.org/item?id=SAD_1982__7_1_32_0

© Association pour la statistique et ses utilisations, 1982, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

ESTIMATION NON PARAMÉTRIQUE
DU α -QUANTILE CONDITIONNEL ET SES DÉRIVÉES PARTIELLES

Walter SCHLEE

Institut für Statistik und Unternehmensforschung
Arcisstrasse 21
Technische Universität - D-8000 München 2

Résumé : En Statistique Appliquée on a souvent besoin du α -quantile conditionnel q_{α} au lieu de l'espérance mathématique conditionnelle. Des estimateurs non paramétriques pour q_{α} et ses dérivées partielles sont considérées ici. Les estimateurs proposés estiment les quantités théoriques directement: ils ne sont pas obtenus comme solutions des relations empiriques. Les propriétés statistiques, algorithmiques et numériques des estimateurs sont énoncées et démontrées. On donne un exemple en contrôle de qualité de matière minérale granulée où de tels estimateurs peuvent être utilisés avec avantage.

Abstract : In the field of practical applications it is often more advantageous to use the conditional α -quantile q_{α} instead of the conditional expectation. There are given nonparametric estimators for q_{α} and its partial derivatives. These are kernel-like estimators. They are not defined as solutions of empirical relations. The statistical, algorithmic and numerical properties of the estimators are given. Furthermore an example concerning the quality control of the particle size distribution of a mineral aggregate is described.

Mots clés : quantilogramme lissé, estimateur à noyau discret

1 - INTRODUCTION

Dans cet article des estimateurs non paramétriques du α -quantile conditionnel et de ses dérivées partielles sont proposés et leurs propriétés statistiques, algorithmiques et numériques sont énoncées.

Le α -quantile conditionnel $q_\alpha(z)$ est défini par

$$(1.1) \quad q_\alpha(z) = \sup \{y | P(Y \leq y | X=z) \leq \alpha\}$$

relativement au vecteur aléatoire (X, Y) où X est un vecteur de dimensions p et Y est de dimension un.

La fonction $q_\alpha : \mathbb{R}^p \rightarrow \mathbb{R}$ est considérée pour $\alpha \in]0, 1[$ fixé. Ce sujet ne doit pas être confondu avec le problème de Csörgö et Révész (1978) qui considèrent la fonction $\alpha \rightarrow q(\alpha) = \sup \{y | P(Y \leq y) \leq \alpha\}$.

Si la loi de probabilité conditionnelle de Y est symétrique, la fonction $q_{0.5}$ est identique à la fonction de régression $E(Y|X=z)$. Toutefois $q_{0.5}$ existe aussi dans le cas où $E(Y|X=z)$ n'existe pas. Déjà Brown et Mood (1951) utilisent la correspondance entre $E(Y|X=z)$ et $q_{0.5}(z)$ pour construire des tests non paramétriques pour l'hypothèse linéaire de la régression. Angers (1979), Griffiths et Willcox (1978), Hogg (1975), Turner et Bowden (1977) décrivent des problèmes pratiques où on a besoin du α -quantile conditionnel, aussi pour $\alpha \neq 0.5$, au lieu de l'espérance mathématique conditionnelle. Je suis moi-même confronté à un problème pratique où il faut estimer $q_\alpha(z)$.

Ce problème est un problème de contrôle de qualité de matière minérale granulée. La description complète se trouve dans Schlee (1980b). Ici je le décris en bref. Le revêtement des routes se compose d'un mélange de cailloux et de goudron. La solidité de cette surface dépend essentiellement de la distribution de la taille des cailloux. Il faut donc surveiller cette distribution. Ce contrôle est effectué grâce à l'expérimentation suivante:

m cribles sont utilisés. Le crible numéro m a les plus grands trous. Les cribles en dessous ont des trous de grosseur diminué. Une masse de cailloux est jeté au-dessus du crible supérieur, le crible numéro m . Cette masse est alors criblée. La masse qui est passée à travers ce crible est criblée avec le prochain crible. On continue jusqu'au dernier crible, le crible numéro 1. Soit M_i la masse restante actuellement sur le crible numéro i . M_0 dénote la masse qui est passée à travers le crible numéro 1. On peut normaliser la somme des M_i : $\sum_{i=0}^m M_i = 1$. L'influence de la manière de

crible et la déviation des trous de crible de leurs valeurs théoriques fait que les M_i sont des variables aléatoires qui oscillent près de leurs valeurs moyennes théoriques R_i . Le problème est compliqué par la dépendance stochastique des variables aléatoires M_i et en outre par la condition que la somme de M_i est 1.

Soit A_i la masse qui ne passe pas faussement le crible numéro i . Cette masse A_i manque le crible ci-dessous. De là A_i dépend du cours de cribler et des masses M_m, M_{m-1}, \dots, M_1 ci-dessus de manière inconnue. Alors un estimateur non paramétrique pour le α -quantile conditionnel de A_i par rapport à M_j peut être utilisé avec avantage, ce qui est fait avec des données provenant du Prüfamf für Bituminöse Baustoffe und Kunststoffe der Technischen Universität München. Le résultat se trouve dans Schlee (1980b). L'estimateur doit être non paramétrique parce qu'un modèle raisonnable n'est pas disponible.

Hogg (1975), Griffiths et Willcox (1978), Turner et Bowden (1977) regardent seulement des estimateurs paramétriques. C'est qu'on estime les paramètres dans un modèle préconçu. Deux estimateurs non paramétriques se trouvent dans Collomb (1978). Un de ces estimateurs est l'estimateur qui est appelé le quantilogramme ci-dessous. L'autre estimateur est comme suit: Un estimateur à noyau pour la répartition conditionnelle de Y est défini et l'estimateur pour q_α est la solution de (1.1) si on remplace la probabilité dans (1.1) par la valeur propre de la répartition conditionnelle estimée. Collomb (1978) donne la convergence presque sûre en tous points de ces deux estimateurs. Mais le calcul d'une solution d'une telle équation peut être très compliqué.

En conséquence nous cherchons ici à estimer directement la quantité q_α . En outre, d'une manière naturelle il est aussi possible d'estimer les dérivées partielles de q_α (si elles existent). Les dérivées sont importantes pour juger de la forme de la fonction q_α , par exemple la constance, la linéarité ou la convexité ou des autres propriétés définissable par des dérivées. Les estimateurs proposés sont semblables aux estimateurs à noyau. Ces estimateurs sont appelés ici estimateurs à noyau discret parce qu'ils comprennent une partition en intervalles du domaine de X . Cette définition est analogue à la définition de l'estimateur de la densité proposé par Schlee (1978).

Nous démontrons la loi asymptotique Gaussienne des estimateurs proprement normalisés avec des variances asymptotiques qui ne dépendent pas des quantités inconnues (comme par exemple, de la densité de probabilité). Nos estimateurs sont aussi constants.

Collomb (1981) décrit aussi des autres méthodes d'estimation non paramétrique (pour la régression), par exemple: la méthode de fonctions orthogonales et de fonctions splines. La recherche suivante montre que chaque estimateur direct pour la quantité q_α utilise essentiellement le groupement des données. Tel groupement n'insère pas dans ces méthodes (aussi on essaie d'adapter elles au problème de q_α). À présent il n'est pas connue comment on peut éviter le groupement des données dans

la définition d'un estimateur direct de q_α . De là les estimateurs à noyau discret ont une importance supplémentaire aux propriétés de paragraphe cinq ci-dessous.

2 - DESCRIPTION DES ESTIMATEURS

Les estimateurs sont construits comme les estimateurs dans Schlee (1978, 1979, 1980a) pour la régression et la densité de la probabilité. Il est défini une fonction, que j'appelle le quantilogramme, et ce quantilogramme est alors lissé d'une manière convenable.

Dans ce qui suit nous considérons des vecteurs aléatoires (X, Y) où X est un vecteur à p dimensions et Y est de dimension un. Soit un échantillon (X_i, Y_i) $i = 1, 2, \dots, n$ de taille n . Soit une partition du \mathbb{R}^p en intervalles

$$I(i) = \prod_{l=1}^p](i_l - 1)\beta, i_l\beta]$$

dont la longueur du côté est β , $i = (i_1, i_2, \dots, i_p)$ dénote un vecteur d'indices. Par ailleurs, $i(X_1) = (i_1(X_1), \dots, i_p(X_1))$ est le vecteur d'indices de l'intervalle $I(i)$ où se trouve l'épreuve X_1 . $w_n(i)$ est la fréquence relative de l'intervalle $I(i)$ concernant les X_1, X_2, \dots, X_n . $J_i(z)$ est la fonction indicatrice de $I(i)$. $Pe(z)$, $z \in \mathbb{R}$, dénote le plus grand nombre entier u qui satisfait $u \leq z$.

Soit $l = Pe(n w_n(i)\alpha)$. Alors $\bar{q}_{\alpha, i}$ est dans l'ordre numérique le l -ième des Y_j tels que X_j appartienne à $I(i)$, $j = 1, 2, \dots, n$.

La quantité suivante est le quantilogramme

$$\hat{q}_{n, \alpha}(z) = \sum_i \bar{q}_{\alpha, i} J_i(z).$$

Une représentation équivalente est

$$\hat{q}_{n, \alpha}(z) = \frac{\sum_i \bar{q}_{\alpha, i} J_i(z) \beta^{-q} w_n(i)}{\sum_i w_n(i) J_i(z) \beta^{-q}}$$

au moins dans le cas où le dénominateur n'est pas égal à zéro.

Si q_α est une fonction continue ou même dérivable, il peut être utile de disposer d'un estimateur possédant cette même propriété. Un tel estimateur est obtenu en remplaçant

$$\beta^{-p} J_i(z) \quad \text{par} \quad h^{-p} \hat{K}(z, i) \quad , \quad \text{où} \quad \hat{K}(z, i) = \prod_{j=1}^p K\left(\frac{z_j - i_j \beta}{h}\right) \quad ,$$

et la fonction $K: \mathbb{R} \rightarrow \mathbb{R}$ est au moins continue.

Alors l'estimateur régulier pour $q_\alpha(z)$ est donné par

$$\hat{q}_{n,\alpha}(z) = \begin{cases} \bar{q}_{n,\alpha}(z) / \hat{f}_n(z) & \text{si } \hat{f}_n(z) \neq 0 \\ 0 & \text{si } \hat{f}_n(z) = 0 \end{cases}$$

où

$$\hat{f}_n(z) = h^{-p} \sum_i \hat{K}(z, i) w_n(i)$$

$$\bar{q}_{n,\alpha}(z) = h^{-p} \sum_i \bar{q}_{\alpha,i} \hat{K}(z, i) w_n(i)$$

On voit que cet estimateur est analogue à l'estimateur de la fonction de régression défini par Schlee(1979), si on remplace $\bar{q}_{\alpha,i}$ par la moyenne arithmétique des Y_j tels que X_j appartienne à l'intervalle $I(i)$, $j=1,2,\dots,n$.

Pour définir les estimateurs des dérivées partielles de q_α on utilise les notations suivantes: $r = (r_1, r_2, \dots, r_p)$ est un vecteur d'indices nonnégatif, $\vec{1} = (1, \dots, 1)$,

$|r| = \sum_{j=1}^p r_j$. Il s'ensuit que la r -ième dérivée partielle d'une fonction $u: \mathbb{R}^p \rightarrow \mathbb{R}$

est notée $u^{(r)} = \frac{\partial^r u}{\partial z_1^{r_1} \dots \partial z_p^{r_p}}$.

Nous proposons $\hat{q}_{n,\alpha}^{(r)}$ comme estimateur de $q_\alpha^{(r)}$, si cette dérivée existe. $\hat{q}_{n,\alpha}^{(r)}$ contient les quantités

$$\hat{f}_n^{(v)}(z) = h^{-p-|v|} \sum_i \hat{K}^{(v)}(z, i) w_n(i)$$

$$\tilde{Q}_{n,\alpha}^{(v)}(z) = h^{-p-|v|} \sum_i \bar{q}_{\alpha,i} \hat{K}^{(v)}(z,i) w_n(i), \quad 0 \leq v \leq r.$$

$\hat{f}_n^{(v)}$ est un estimateur pour $f^{(v)}$ et $\tilde{Q}_{n,\alpha}^{(v)}$ est un estimateur pour $(q_\alpha f)^{(v)}$.

Pour obtenir un estimateur de la variance de $\hat{q}_{n,\alpha}^{(r)}$ et de là des intervalles de confiance non paramétriques on utilise aussi un estimateur \hat{f}_n^{XY} pour la densité commune de X et Y; notée f^{XY} :

$$\hat{f}_n^{XY}(z, z_{p+1}) = h^{-p-1} \sum_i \hat{K}(z, z_{p+1}, i, i_{p+1}) w_n(i, i_{p+1})$$

Cet estimateur est identique à l'estimateur \hat{f}_n : il suffit de tenir compte du fait que le domaine de (X,Y) a la dimension p+1.

Dans le contexte présent on peut aussi définir un estimateur lissé pour la répartition conditionnelle $F^{Y|X}$ de Y. Si le vecteur d'indices $r = (r_1, r_2, \dots, r_p)$ signifie dans le cas $r_j < 0$ que la fonction considérée est intégrée $(-r_j)$ fois par rapport à la j-ième variable, l'estimateur pour $F^{Y|X}$ est

$$\hat{F}_n^{Y|X}(z_{p+1}|X=z) = \begin{cases} \hat{f}_n^{XY(\vec{0}, -1)}(z, z_{p+1}) / \hat{f}_n(z) & \text{si } \hat{f}_n(z) \neq 0 \\ 0 & \text{si } \hat{f}_n(z) = 0 \end{cases}$$

On obtient alors un estimateur analogue à celui de Collomb (1978). Il s'ensuit que nous estimons q_α à partir de l'équation

$$\hat{F}_n^{Y|X}(q_{n,\alpha}^*(z)|X=z) = \alpha.$$

Cet estimateur $q_{n,\alpha}^*$ est aussi un estimateur non paramétrique, le calcul de ses valeurs repose sur une méthode numérique, par exemple la méthode de Newton pour la solution d'une équation. Evidemment cette opération est plus complexe que le calcul de la quantité $\bar{q}_{\alpha,i}$.

Les estimateurs proposés ci-dessus évitent cette difficulté.

3 - LES PROPRIÉTÉS STATISTIQUES DES ESTIMATEURS

Dans le paragraphe là nous donnons la liste des hypothèses et les résultats pour

l'estimation des dérivées partielles de l'ordre r de q_{α} , $0 < \alpha < 1$, en point $z \in \mathbb{R}^p$. Ces conditions assurent la consistance statistique et la propriété de convergence asymptotiquement normale de ces estimateurs. En effet ces propriétés sont des justifications pour l'usage pratique des estimateurs. Les hypothèses sont divisées en trois parties: la première partie concerne la loi commune de la probabilité de X et Y , la deuxième partie la fonction du noyau K , la troisième partie les paramètres h , β et la taille n de l'échantillon.

Les hypothèses suivantes sont valables pour tous les résultats énoncés:

(3.1.1) La densité commune f^{XY} des variables aléatoires X et Y existe (f dénote la densité de X).

Pour tous x dans un voisinage de z on a:

(3.1.2) $f(x) > 0$

(3.1.3) f a des dérivées continues jusqu'à l'ordre $(r+1)$

(3.1.4) $f^{XY}(x, q_{\alpha}(x)) > 0$.

Soient $B(y|x) = P(Y \leq y | X=z)$, $B^{-1}(u|x) = \sup\{y | B(y|x) \leq u\}$ pour $0 < u < 1$.

(3.1.5) B^{-1} a des dérivées continues par rapport à (u, x) jusqu'à l'ordre $(2, r+1)$ dans un voisinage de (α, z) .

(3.2.1) $K: \mathbb{R} \rightarrow \mathbb{R}$ a des dérivées continues jusqu'à l'ordre $d + \max_j r_j$ avec $d = 2 \text{Pe}((p+|r|+3)/2)$

(3.2.2) $K(t) = 0$ pour $t \notin]-A, A[$ où A est une constante positive.

(3.2.3) $\int_{\mathbb{R}} K(t) dt = 1$

(3.3.1) $h(n) > 0$, $\beta(n) > 0$

(3.3.2) $\lim_{n \rightarrow \infty} h(n) = 0$ $\lim_{n \rightarrow \infty} \beta(n) = 0$ $\lim_{n \rightarrow \infty} \beta/h = 0$

(3.3.3) $\lim_{n \rightarrow \infty} (\beta/h)^d h^{-|r|} = 0$

(3.3.4) $\lim_{n \rightarrow \infty} nh^{p+1} = \infty$, $\lim_{n \rightarrow \infty} nh^{p+2|r|} = \infty$, $\lim_{n \rightarrow \infty} nh^{p+2|r|+2} = 0$

(3.3.5) $\lim_{n \rightarrow \infty} nh^p \left(\frac{\beta}{h}\right)^{2d} = 0$

(3.3.6) $\lim_{n \rightarrow \infty} n\beta^p = \infty$

(3.3.7) $\lim_{n \rightarrow \infty} n\beta^p h = \infty$

Le lemme suivant montre que les hypothèses sur h , β , n peuvent être satisfaites.

LEMME 1.

Les conditions (3.3.1) - (3.3.7) sont réalisées si on choisit:

$\beta(n) = Cn^{-\epsilon}$ où C est une constante positive

$h(n) = \beta Pe(n^\Delta)$

Si $|r| = 0$:

$\epsilon = (p+1)^{-1}$, $[(p+1)(2d-p)]^{-1} < \Delta < [(p+1)(p+2)]^{-1}$

Si $|r| > 1$:

$\epsilon = (p+|r|)^{-1}$, $|r|[(p+|r|)(p+2|r|)]^{-1} < \Delta < (|r|+2)[(p+|r|)(p+2|r|+2)]^{-1}$

Le lemme est démontré seulement par calcul. Le choix de h et β selon le lemme 1 est lié à la propriété de calculabilité complète qui est nécessaire pour l'application pratique des estimateurs (voir le paragraphe 5).

THÉORÈME 1.

Soit v un vecteur d'indices $0 \leq v \leq r$.

$$E \hat{f}_n^{(v)}(z) - f^{(v)}(z) = O(h) + O\left(\left(\frac{\beta}{h}\right)^d h^{-|v|}\right)$$

$$\lim_{n \rightarrow \infty} \text{Var} (\sqrt{nh^p} h^{|v|} \hat{f}_n^{(v)}(z)) = f(z) \prod_{j=1}^p \int (K_j^{(v_j)}(t))^2 dt$$

$$E \tilde{Q}_{n,\alpha}^{(v)}(z) - (q_\alpha(z)f(z))^{(v)} = O\left(\left(\frac{\beta}{h}\right)^d h^{-|v|}\right) + O(h)$$

$$\lim_{n \rightarrow \infty} \text{Var} (\sqrt{nh^p} h^{|v|} \tilde{Q}_{n,\alpha}^{(v)}(z)) = \frac{\alpha(1-\alpha)}{\left(\frac{\partial B}{\partial u}(q_\alpha(z)|z)\right)^2} + (q_\alpha(z))^2 f(z) \prod_{j=1}^p \int (K_j^{(v_j)}(t))^2 dt$$

Le théorème 1 est un résultat préliminaire au théorème 2. Toutefois il énonce immédiatement la convergence en probabilité de $\hat{f}_n^{(v)}$ et $\tilde{Q}_{n,\alpha}^{(v)}$ et l'ordre de cette convergence, qui est $O(h + (\frac{\beta}{h})^d h^{-|v|} + (nh^p)^{-1/2} h^{-|v|})$.
Utilisant le lemme 1 cet ordre peut être exprimé par une puissance (de valeur négative) de n .

THÉOREME 2.

Les estimateurs $\hat{f}_n^{(v)}$, \hat{f}_n^{XY} , $\tilde{q}_{n,\alpha}^{(v)}$ et $\hat{q}_{n,\alpha}^{(v)}$ $0 \leq v \leq r$, sont consistents au sens suivant:

Pour tout z qui satisfait les conditions (3.1.2) - (3.1.5) on a, pour tout $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\hat{f}_n^{(v)}(z) - f^{(v)}(z)| > \epsilon) = 0$$

$$\lim_{n \rightarrow \infty} P(|\tilde{q}_{n,\alpha}^{(v)}(z) - (q_\alpha(z)f(z))^{(v)}| > \epsilon) = 0$$

$$\lim_{n \rightarrow \infty} P(|\hat{f}_n^{XY}(z, q_\alpha(z)) - f^{XY}(z, q_\alpha(z))| > \epsilon) = 0$$

$$\lim_{n \rightarrow \infty} P(|\hat{q}_{n,\alpha}^{(v)}(z) - q_\alpha^{(v)}(z)| > \epsilon) = 0$$

Les quatre énoncés du théorème reflètent les phases de sa démonstration.

En effet le dernier résultat signale la convergence en probabilité de nos estimateurs $\hat{q}_{n,\alpha}^{(v)}$ $0 < \alpha < 1$, $0 \leq v \leq r$.

Le théorème 2 est une conséquence des résultats du théorème 1 sur les biais et les variances et de la définition de $\hat{q}_{n,\alpha}$ comme quotient.

Le théorème 3 énonce la propriété de convergence asymptotiquement normale.

THÉOREME 3.

(3A) La variable aléatoire

$U_r(z) = \sqrt{nh^p} h^{|r|} (\hat{q}_{n,\alpha}^{(r)}(z) - q_\alpha^{(r)}(z))$ est asymptotiquement gaussienne avec espérance 0 et variance

$$\rho(z) = \frac{\alpha(1-\alpha)}{f(z) \left(\frac{\partial B}{\partial u} (q_\alpha(z)|z) \right)^2} \prod_{j=1}^p \int (K^{(r,j)}(t))^2 dt$$

(3B) Par ailleurs

$$\tilde{U}_r(z) = (\hat{f}_n(z))^{-1/2} \hat{f}_n^{XY}(z, \hat{q}_{n,\alpha}(z)) U_r(z)$$

est de même asymptotiquement gaussienne avec espérance 0 et variance

$$\tilde{\rho}(z) = \alpha(1-\alpha) \prod_{j=1}^p \int (K^{(v_j)}(t))^2 dt$$

(3C) Soient z_1, \dots, z_1 des points différents de \mathbb{R}^p qui satisfont (3.1.2) - (3.1.5). Alors le vecteur aléatoire $(\tilde{U}_r(z_1), \dots, \tilde{U}_r(z_1))$ est asymptotiquement gaussien avec espérance 0 et une matrice diagonale des covariances avec éléments diagonaux $\tilde{\rho}(z_i)$.

Les trois parties du théorème 3 reflètent les trois phases de la démonstration. La partie (3C) du théorème 3 permet de définir des intervalles asymptotiques de confiance pour le vecteur aléatoire $(q_\alpha^{(v)}(z_1), \dots, q_\alpha^{(v)}(z_m))$, m un nombre entier quelconque. Il ne suffit pas à utiliser $\tilde{U}_r(z)$. Cette variable aléatoire n'effectue pas tels intervalles de confiance non paramétrique parce que la variance de $\tilde{U}_r(z)$ n'est pas connue. Toutefois la variance de $\tilde{U}_r(z)$ est connue par la partie (3B). De là on peut juger des propriétés de constance, linéarité et convexité de q_α .

Mais testant ces propriétés en le sens plus strictement mathématique il faut utiliser un résultat sur la loi limite de la variable aléatoire $\sup_z |\tilde{U}_r(z)|$.

Tel résultat pour l'estimateur à noyau discret de la régression se trouve dans Schlee (1980c). De même les tests statistiques sur la fonction de régression sont décrites. Les résultats correspondants pour q_α sont réservés à un article prochain. Pour le cas de la régression des autres références sont mentionnées par Collomb (1981).

4 - UN RESUMÉ DES DÉMONSTRATIONS DES THÉORÈMES

Les lemmes suivants 2 - 5 sont souvent utilisés dans les démonstrations des théorèmes du paragraphe 3.

Les hypothèses de ces lemmes sont (correspondant aux hypothèses (3.1.1)-(3.3.7)):

$K: \mathbb{R} \rightarrow \mathbb{R}$ est une fonction telle que

(4.1) K est continue pour tous $t \in \mathbb{R}$ et

(4.2) $K(t) = 0$ pour $t \notin]-A, A[$, où A est une constante positive.

$u: \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction telle que

(4.3) u est continue dans un voisinage d'un certain point $z \in \mathbb{R}^p$.

Le paramètre $\beta > 0$ est une fonction d'un paramètre $h > 0$ tel que

$$(4.4) \quad \lim_{h \rightarrow 0} \beta/h = 0$$

(4.5) ξ_i est un point arbitraire de l'intervalle $I(i)$.

LEMME 2.

$$(4.6) \quad \lim_{h \rightarrow 0} \sum_i \widehat{K}(z,i) (\beta/h)^p = (\int K(t) dt)^p$$

$$(4.7) \quad \lim_{h \rightarrow 0} \sum_i \widehat{K}(z,i) u(\xi_i) (\beta/h)^p = (\int K(t) dt)^p u(z)$$

(4.8) Si $z \neq y$

$$\lim_{h \rightarrow 0} \sum_i \widehat{K}(z,i) \widehat{K}(y,i) u(\xi_i) (\beta/h)^p = 0$$

La démonstration se trouve dans Schlee(1979).

LEMME 3.

Si en outre de (4.1) - (4.5) $K: \mathbb{R} \rightarrow \mathbb{R}$ a des dérivées continues jusqu'à l'ordre (2l) pour tous $t \in \mathbb{R}$ et pour un certain $l \geq 0$, il suit

$$(4.9) \quad \sum_i \widehat{K}(z,i) (\beta/h)^p = (\int K(t) dt)^p + o((\beta/h)^{2l}) .$$

La démonstration repose sur l'utilisation de la formule bien connue

$$\begin{aligned} & -A \int_A^A K(t) dt - T(\beta/h) = \\ & = \sum_{v=1}^{l-1} (\beta/h)^{2v} \frac{B_{2v}}{(2v)!} \{ K^{(2v-1)}(A) - K^{(2v-1)}(-A) \} - (\beta/h)^{2l} \frac{B_{2l}}{(2l)!} {}_{2A} K^{(2l)}(\xi), \end{aligned}$$

où $T(\beta/h)$ est la somme trapézoïdale de l'intégrale avec longueur des intervalles β/h et $-A < \xi < A$ et B_{2v} sont les nombres de Bernoulli.

LEMME 4.

Si en outre de (4.1) - (4.5) $K: \mathbb{R} \rightarrow \mathbb{R}$ a des dérivées continues jusqu'à l'ordre d et $u: \mathbb{R}^p \rightarrow \mathbb{R}$ a des dérivées continues jusqu'à l'ordre $(r+1)$ dans un voisinage de z et $d \geq (\max_j r_j + 21)$, il suit pour $v \leq r$

$$\begin{aligned}
 (4.10) \quad h^{-|v|} \sum_i \hat{K}^{(v)}(z, i) u(\xi_i) (\beta/h)^p &= u^{(v)}(z) (\int K(t) dt)^p + \\
 &+ \sum_{s: |s|=|v|+1}^{|r|} u^{(s)}(z) h^{|s|-|v|} \left[\prod_{j=1}^p \frac{v_j!}{s_j!} \int t^{s_j - v_j} K(t) dt + \right. \\
 &\left. + O(\beta/h) \prod_{j=1}^p \int K^{(v_j)}(t) dt \right] + O(h^{|r|-|v|+1}) + O((\beta/h)^{21} h^{-|v|}) .
 \end{aligned}$$

La démonstration s'effectue en utilisant la formule de Taylor pour la fonction u ainsi que les lemmes 2 et 3.

LEMME 5.

Soit donnée la partition en intervalles du paragraphe 1. Soit h une fonction de n et de là même β . Soit $w(i)$ la probabilité telle que $X \in I(i)$. Si la densité f de X est continue dans un voisinage de $I(i)$, alors

$$(4.11) \quad w(i) = \beta^p f(\xi_i) .$$

Soit $z \in \mathbb{R}^p$ et $f(x) \geq \delta > 0$ dans un voisinage de z contenant $I(i)$ et $I(1)$ pour n suffisamment grand. Si $\lim_{n \rightarrow \infty} n \beta^p = \infty$, il suit pour des nombres entiers non négatifs j_1, j_2 :

$$(4.12) \quad \lim_{n \rightarrow \infty} E((w(i)/w_n(i))^{j_1} (w(1)/w_n(1))^{j_2}) = 1 .$$

La démonstration de (4.11) est évidente, la démonstration de (4.12) suit de la convergence en probabilité de la fréquence relative $w_n(i)$.

La démonstration du théorème 3 utilise les deux résultats intermédiaires suivants:

LEMME 6.

Sous les conditions du paragraphe 3 le vecteur aléatoire

$$\sqrt{nh^p} h^{|v|} (\hat{q}_{n,\alpha}^{(v)}(z) - E\hat{q}_{n,\alpha}^{(v)}(z) , \hat{f}_n^{(v)}(z) - E\hat{f}_n^{(v)}(z))$$

est asymptotiquement gaussien avec espérance mathématique 0 et matrice de covariance

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{(\frac{\partial B}{\partial u}(q_\alpha(z)|z))^2} + (q_\alpha(z))^2 & q_\alpha(z) \\ q_\alpha(z) & 1 \end{pmatrix} f(z) \prod_{j=1}^p \int (K^{(v_j)}(t))^2 dt$$

LEMME 7.

Sous les conditions du paragraphe 3 la variable aléatoire

$$\sqrt{nh^p} h^{|r|} (\hat{q}_{n,\alpha}^{(r)}(z) - q_\alpha^{(r)}(z))$$

a la même loi asymptotique que la variable aléatoire

$$\sqrt{nh^p} h^{|r|} f^{-1}(z) [(\hat{q}_{n,\alpha}^{(r)}(z) - E\hat{q}_{n,\alpha}^{(r)}(z)) - q_\alpha(z)(\hat{f}_n^{(r)}(z) - E\hat{f}_n^{(r)}(z))] .$$

La démonstration du lemme 6 utilise le lemme 4 de Schlee(1978) et les lemmes précédentes. La démonstration du lemme 7 utilise essentiellement des résultats bien connus (voir par exemple Billingsley(1968)) sur la convergence des variables aléatoires et les conditions (3.3.1) - (3.3.7) et le lemme 4.

5 - PROPRIÉTÉS ALGORITHMIQUES ET NUMÉRIQUES

L'ensemble des données $(X_1, Y_1) 1=1,2,\dots,n$ présente l'information complète. Si la taille n est très grande, cette forme de l'information n'est pas pratique. À cause de cela on ramène les données à la moyenne arithmétique, la déviation quadratique moyenne, les fréquences relatives ou à un estimateur pour le α -quantile conditionnel ou à un estimateur pour la fonction de régression ou à d'autres quantités. La valeur essentielle des estimateurs est la caractérisation de l'information contenue dans les données par des quantités concises. En effet il s'agit d'une réduc-

tion des données de taille n à des quantités moins nombreuses. On ne peut pas renoncer à cette propriété de réduction dans le cas du α -quantile conditionnel et de la fonction de régression au raison de la pratique. Jusqu'à présent on n'a pas fait attention suffisante à ce problème. Il n'est pas trivial parceque par exemple les estimateurs à noyau pour la régression et pour la densité de probabilité ne comprennent pas une telle réduction des données. Cela reste aussi vrai quand on considère l'estimateur seulement dans un intervalle fini. En outre il n'est pas possible de choisir une fonction du noyau pour obtenir cette réduction. Il est seulement possible à calculer les valeurs des estimateurs pour tous points utilisant chaque fois à nouveau l'ensemble des données $(X_1, Y_1) \quad l=1, 2, \dots, n$. Par contre les estimateurs proposés ici comprennent une réduction des données si on pose deux conditions supplémentaires:

(5.1) K est une fonction polynôme par morceaux

(5.2) $h=l\beta$ avec l entier et $l > 0$.

Sous ces conditions les estimateurs $\hat{f}_n^{(v)}$, \hat{f}_n^{XY} , $\hat{Q}_{n,\alpha}^{(v)}$, $0 \leq v \leq r$, sont aussi des fonctions polynôme par morceaux parceque ces estimateurs ont la forme

$$(5.3) \sum_i \Gamma_i \hat{K}^{(v)}(z, i)$$

et $\hat{K}^{(v)}(z, i)$ est une fonction polynôme par morceaux par (5.1) et (5.2). Γ_i est independant de z et il est un produit de $h^{-1}, w_n(i), \bar{q}_{\alpha, i}$ correspondant à l'estimateur considéré. Comme les estimateurs $\hat{f}_n^{(v)}$, \hat{f}_n^{XY} , $\hat{Q}_{n,\alpha}^{(v)}$ sont des fonctions polynôme par morceaux il suffit de calculer ces estimateurs pour les points de la partition en intervalles. Les valeurs en d'autres points peuvent être obtenues exactement par ces valeurs.

Comme $\hat{q}_{n,\alpha}^{(r)}$ est une fonction rationnelle des estimateurs $\hat{f}_n^{(v)}$ et $\hat{Q}_{n,\alpha}^{(v)}$ il peut être calculé par multiplication, sommation et division. Elles n'existent pas des erreurs causées par une méthode ayant besoin des nombres infinis des démarches et qui en conséquences doit être rompue après des nombres finis des démarches. Je voudrais appeler les deux propriétés de réduction des données et de calculabilité sauf erreurs numériques d'un estimateur la propriété de calculabilité complète. Par rapport à la pratique je pense, que cette propriété est une propriété très désirable.

Si on choisit une propre fonction du noyau les estimateurs à noyau peuvent être aussi calculés pour chaque point sauf erreurs numériques. Mais la réduction des données peut être accomplie seulement par une modification. Cette modification est

comme il suit:

L'estimateur à noyau est calculé pour des points d'un treillis assez fin. Les valeurs pour les autres points sont obtenues par interpolation.

Mais on n'a pas recherché les propriétés statistiques de cet estimateur modifié, on a recherché seulement les propriétés d'un estimateur théorique, l'estimateur à noyau qui n'est pas réellement utilisé.

Cette modification par interpolation a les mêmes effets que les discretisations des estimateurs proposés ici. Pourquoi ne pas opérer directement? En outre les propriétés statistiques des estimateurs à noyau discret sont données dans le présent article pour le α -quantile conditionnel. Le cas de la régression et de la densité de probabilité se trouvent respectivement dans Schlee (1979, 1980a) et Schlee (1978). Comment doit on choisir K , β et h de manière à satisfaire (5.1), (5.2) et (3.2.1) - (3.3.7)? Si on prend h et β selon le lemme 1 les conditions (5.2) et (3.3.1) - (3.3.7) sont réalisées et la quantité 1 est égale à $Pe(n^{\Delta})$.

Pour la fonction K on a deux possibilités:

(5.3.1) fixer A (de 3.2.2) et augmenter l'ordre des polynômes

ou

(5.3.2) fixer l'ordre des polynômes et augmenter la quantité A .

La possibilité (5.3.2) est décrite dans Schlee (1979). La possibilité (5.3.1) peut être décrite comme suit. Si K possède des dérivées jusqu'à l'ordre 1 (d'après (3.2.1)), alors K doit être choisi comme fonction du polynôme d'ordre $l+1$ par morceaux. Les coefficients des polynômes sont les solutions des équations linéaires données par la condition que K a une dérivée continue d'ordre $l+1$.

Nous donnons deux exemples:

un polynôme d'ordre 2

$$K(t) = \begin{cases} -t^2/4 + 1/2 & 0 \leq t \leq 1 \\ (t-1)^2/4 - (t-1)/2 + 1/4 & 1 \leq t \leq 2 \\ 0 & t \geq 2 \end{cases}$$

$$K(t) = K(-t) \quad t \leq 0$$

un polynôme d'ordre 3

$$K(t) = \begin{cases} t^3/2 - t^2 + 2/3 & 0 \leq t \leq 1 \\ -1/6(t-1)^3 + 1/2(t-1)^2 - 1/2(t-1) + 1/6 & 1 \leq t \leq 2 \\ 0 & t \geq 2 \end{cases}$$

$$K(t) = K(-t) \quad t \leq 0$$

6 - RÉFÉRENCES

- Angers, C. (1979) "Simultaneous estimation of percentile curves with application to salary data" *Journal Amer. Statist. Assoc.* 74, 621-625
- Billingsley, P. (1968) "Convergence of Probability Measures" John Wiley & Sons, New York
- Brown, G.W. et Mood, A.M. (1951) "On median tests for linear hypotheses" *Proc. Second Berkeley Symp. Math. Statist. Prob.* 1, 159-166
- Collomb, G. (1978) "Estimation non paramétrique de la régression: régressogramme et methode du noyau" Publication du Laboratoire de Statistique de l'Université Paul Sabatier, No 07-78, Vannes 1978
- Collomb, G. (1981) "Estimation non paramétrique de la régression: revue bibliographique" *International Statistical Review* 49, 75-93
- Csörgö, M. et Révész, P. (1978) "Strong approximation of the quantile process" *The Annals of Statistics* 6, 882-894
- Griffiths, D. et Willcox, M. (1978) "Percentile regression: A parametric approach" *Journal Amer. Statist. Assoc.* 73, 496-498
- Hogg, R.V. (1975) "Estimates of percentile regression lines using salary data" *Journal Amer. Statist. Assoc.* 70, 56-59
- Schlee, W. (1979) "Nonparametric estimation of curves" *Serdica Bulg. math. publ.* 5 186-203
- Schlee, W. (1980a) "Estimation non paramétrique des propriétés de la courbe de régression" Dans: *Data Analysis and Informatics* (E. Diday et al., Eds), 109-119 North-Holland Publishing Company
- Schlee, W. (1980b) "Qualitätskontrolle körniger Mineralstoffe" Dans: *Operations Research Proceedings 1980*, 181-187, (G. Fandel et al., Eds), Springer-Verlag Berlin 1981
- Schlee, W. (1980c) "Nonparametric tests of the monotony and convexity of regression" *Colloquium on Nonparametric Statistical Inference*, Budapest 23-28 Juin 1980, à paraître dans les actes du colloque
- Turner, D.L. et Bowden, D.C. (1977) "Simultaneous confidence bands for percentile lines in the general linear model" *Journal Amer. Statist. Assoc.* 72, 886-889