L-ESTIMATION DES QUANTILES CONDITIONNELS

Ines JLASSI [1] & Ali GANNOUN [2] & Salah KHARDANI [3]

[1] *Faculté des Sciences de Monastir, jlassi.ines.fsm@gmail.com*
[2] *Université Montpellier 2, gannoun@math.univ-montp2.fr*
[3] *Ecole Nationale d'Ingénieurs de Monastir, khardani_salah@yahoo.fr*

Résumé. Un intéressant problème dans l'étude de l'interdépendance entre deux variables aléatoires X et Y est l'estimation des quantiles conditionnels. L'idée, alors, est de construire un estimateur non paramétrique des quantiles conditionnels de Y sachant X basé sur la méthode des noyaux en utilisant des statistiques d'ordre. Celui-ci a été implémenté dans R afin d'évaluer le comportement numérique. Une comparaison des performances de cet estimateur avec autres estimateurs est établie.

Mots-clés. Estimation non-paramétrique, Quantile conditionnel, Statistique d'ordre, ...

Abstract. An interesting problem in a study of the interdependence between two random variables X and Y is to estimate the conditional quantiles. The aim of this Work is to give a new nonparametric estimation of conditional quantile of Y for a given value of X using a kernel type of estimator based on order statistics. We have conducted a numerical study to evaluate the numerical behavior of the estimator. A performance comparison with an other existing estimator is also made.

Keywords. Conditional quantile, Nonparametric estimation, Order statistics, ...

1 Introduction

L'idée est de construire des estimateurs non paramétrique des quantiles conditionnels de Y sachant X. Les quantiles conditionnels sont particulièrement intéressants lorsqu'il apparaît que la moyenne conditionnelle seule ne permet pas de réprésenter convenablement l'impact de la covariable X sur la variable dépendante Y.

On considère deux variables quantitatives: une variable d'interêt Y et une covariable X.

Le **quantile conditionnel** d'ordre λ ($\lambda \in (0,1)$) de la variable Y sachant que $X = x$ est définie de la manière suivante:

$$q_\lambda(x) = F^{-1}(\lambda|x) = \inf\{y \in \mathbb{R} | F(y|x) \geq \lambda\}$$

où $F(.|x)$ désigne la fonction de répartition conditionnelle de Y sachant $X = x$.

2 Estimateur des quantiles conditionnels

Une principale application du calcul des quantiles conditionnels est la construction des courbes de référence qui sert à comparer un individu i à la population de référence et détecter si cet individu est "hors-norme". Les courbes de référence sont utilisées dans les études biomédicales ou biométriques, les applications industrielles, ...

Plusieurs approches ont été développées pour l'estimation des quantiles conditionnels. On va s'intéresser à l'estimation non paramétrique des quantiles conditionnels basée sur la méthode des noyaux.

L'estimation non paramétrique des quantiles conditionnels a reçu un grand intérêt depuis (1969) quand Roussas (1969) a montré la convergence et la normalité asymptotique des estimations à noyau. On peut citer les travaux de Rosenblatt (1956), (1969), Parzen (1962) et Nadaraya (1964). Dans le même esprit, on propose une classe d'estimateurs non paramétriques des quantiles conditionnels de Y pour une valeur donnée de X.

On considère m observations $(X_1, Y_1), (X_2, Y_2), ..., (X_m, Y_m)$ d'un couple de variables (X, Y) à valeurs dans $I\!R \times I\!R$ telque Y_i dépend de X_i pour $i = 1, ..., m$.

En s'inspirant de l'idée de Shie-Shien-YANG (1985) on propose l'estimateur des quantiles conditionnels suivant:

$$Q_{m,n}^{\lambda}(x) = \frac{1}{n} \sum_{i=1}^{n} Z_i^x h_n^{-1} K\left(\frac{\frac{i}{n} - \lambda}{h_n}\right)$$

avec K, appelée noyau, est une densité de probabilité, le paramètre h_n permet de contrôler le lissage appliqué aux données et où les $(Z_i^x)_{i=1,\ldots,n}$ sont les quantiles d'ordre $\frac{i}{n}$ vérifiant l'équation suivante:

$$F_m^{-1}\left(\frac{i}{n}|x\right) = Z_i^x$$

avec $F_m(.|x)$ est la fonction de répartition conditionnelle empirique de Y sachant $X = x$ définie comme suit: (Nadaraya Watson (1964))

$$F_m(y|x) = \frac{\sum_{i=1}^{m} 1_{\{Y_i \le y\}} K\left(\frac{x - X_i}{h_m}\right)}{\sum_{i=1}^{m} K\left(\frac{x - X_i}{h_m}\right)}.$$

3 Etude pratique

A partir de l'écriture de l'estimateur proposé, une première étude sur simulation a été faite afin de regarder les performances numériques de cet estimateur avec différents modèles et différentes tailles d'échantillons. Un tableau d'erreur quadratique moyenne (MSE) nous a permis de faire une comparaison des performances de l'approche préconisée avec celles obtenues avec l'estimateur empirique des quantiles conditionnels et on a observé que notre estimateur est bien plus performant.

Bibliographie

[1] Nadaraya, E.A. (1964), Some new estimates for distribution functions, *Theory of Probability and Its Applications* 9, 497-500.
[2] Parzen, E. (1962), On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33, 1065-1076.
[3] Rosenblatt, M. (1956), Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* 27, 832-837.
[4] Rosenblatt, M. (1969), Conditional probability density and regression estimators, in: P.R. Krishnaiah, ed., *Multivariate Analysis II (Academic Press, New York and London)* pp. 25-31.
[5] Roussas,G.G., (1969), Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.*
[6] Yang, S.S., (1985), A Smooth Nonparametric Estimator of a Quantile Function, *Journal of the American Statistical Association*, 1004-1011