Distribution-Free Estimation of Conditional Quantile with Different Types of Conditioning Attributes

P. Kulczycki[1,4], M. Charytanowicz[2,4], A.L. Dawidowicz[3]

[1] Cracow University of Technology,
Department of Automatic Control and Information Technology
Warszawska 24, 31-155 Cracow, Poland
E-mail: kulczycki@pk.edu.pl

[2] John Paul II Catholic University of Lublin,
Institute of Mathematics and Computer Science
Konstantynów 1H, 20-708 Lublin, Poland

[3] Jagiellonian University, Institute of Mathematics
Łojasiewicza 6, 30-348 Cracow, Poland

[4] Polish Academy of Sciences, Systems Research Institute,
Center for Information Technology Data Analysis
Newelska 6, 01-447 Warsaw Poland

Abstract: The paper presents a distribution-free procedure for calculating the value of a conditional quantile estimator. Thanks to a clear, near intuitive interpretation, the practical implementation of this method is very simple and it can easily be modified or generalized depending on the individual needs of atypical applications. In particular, conditioning variables can be taken into account – not only continuous (real), but also binary, discrete and categorized, or any of their combinations.

Keywords: *conditional quantile; nonparametric estimation; conditioning variables of continuous, binary, discrete and categorized types; numerical algorithm*

1. Introduction

Consider the one-dimensional random variable Y, termed below as a describing variable. Let also be given the n_W-dimensional random variable W, called hereinafter a conditioning variable. Their composition $Z = \begin{bmatrix} Y \\ W \end{bmatrix}$ is a random variable of the dimension $n_W + 1$. Assume that distributions of the variables Z and, in consequence, W have densities, denoted below as $f_Z : \mathbb{R}^{n_W+1} \to [0, \infty)$ and $f_W : \mathbb{R}^{n_W} \to [0, \infty)$, respectively. Let also be given the so-called conditioning value, that is the fixed value of a conditioning random variable $w^* \in \mathbb{R}^{n_W}$, such that

$$f_W(w^*) > 0 \quad . \tag{1}$$

Then the function $f_{Y|W=w^*} : \mathbb{R} \to [0,\infty)$ given by

$$f_{Y|W=w^*}(y) = \frac{f_X(y,w^*)}{f_W(w^*)} \qquad \text{for every} \quad y \in \mathbb{R} \tag{2}$$

constitutes a conditional density of probability distribution of the random variable Y for the conditioning value w^*. A quantile of the order $r \in (0,1)$ with the condition $w^* \in \mathbb{R}^{n_W}$ is every number $q_{r|w^*} \in \mathbb{R}$, such that

$$\int_{-\infty}^{q_{r|w^*}} f_{Y|W=w^*}(y)\,\mathrm{d}y = r \quad . \tag{3}$$

If the support of the function $f_{Y|W=w^*}$ is connected, then the quantile is unique. The conditional quantile $q_{r|w^*}$ constitutes therefore the refinement of the “classic” quantile by using the information that the conditioning random variable, in a specific situation, has taken the value w^*.

This paper presents a procedure for calculating the estimator of a conditional quantile, based on the statistical kernel estimator methodology. Its nonparametric nature implies the worked out procedure is independent of types of random variable distributions. The key advantage is, however, its simplicity and ease of interpretation and possibilities of creating individual modifications suitable in practice for specific particular applications, especially the potential generalization of conditioning variables, not only the continuous (real) but also – as should be clearly underlined – binary, discrete and categorical (ordered and unordered), as well as their compositions.

Classic methods for the conditional quantile estimation, together with reach subject literature can be found in the monograph [7]. A survey of procedures for the basic unconditional case is contained in the articles [3-4]. As an example of the non statistical approach, see e.g. the paper [15].

2. An algorithm for conditional quantile estimation

Let the *n*-dimensional random variable X be given, with a distribution characterized by the density f. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0,\infty)$, calculated using experimentally obtained values for the *m*-element random sample

$$x_1,\ x_2, \ldots, x_m \quad , \tag{4}$$

in its basic form is defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \quad , \tag{5}$$

where $m \in \mathbb{N} \setminus \{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K : \mathbb{R}^n \to [0, \infty)$ of unit integral $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. The choice of form of the kernel K and the calculation of the smoothing parameter h is made most often with the criterion of the mean integrated square error.

Thus, the choice of the kernel form has – from a statistical point of view – no practical meaning and thanks to this, it becomes possible to take into account primarily properties of the estimator obtained, or calculational aspects, both advantageous from the point of view of the applicational problem under investigation. In practice, for the one-dimensional case, the function K is assumed most often to be the density of a common probability distribution. In the multidimensional case, two natural generalizations of the above concept are used**:** radial and product kernels. However, the former is somewhat more effective, although from an applicational point of view, the difference is immaterial and the product kernel – significantly more convenient in analysis – is often favored in practical problems. The n-dimensional product kernel K can be expressed as

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = K_1(x_1) K_2(x_2) \ldots K_n(x_n) \quad , \tag{6}$$

where K_i for $i = 1, 2, \ldots, n$ denotes the previously-mentioned one-dimensional kernels, while the expression h^n appearing in the basic formula (5) should be replaced by $h_1 \cdot h_2 \cdot \ldots \cdot h_n$, the product of the smoothing parameters for particular coordinates.

The fixing of the smoothing parameter h has significant meaning for quality of estimation. Fortunately – from the applicational point of view – many suitable procedures for calculating the value of the parameter h on the basis of random sample (4) have been worked out.

For broader discussion of the above tasks see [8, 12-13].

The kernel estimators technique will now be used below for the task of conditional quantile estimation. Let a one-dimensional describing random variable Y as well as the n_W-dimensional conditioning random variable W be given. Suppose also the random sample

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \ldots, \begin{bmatrix} y_m \\ w_m \end{bmatrix} \quad , \tag{7}$$

obtained from the variable $Z = \begin{bmatrix} Y \\ W \end{bmatrix}$. The particular elements of this sample are interpreted as the values y_i taken in measurements from the random variable Y, when the conditioning variable W assumes the respective values w_i. Using the methodology presented above, on the basis of sample (7) one can calculate $\hat{f}_Z$, i.e. the kernel estimator of density of the random variable Z probability distribution, while the sample

$$w_1 ,\ w_2 , \dots , w_m \tag{8}$$

gives $\hat{f}_W$ – the kernel density estimator for the conditioning variable W. The kernel estimator of conditional density of the random variable Y probability distribution for the conditioning value w^*, is defined then – as a natural consequence of formula (2) – as the function $\hat{f}_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0,\infty)$ given by

$$\hat{f}_{Y|W=w^*}(y) = \frac{\hat{f}_X(y, w^*)}{\hat{f}_W(w^*)} \quad . \tag{9}$$

If for the estimator $\hat{f}_W$ one uses a kernel with positive values, then the inequality $\hat{f}_W(w^*) > 0$ implied by condition (1) is fulfilled for any $w^* \in \mathbb{R}^{n_W}$. In the case when for the estimators $\hat{f}_Z$ and $\hat{f}_W$ the product kernel (6) is used, applying in pairs the same positive kernels to the estimator $\hat{f}_W$ and to the last n_W coordinates of the estimator $\hat{f}_Z$, then the formula (9) can then be given in the form particularly helpful for practical applications:

$$\hat{f}_{Y|W=w^*}(y) = \frac{1}{h_0 \sum_{i=1}^{m} d_i} \sum_{i=1}^{m} d_i K_0\left(\frac{y - y_i}{h_0}\right) \quad , \tag{10}$$

while h_0, $h_1, \dots, h_{n_W}$ represent smoothing parameters mapped to particular coordinates of the random variable Z (the first h_0 connotes with the describing variable Y, and the rest $h_1, \dots, h_{n_W}$ with subsequent coordinates of the conditioning variable W), the so-called conditioning parameters d_i for $i = 1, 2, \dots, m$ are defined by the following formula:

$$d_i = K_1\left(\frac{w_1^* - w_{i,1}}{h_1}\right) K_2\left(\frac{w_2^* - w_{i,2}}{h_2}\right) \dots K_{n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_W}}\right) \quad , \tag{11}$$

where the particular coordinates of the vectors w^* and w_i have been denoted in a natural manner

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W}^* \end{bmatrix} \quad \text{and} \quad w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{for} \quad i = 1, 2, \dots, m \quad . \tag{12}$$

If one uses the kernels K_1, $K_2, \dots, K_{n_W}$ with positive values, conditioning parameters (11) are also positive. The value of the parameter d_i characterizes the "distance" of the given conditioning value w^* from w_i – that of the conditioning variable for which the *i*-th element of the random sample was obtained. Then estimator (10) can be interpreted as the linear combination of kernels mapped to particular elements of a random sample obtained for the variable Y, when the coefficients of this combination characterize how representative these elements are for the given value w^*.

With respect to the definition of a conditional quantile (3), its natural estimator is the solution of the following equation with the argument $\hat{q}_{r|w^*}$:

$$\int_{-\infty}^{\hat{q}_{r|w^*}} \hat{f}_{Y|W=w^*}(y)\,\mathrm{d}y = r \quad . \tag{13}$$

For estimation of the conditional density $\hat{f}_{Y|W=w^*}$ appearing above, the kernel estimator given in the form (10) will be used. Moreover, a continuous function of positive values should be chosen as the kernel K_0, also so that the function $I : \mathbb{R} \to \mathbb{R}$ such that $I(w) = \int_{-\infty}^{w} K_0(u)\,\mathrm{d}u$ can be expressed by a relatively simple analytical formula. Equation (13) is equivalently described then in the following form:

$$\sum_{i=1}^{m} d_i I\left(\frac{\hat{q}_{r|w^*} - y_i}{h_0} \right) - r \sum_{i=1}^{m} d_i = 0 \quad . \tag{14}$$

If the left side of the above equation is denoted by L, i.e.

$$L(\hat{q}_{r|w^*}) = \sum_{i=1}^{m} d_i I\left(\frac{\hat{q}_{r|w^*} - y_i}{h_0} \right) - r \sum_{i=1}^{m} d_i \quad , \tag{15}$$

then $\lim_{\hat{y}_{w^*} \to -\infty} L(\hat{q}_{r|w^*}) < 0$, $\lim_{\hat{y}_{w^*} \to \infty} L(\hat{q}_{r|w^*}) > 0$, the function L is strictly increasing and its derivative is simply expressed by

$$L'(\hat{q}_{r|w^*}) = \frac{1}{h_0} \sum_{i=1}^{m} d_i\, K_0\left(\frac{\hat{q}_{r|w^*} - y_i}{h_0} \right) \quad . \tag{16}$$

In this situation, the solution of equation (14) can be effectively calculated on the basis of Newton's algorithm [6] as the limit of the sequence $\{\hat{q}_{r|w^*,j}\}_{j=0}^{\infty}$ defined by

$$\hat{q}_{r|w^*,0} = \frac{\sum_{i=1}^{m} d_i \, y_i}{\sum_{i=1}^{m} d_i} \tag{17}$$

$$\hat{q}_{r|w^*,j+1} = \hat{q}_{r|w^*,j} - \frac{L(\hat{q}_{r|w^*,j})}{L'(\hat{q}_{r|w^*,j})} \quad \text{for} \quad j = 0, 1, \ldots \quad , \tag{18}$$

with the functions L and L' being given by dependencies (15)-(16), whereas a stop criterion takes on the form

$$|\, \hat{q}_{r|w^*,j} - \hat{q}_{r|w^*,j-1} \,| \le 0.01 \hat{\sigma}_Y \quad , \tag{19}$$

while $\hat{\sigma}_Y$ denotes the estimator of the standard deviation of the random variable Y.

3. Final remarks

The correct functioning and positive properties of the algorithm presented in this paper were confirmed with detailed numerical verification. It is worth noting that in any case as the sample size increased, the obtained parameter value converged to the theoretical, and the standard deviation to zero. The above asymptotical features are of fundamental significance from an applicational point of view, as they prove that it is possible to obtain any precision wished, although this requires the assurance of a sufficient random sample size. In practice, therefore, the necessity of the right compromise between these quantities is called for.

The procedure presented in this paper has been given in its basic form, easier to implement and calculationally more convenient. A clearer interpretation means it is possible to make individual modifications and generalizations, which may be useful in particular atypical tasks. In particular, one can introduce the different types of particular coordinates of the conditioning random variable W. Namely, the kernel estimator's definition (5) was presented in Section 2 for the most often used in practice continuous (real) random variables, but the same can be also made for binary, discrete and categorical (ordered, too) variables as well as their compositions. The literature concerning this subject is quite broad and varied. For the first case, it is worth quoting the monographs [8 – Section 3.1.8; 12 – Section 6.1.4] as well as the classic paper [2], and for the second [1, 14]. Issues connected with categorical variables can be found in the publications [5, 10-11]. After introducing a binary, discrete and/or categorized variable to the algorithm worked out here, it undergoes practically no changes, apart from technical ones resulting from calculational differences**:** the conditioning parameters d_i consists then only from factors of different types, not only continuous like in formula (11). This property particularly should be underlined considering the modern data analysis tasks, which more and more often take advantage of the many different configurations for particular types of attributes.

A detailed description of the methodology presented here and verified experimentations will be found in the paper [9] currently undergoing publication. The algorithm is given there in the ready-to-use form, together with all formulas, concrete analytical forms of functions used and rules defining parameter values. It can be applied directly without detailed subject knowledge and laborious research.

References

[1] Ahmad, I.A., Cerrito, P.B.: Nonparametric estimation of joint discrete-continuous probability densities with applications, Journal of Statistical Planning and Inference, vol. 41, pp. 349–364, 1994

[2] Aitchison, J., Aitken, C.: Multivariate Binary Discrimination by the Kernel Method, Biometrika, vol. 63, pp. 413–420, 1976

[3] David, C.E., Steinberg, S.M.: *Quantile estimation*; in Encyclopedia of Statistical Sciences, Editors: Kotz, S., Johnson, N.L., Wiley, New York, vol. 7, pp. 408–412, 1986

[4] Dielman, T., Lowry, C., Pfaffenberger, R.: A comparison of quantile estimators, Communications in Statistics – Simulation and Computation, vol. 23, pp. 355–371, 1994

[5] Gaosheng, J., Rui, L., Zhongwen, L.: Nonparametric Estimation of Multivariate CDF with Categorical and Continuous Data, Advances in Econometrics, vol. 25, pp. 291–318, 2009

[6] Kincaid, D., Cheney, W.: *Numerical Analysis*, Brooks/Cole, Pacific Grove, 2002

[7] Koenker, R.: *Quantile Regression*, Cambridge University Press, Cambridge, 2005

[8] Kulczycki, P.: *Estymatory jadrowe w analizie systemowej*, WNT, Warsaw, 2005

[9] Kulczycki, P., Charytanowicz, M., Dawidowicz, A.L.: A Convenient Ready-to-Use Algorithm for a Conditional Quantile Estimator, An International Journal of Applied Mathematics & Information, 2014, accepted for publication

[10] Li, Q., Racine, J.S.: Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, Journal of Business and Economic Statistics, vol. 26, pp. 423–434, 2008

[11] Ouyang, D., Li, Q., Racine, J.: Cross-validation and the estimation of probability distributions with categorical data, Journal of Nonparametric Statistics, vol. 18, pp. 69–100, 2006

[12] Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986

[13] Wand, M.P., Jones, M.C.: *Kernel Smoothing*, Chapman and Hall, London, 1995

[14] Wang, M., van Ryzin, J.: A class of smooth estimators for discrete distributions, Biometrica, vol. 68, pp. 301–309, 1981

[15] White, H.: Nonparametric Estimation of Conditional Quantiles Using Neural Networks, Computing Science and Statistics, vol. 24, pp. 190–199, 1992