UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA, FASCICULUS XXXVII
1999

KERNEL ESTIMATOR OF QUANTILE

BY PIOTR KULCZYCKI AND ANTONI LEON DAWIDOWICZ

Abstract. The subject of the paper is the issue of quantile estimation. To solve this problem, the kernel estimators technique has been applied. The strong consistency of the obtained estimator is shown. The considerations result in a complete usable algorithm for numerically calculating the value of the estimator on the basis of a random sample.

1. Introduction. For statistical purposes, distributions of random variables are most often reported through characteristic parameters describing their fundamental features. Moments, especially mean value and variance, constitute well known examples of such parameters. Another group of characteristic quantities consists of the so-called positional parameters, including quantiles and their functions, which are more directly connected to the distribution function by relating certain points to its assumed values. Frequently, the median (quantile of order 0,5) is treated like the mean, and the quantile deviation – i.e., the difference between quantiles of order 0,75 and 0,25 – can be interpreted similarly to the variance. Special quantiles such as quadriles, deciles and percentiles also appear often in statistical applications [3].

In this paper, the kernel estimators technique will be used to calculate the estimator of the quantile. Presently, due to the expansion of numerical methods, that technique is finding ever wider application. Since kernel estimators are predominantly used when the sample size is rather large, the property of (strong) consistency will be proved. The final result of the considerations presented here will be a complete usable algorithm for specifying the value of the quantile estimator on the basis of a random sample. Because any sort of

Key words and phrases. Estimation of quantile, kernel estimators, numerical calculations.

discretionary statistical research has been eliminated here through the application of optimizing criteria, the proposed method may be successfully adapted to numerical procedures.

A review of alternative methods of quantile estimation can be found in survey papers: [4] for order statistics, and [6] where other kernel concepts are considered.

2. Kernel estimators.

2.1. Kernel estimator of density function. Let (Ω, Σ, P) be a probability space. Consider a real random variable $X : \Omega \longrightarrow \mathbf{R}$, whose distribution has the density function f. In practice, its estimator $\widehat{f}$ is calculated on the basis of the value of an m-element simple random sample $x_1, x_2, \ldots, x_m \in \mathbf{R}$. The fundamental form of the kernel estimator can then be defined by the formula

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

where

$$h > 0, \tag{2}$$

while the measurable function $K : \mathbf{R} \longrightarrow [0, \infty)$ fulfills the condition

$$\int_{\mathbf{R}} K(x)\, dx = 1 \tag{3}$$

and for every $x \in \mathbf{R}$:

$$K(x) = K(-x), \tag{4}$$

$$K(0) \geq K(x). \tag{5}$$

The function K is called the kernel, whereas the constant h is known as the smoothing parameter.

If the function K is Borel and fulfills the condition

$$\lim_{x \to \infty} x\, K(x) = 0, \tag{6}$$

while the value of the smoothing parameter h is selected in such a way that

$$\lim_{m \to \infty} h = 0, \tag{7}$$

$$\lim_{m \to \infty} m\, h = \infty, \tag{8}$$

then at every point of continuity x of the density function f, the kernel estimator is strongly consistent, i.e.,

$$P\left(\lim_{m\to\infty}\widehat{f}(x)=f(x)\right)=1, \tag{9}$$

and therefore also consistent:

$$\lim_{m\to\infty}P\left(|\widehat{f}(x)-f(x)|\geq\varepsilon\right)=0 \quad \text{for every } \varepsilon>0. \tag{10}$$

Detailed discussions of this subject, especially the procedures for choosing the form of the function K and calculating the value of the smoothing parameter h, can be found in books [2,5,7].

2.2. Kernel estimator of distribution function. To carry forward the concept sketched in the previous subsection, the natural estimator of the distribution function, denoted hereinafter as $\widehat{F}$, is defined by the formula

$$\widehat{F}(x)=\int_{-\infty}^{x}\widehat{f}(y)\,dy. \tag{11}$$

Condition (3) guarantees the existence of the primitive $I:\mathbf{R}\longrightarrow[0,1]$ of the kernel K, i.e.,

$$I(x)=\int_{-\infty}^{x}K(y)\,dy. \tag{12}$$

The kernel estimator of the distribution function can therefore be expressed as

$$\widehat{F}(x)=\frac{1}{m}\sum_{i=1}^{m}I\left(\frac{x-x_i}{h}\right). \tag{13}$$

The property of its (strong) consistency, under very mild assumptions, will be shown below. First, however, the notion of empirical ergodicity will be presented.

Consider the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$ defined on a common probability space (Ω,Σ,P), as well as the corresponding sequence of its realizations $\{x_i\}_{i=1}^{\infty}$. For an arbitrarily fixed $m\in\mathbf{N}\setminus\{0\}$, the mapping $\mathcal{P}_m:\mathcal{B}(\mathbf{R})\longrightarrow[0,1]$ given by the formula

$$\mathcal{P}_m(B)=\frac{1}{m}\#\left\{i\in\{1,2,\ldots,m\}:x_i\in B\right\}, \tag{14}$$

where $\#(A)$ denotes the power of the set A and $\mathcal{B}(\mathbf{R})$ represents the family of real Borel sets, is known as the empirical distribution of the sequence $\{X_i\}_{i=1}^{\infty}$. Let also $\underset{\sim}{\mathcal{P}} : \mathcal{B}(\mathbf{R}) \longrightarrow [0,1]$ be the distribution of a probability measure. The sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is called the empirically ergodic sequence with the limit $\underset{\sim}{\mathcal{P}}$, if the condition

$$\lim_{m \to \infty} \mathcal{P}_m(E) = \underset{\sim}{\mathcal{P}}(E) \tag{15}$$

is fulfilled with probability 1 (with respect to the measure P) for every set E of the form $(-\infty, e]$, where $\underset{\sim}{\mathcal{P}}(\{e\}) = 0$.

As results from The Glivenko-Cantelli Theorem [1], this condition is more general than the assumption frequently formulated in the theory of estimation concerning the identity of the distributions and the independence of the random variables X_i that represent the random sample. In the case when such an assumption is accepted, the measure $\underset{\sim}{\mathcal{P}}$ is nothing other than the distribution of the variables X_i, i.e.

$$\underset{\sim}{\mathcal{P}}(B) = P(x_i \in B) \tag{16}$$

for any $i = 1, 2, \ldots$ and $B \in \mathcal{B}(\mathbf{R})$.

Lemma 1. *Let the sequence of real random variables $\{X_i\}_{i=1}^{\infty}$, defined on a common probability space (Ω, Σ, P), be empirically ergodic with the limit $\underset{\sim}{\mathcal{P}}$, which has the distribution function F. If the kernel estimator of this function $\widehat{F}$ is given by formula* (13), *and conditions* (2)–(3) *with* (12) *are fulfilled, then for every $x^* \in \mathbf{R}$ such that*

$$\underset{\sim}{\mathcal{P}}(\{x^*\}) = 0, \tag{17}$$

the equality

$$\lim_{h \to 0} \lim_{m \to \infty} \widehat{F}(x^*) = F(x^*) \tag{18}$$

is true with probability 1 (with respect to the measure P).

Proof. From condition (14) it can be directly obtained that

$$\frac{1}{m} \sum_{i=1}^{m} \chi_B(x_i) = \int_{\mathbf{R}} \chi_B(x)\, d\mathcal{P}_m(x) \quad \text{for any } B \in \mathcal{B}(\mathbf{R}), \tag{19}$$

where χ_B denotes the characteristic function of the set B. Since linear and continuous operators equal on dense spaces are identical, for any measurable function $g : \mathbf{R} \longrightarrow \mathbf{R}$ the following is true:

$$\frac{1}{m} \sum_{i=1}^{m} g(x_i) = \int_{\mathbf{R}} g(x)\, d\mathcal{P}_m(x). \tag{20}$$

In particular, the formula

$$\widehat{F}(x^*) = \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_m(x) \tag{21}$$

can be obtained on the basis of definition (13). Therefore, due to the properties of weak convergence of the distribution functions, condition (15) implies that with probability 1

$$\lim_{m\to\infty} \widehat{F}(x^*) = \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x). \tag{22}$$

The consequences of formulas (3) and (12) are:

$$\lim_{x\to-\infty} I(x) = 0 \tag{23}$$

and

$$\lim_{x\to\infty} I(x) = 1, \tag{24}$$

which, thanks to condition (2), gives

$$\lim_{h\to 0} I\left(\frac{x^* - x}{h}\right) = \begin{cases} 1 & \text{for } x < x^*, \\ I(0) & \text{for } x = x^*, \\ 0 & \text{for } x > x^*. \end{cases} \tag{25}$$

In turn, the following equality is true:

$$\begin{aligned} \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x) &= \int_{(-\infty,x^*)} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x) \\ &+ I(0)\underset{\sim}{\mathcal{P}}(\{x^*\}) + \int_{(x^*,\infty)} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x). \end{aligned} \tag{26}$$

Therefore, it results from The Lebesgue Dominated Convergence Theorem that

$$\lim_{h\to 0} \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x) = \int_{(-\infty,x^*)} d\underset{\sim}{\mathcal{P}}(x) + I(0)\underset{\sim}{\mathcal{P}}(\{x^*\}), \tag{27}$$

i.e., taking into account assumption (17):

$$\lim_{h\to 0} \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\underset{\sim}{\mathcal{P}}(x) = \int_{(-\infty,x^*]} d\underset{\sim}{\mathcal{P}}(x). \tag{28}$$

Applying equality (22) to the above formula, one ultimately obtains the claim of Lemma 1. □

THEOREM 1. *Let the sequence of real random variables* $\{X_i\}_{i=1}^{\infty}$, *defined on a common probability space* (Ω, Σ, P), *be empirically ergodic with the limit* $\mathcal{P}_{\sim}$, *which has the distribution function* F. *If the kernel estimator of this function* $\widehat{F}$ *is given by formula* (13), *and conditions* (2)–(3) *with* (12), *as well as*

$$\lim_{m \to \infty} h = 0 \tag{29}$$

are fulfilled, then for every $x^* \in \mathbf{R}$ *such that*

$$\mathcal{P}_{\sim}(\{x^*\}) = 0, \tag{30}$$

the equality

$$\lim_{m \to \infty} \widehat{F}(x^*) = F(x^*) \tag{31}$$

is true with probability 1 (with respect to the measure P*), which gives the strong consistency, therefore also the consistency, of the kernel estimator of the distribution function at the points of its continuity.*

PROOF. It suffices to demonstrate that the convergence, when $m \longrightarrow \infty$, occurring in formula (18) is uniform with respect to the variable h.

Let F_m denote the distribution function of the measure $\mathcal{P}_m$. For an arbitrarily fixed $m \in \mathbf{N} \setminus \{0\}$, it is obvious that

$$\lim_{x \to \infty} I\left(\frac{x^* - x}{h}\right)(F_m - F)(x) = 0, \tag{32}$$

$$\lim_{x \to -\infty} I\left(\frac{x^* - x}{h}\right)(F_m - F)(x) = 0. \tag{33}$$

Applying to the Stielties integral $⨏$ the integration by parts procedure, one obtains

$$\begin{aligned} &\int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_m(x) - \int_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d\mathcal{P}_{\sim}(x) \\ &= ⨏_{\mathbf{R}} I\left(\frac{x^* - x}{h}\right) d(F_m - F)(x) \\ &= -⨏_{\mathbf{R}} (F_m - F)(x)\, dI\left(\frac{x^* - x}{h}\right). \end{aligned} \tag{34}$$

Since, regardless of the value of the variable h, the saltus of the function I equals 1 (is finite), while from The Glivenko-Cantelli Theorem it results that

$$\sup_{x\in\mathbf{R}} |(F_m - F)(x)| \xrightarrow{m\to\infty} 0, \tag{35}$$

formulas (21)–(22) and (34) finally prove Theorem 1. □

2.3. Kernel estimators of quantile. To carry on the considerations given in the previous subsection: if

$$K(x) > 0 \quad \text{for every } x \in \mathbf{R}, \tag{36}$$

then the kernel estimator of the quantile of order r, denoted hereinafter as $\widehat{q}$, may be uniquely defined by the equation

$$\widehat{F}(\widehat{q}) = r; \tag{37}$$

therefore, given formula (13), one finally obtains

$$\sum_{i=1}^{m} I\left(\frac{\widehat{q} - x_i}{h}\right) = mr. \tag{38}$$

As before, the (strong) consistency of the estimator defined above will be shown under very mild assumptions.

LEMMA 2. *Let the sequence of real random variables* $\{X_i\}_{i=1}^{\infty}$, *defined on a common probability space* (Ω, Σ, P), *be empirically ergodic with the limit* $\mathcal{P}$. *If the quantile of order* r *is given uniquely (with respect to the measure* $\mathcal{P}$*), while its kernel estimator is defined by formula* (38), *and conditions* (2)–(3) *with* (12) *are fulfilled, then the equality*

$$\lim_{h\to 0}\lim_{m\to\infty} \widehat{q} = q \tag{39}$$

is true with probability 1 (with respect to the measure P*).*

PROOF. In order to show formula (39), it is sufficient to prove that

$$\forall \varepsilon > 0 \ \exists h_* > 0 : \forall h < h_* \ \exists m_* \in \mathbf{N} \setminus \{0\} : \forall m > m_* \qquad |\widehat{q} - q| < \varepsilon. \tag{40}$$

Let some $\varepsilon > 0$ be fixed. Since the measure $\mathcal{P}$ is finite, the set of real numbers of positive measure can be at most countable. Thus there exist $\tilde{x}, \tilde{\tilde{x}} \in \mathbf{R}$ of zero measure $\mathcal{P}$ and fulfilling the inequalities

$$q - \varepsilon < \tilde{x} < q < \tilde{\tilde{x}} < q + \varepsilon. \tag{41}$$

The distribution function F is an increasing mapping; therefore, due to the assumed uniqueness of the quantile, it can be inferred that there exists $\delta > 0$ such that

$$F(\tilde{x}) + \delta < F(q) < F(\tilde{\tilde{x}}) - \delta, \tag{42}$$

where F denotes the distribution function of the measure $\underset{\sim}{P}$. Lemma 1 states that

$$\begin{aligned} &\forall \varepsilon > 0\ \exists h_* > 0 : \forall h < h_*\ \exists m_* \in \mathbf{N} \setminus \{0\} : \forall m > m_* \\ &\widehat{F}(\tilde{x}) < F(\tilde{x}) + \delta, \\ &\widehat{F}(\tilde{\tilde{x}}) < F(\tilde{\tilde{x}}) - \delta; \end{aligned} \tag{43}$$

therefore, by combining the latest two conditions one obtains

$$\begin{aligned} &\forall \varepsilon > 0\ \exists h_* > 0 : \forall h < h_*\ \exists m_* \in \mathbf{N} \setminus \{0\} : \forall m > m_* \\ &\widehat{F}(\tilde{x} < F(q) < F(\tilde{\tilde{x}}). \end{aligned} \tag{44}$$

Thus, due to the monotonicity of the function $\widehat{F}$ and to formula (41), inequality (44) implies that condition (40) is true, which concludes Lemma 2. □

THEOREM 2. *Let the sequence of real random variables* $\{X_i\}_{i=1}^{\infty}$, *defined on a common probability space* (Ω, Σ, P), *be empirically ergodic with the limit* $\underset{\sim}{P}$. *If the quantile of order* r *is defined uniquely (with respect to the measure* $\underset{\sim}{P}$*), while its kernel estimator is given by equation* (38), *and formulas* (2)–(3) *with* (12), *as well as the condition*

$$\lim_{m \to \infty} h = 0 \tag{45}$$

are fulfilled, then the equality

$$\lim_{m \to \infty} \widehat{q} = q \tag{46}$$

is true with probability 1 (with respect to the measure P*), which means that the kernel estimator of the quantile is strongly consistent, therefore, also consistent.*

PROOF. As results from the proof of Theorem 1, the value m_* introduced by formula (43) does not depend on the variable h. This implies that, in Lemma 2, the convergence when $m \longrightarrow \infty$ is uniform, which finally proves Theorem 2. □

Note that if the distribution of the measure P has a density function with a connected support, then its quantile is uniquely defined.

It should also be emphasized that condition (8), required in the estimation of the density function in order to assure the consistency property, proves to be superfluous in the cases of the distribution function and quantile.

3. An algorithm for numerical computations. Below, an algorithm for numerically calculating the value of the kernel estimator of the quantile will be presented, as an applicational corollary from the theoretical material presented in the previous section.

In practice, the form of the kernel K and the value of the smoothing parameter h can be chosen on the basis of the minimum mean squared error criterion (see [5,7]). It is then additionally assumed that $f \in C^2$, and that the functions f and f'' are bounded. In particular, the approximate value of the optimal smoothing parameter can be calculated by assuming the normal distribution; one then obtains

$$h = \left(V_K \frac{8}{3} \sqrt{\pi} \frac{1}{m}\right)^{1/5} \widehat{\sigma}, \tag{47}$$

while

$$V_K = \int_{-\infty}^{\infty} K(x)^2 \, dx \cdot \left(\int_{-\infty}^{\infty} x^2 K(x) \, dx\right)^{-2}, \tag{48}$$

$$\widehat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^{m} x_i^2 - \left(\frac{1}{m} \sum_{i=1}^{m} x_i\right)^2. \tag{49}$$

On the other hand, the choice of the type of the kernel K does not have a great impact on the statistical quality of estimation, and in practice it becomes possible to take into account primarily the desired properties of the estimator obtained, e.g. the simplicity of calculation, the finiteness of the support, etc.

In many applications, it proves to be particularly advantageous to introduce the concept of modification of the smoothing parameter. The estimator can then be constructed in the following manner:

(a) the kernel estimator $\widehat{f}$ is calculated in accordance with basic definition (1);

(b) the modifying parameters $s_i > 0$ $(i = 1, 2, \ldots, m)$ are stated in the form

$$s_i = \left(\frac{\widehat{f}(x_i)}{b}\right)^{-1/2}, \tag{50}$$

where b denotes the geometric mean of the numbers $\widehat{f}(x_1)$, $\widehat{f}(x_2)$, $\dots$,$\widehat{f}(x_m)$, given in the form of the logarithmic equation

(51) $$\log(b) = \frac{1}{m}\sum_{i=1}^{m}\log(\widehat{f}(x_i));$$

(c) the kernel estimators with the modified smoothing parameter are defined as follows:

– for the density function (the counterpart of formula (1)):

(52) $$\widehat{f}(x) = \frac{1}{mh}\sum_{i=1}^{m}\frac{1}{s_i}K\left(\frac{x - x_i}{hs_i}\right),$$

– for the distribution function (the counterpart of formula (13)):

(53) $$\widehat{F}(x) = \frac{1}{m}\sum_{i=1}^{m}I\left(\frac{x - x_i}{hs_i}\right),$$

– for the quantile (the counterpart of formula (38)):

(54) $$\sum_{i=1}^{m}I\left(\frac{\widehat{q} - x_i}{hs_i}\right) = mr.$$

The use of the modification procedure improves the quality of the estimation; however, from the practical point of view another essential feature consists in its slight sensitivity to the exactness of the choice of the constant h. In practice this property is exceptionally advantageous, and when such an estimator is applied, it mostly proves sufficient to accept the approximate value given by condition (47).

Since, thanks to assumption (5), the mapping I is a Lipschitz function with the constant

(55) $$L = K(0),$$

then the estimator of the quantile $\widehat{q}$, given by equation (54), may be calculated recurrently as the limit of the sequence $\{\widehat{q}^k\}_{k=0}^{\infty}$ defined by the formulas

(56) $$\widehat{q}^0 = \frac{1}{m}\sum_{i=1}^{m}x_i$$

(57) $$\widehat{q}^{k+1} = \widehat{q}^k + c\cdot\left(mr - \sum_{i=1}^{m}I\left(\frac{\widehat{q}^k - x_i}{hs_i}\right)\right) \quad \text{for } k = 1, 2, \dots,$$

whereas the global convergence of this algorithm is guaranteed by the condition

$$(58) \qquad 0 < c < \frac{2h}{L}\left(\sum_{i=1}^{m}\frac{1}{s_i}\right)^{-1};$$

furthermore, in the case when the function I is not linear in any restriction of the domain, then also

$$(59) \qquad c = \frac{2h}{L}\left(\sum_{i=1}^{m}\frac{1}{s_i}\right)^{-1}.$$

In practice, value (59) yields the best results in the application of algorithm (56)–(57).

For the purposes of the method elaborated here, the kernel

$$(60) \qquad K(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

can be proposed. It fulfills all the requirements formulated above, and in particular its primitive has a form convenient for calculations, namely:

$$(61) \qquad I(x) = \frac{1}{1+e^{-x}}.$$

In this case, the constant (48) amounts to

$$(62) \qquad V_K = \frac{3}{2\pi^4}.$$

Finally, algorithm (56)–(57) together with

(a) equality (61) defining the function I,
(b) formulas (47), along with (62) and (49), giving the value of the constant h,
(c) modification procedure (50)–(51), along with (1) and (60), the result of which are the parameters s_i,
(d) equalities (59), along with (55) and (60) for coefficient c,

provide a complete set of rules defining the practical procedure used to numerically calculate the quantile estimator of order r on the basis of the m-element random sample $x_1, x_2, \ldots, x_m$.

References

1. Billingsley P., *Probability and Measure*, Wiley, New York, 1979.
2. Devroye L., Gyorfi L., *Nonparametric Density Estimation, The L_1 View*, Wiley, New York, 1985.
3. Fisz M., *Probability Theory and Mathematical Statistics*, Wiley, New York, 1963.
4. Parrish R. S., *Comparison of quantile estimators in normal sampling*, Biometrics **46** (1990), 247–257.
5. Prakasa Rao B. L. S., *Nonparametric Functional Estimation*, Academic Press, Orlando, 1983.
6. Sheather S. J., Marron J. S., *Kernel quantile estimators*, J. Amer. Statist. Assoc. **85** (1990), 410–416.
7. Silverman B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

Received April 22, 1997

Cracow University of Technology
Faculty of Electrical and Computer Engineering
Warszawska 24
PL-31-155 Kraków

Institute of Mathematics
Jagiellonian University
Reymonta 4
PL-30-059 Kraków