საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, 175, №2, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, 175, №2, 2007

Mathematics

On Deheuvels' Nonparametric Estimation of Distribution Density

Elizbar Nadaraya*, Petre Babilua**

* *Academy Member, I. Javakhishvili Tbilisi State University*
** *I. Javakhishvili Tbilisi State University*

ABSTRACT. We establish the limiting distribution of an integral quadratic deviation of Deheuvels' nonparametric estimation of a multidimensional distribution density. © *2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *nonparametric estimation, recursive estimators, convergence in distribution.*

1. Let $X_1, X_2, \dots, X_n$ be a sequence of independent, equally distributed random values in a p-dimensional Euclidean space R_p, $p \ge 1$, that have the distribution density $f(x)$, $x = (x_1, x_2, \dots, x_p)$. As is known, Rosenblatt [1] and Parzen [2] gave the following definition of the nonparametric kernel estimation of the distribution density $f(x)$:

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

where $K(x)$ is a given kernel, while $\{h_n\}$ is a sequence of positive integers monotonically converging to zero.

Deheuvels [3] proposed a recurrent estimation of the density $f(x)$ of the following form:

$$f_n(x) = \frac{1}{H_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_i}\right) = \frac{H_{n-1}}{H_n} f_{n-1}(x) + \frac{1}{H_n} K\left(\frac{x - X_n}{h_n}\right),$$

$$H_n = \sum_{j=1}^{n} h_j .$$

The recurrent definition of estimations $f_n(x)$ of the probability density $f(x)$ has two advantages: a) there is no need to memorize data, i.e. if the estimation $f_{n-1}(x)$ is known, then to calculate $f_n(x)$ we use only the last observation $f(x)$; b) Deheuvels [3] shows that the asymptotic dispersion of the estimation $f_n(x)$ is less than that of estimations $f_n^{(RP)}(x)$. He also obtained asymptotically optimal values of h_j for the criterion of an integral mean-square error.

The aim of this paper is to obtain a limit distribution of the quadratic deviation $\int (f_n(x) - f(x))^2 r(x) dx$, where

$r(x)$ is a weight function which we assume to be piecewise-continuous and bounded. To obtain our results we use the functional limit theorem for a sequence of semimartingales [4], which was previously also used in [5] to solve an analogous problem for the estimation $f_n^{(RP)}(x)$.

2. Let us formulate the assumptions with regard to $f(x)$ and $K(x)$ to be used below in investigating the problem posed.

1°. The kernel

$$K(x)=\prod_{j=1}^{p} K_j(x), \quad x=(x_1,\dots,x_p),$$

and each of the kernels $K_j(u)$, $u\in R_1$, possesses the following properties:

$$0\le K_j(u)\le c<\infty, \quad K_j(-u)=K_j(u), \quad u^2K_j(u)\in L_1(R_1),$$
$$\int K_j(u)\,du=1, \quad K_j^0(ux)\ge K_j^0(x) \text{ for all } u\in[0,1] \text{ and all } x\in R_1,$$

where $K_j^0=K_j*K_j$, * being a convolution operator.

2°. The distribution density $f(x)$ is bounded and has bounded partial derivatives up to the second order.

Notation.

$$L_n=nh_n^p\int (f_n(x)-Ef_n(x))^2 r(x)dx,$$
$$U_n=nh_n^p\int (f_n(x)-f(x))^2 r(x)dx,$$
$$\alpha_i(x,y)=K\left(\frac{x-X_i}{h_i}\right)-EK\left(\frac{x-X_i}{h_i}\right),$$
$$b_n^2=4n^2h_n^{2p}H_n^{-4}\sum_{i=2}^{n}\sum_{j=1}^{i-1}E\left(\int \alpha_i(x,X_i)\alpha_j(x,X_j)r(x)dx\right)^2,$$
$$d_n^2=2n^2h_n^{2p}H_n^{-4}\iint f^2(x)\left(\sum_{j=1}^{n}h_j^pK_0\left(\frac{x-y}{h_j}\right)\right)^2 r(x)r(y)\,dx\,dy,$$
$$\Theta_n^{(1)}=nh_n^p\int (Ef_n(x)-f(x))^2 r(x)dx,$$
$$\Theta_n^{(2)}=nh_n^pH_n^{-2}\sum_{k=1}^{n}\iint K^2\left(\frac{x-u}{h_k}\right)f(u)r(x)\,du\,dx,$$
$$K_0=K*K, \quad \gamma_s(n)=\frac{1}{n}\sum_{i=1}^{n}h_i^{ps}, \quad s=1,2,\dots .$$

Lemma 1. *Let condition* 1^0 *be fulfilled,* $f(x)$ *and* $r(x)\ge 0$ *be piecewise-continuous and bounded functions. If* $h_n\to 0$, $nh_n^p\to\infty$ *and* $h_n^{ps}/\gamma_s(n)\to\gamma_s$, $s=1,2$, $0<\gamma_2\le\gamma_1\le 1$ *as* $n\to\infty$, *then*

$$b_n^2=d_n^2+\gamma_1(n)o(1)+O(1)+O\left(\frac{\gamma_1(n)}{n}\right)$$

and also

$$2\gamma_1^2\int f^2(x)r^2(x)dx\int K_0^2(u)du\le$$

$$\le\underline{\lim}_{n\to\infty}h_n^{-p}d_n\le\overline{\lim}_{n\to\infty}h_n^{-p}d_n\le 2\frac{\gamma_1^3}{\gamma_2}\int f^2(x)r^2(x)dx\int K_0^2(u)du.$$

Theorem 1. *Let the conditions of Lemma 1 be fulfilled. Then*

$$b_n^{-1}(L_n-EL_n)\xrightarrow{d}N(0,1),$$

where d is the convergence in distribution, and $N(0,1)$ *is a random value having a normal distribution with zero mean value and dispersion 1.*

Corollary. $d_n^{-1}(L_n-EL_n)\xrightarrow{d}N(0,1)$ *as* $n\to\infty$.

Lemma 2. *Let* $K(x)$ *and* $f(x)$ *satisfy conditions* 1^0 *and* 2^0*, respectively;* $r(x)\ge 0$ *be a bounded and piecewise-continuous function. If*

$$n^{-1}h_n^{-p}\left(\sum_{j=1}^{n}h_j^{p+2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

$$h_n^{-p/2}\left(U_n-L_n-\Theta_n^{(1)}\right)=o_p(1).$$

Theorem 2. *a) Let* $K(x)$, $f(x)$ *and* $r(x)$ *satisfy the conditions of Lemma 2. If*

$$nh_n^p\to\infty,\qquad \frac{h_n^{ps}}{\gamma_s(n)}\to\gamma_s,\quad s=1,2,\qquad n^{-1}h_n^{-p}\left(\sum_{j=1}^{n}h_j^{p+2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

$$h_n^{-p/2}\sigma_n^{-1}\left(U_n-\Theta_n^{(1)}-\Theta_n^{(2)}\right)\xrightarrow{d}N(0,1),$$

where $\sigma_n^2=h_n^{-p}d_n^2$.

b) Let the following condition be added to the conditions with regard to $f(x)$ *and all second order partial derivatives of the function* $f(x)$ *belong to* $L_2(R_p)$. *If*

$$nh_n^p\to\infty,\qquad \frac{h_n^{ps}}{\gamma_s(n)}\to\gamma_s,\quad s=1,2,\qquad n^{-1}h_n^{-\frac{3p}{2}}\left(\sum_{j=1}^{n}h_j^{p+2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

$$h_n^{-p/2}\sigma_n^{-1}\left(U_n-\Theta_n^{(2)}\right)\xrightarrow{d}N(0,1).$$

c) Let $K(x)$, $f(x)$ *and* $r(x)$ *satisfy the conditions of Lemma 2 and, in addition to this, the second order partial derivatives of the function* $f(x)$ *belong to* $L_1(R_p)$. *If*

$$nh_n^p\to\infty,\qquad \frac{h_n^{sp}}{\gamma_s(n)}\to\gamma_s,\quad s=1,2,$$

also

$$\frac{h_n^p}{\gamma_1(n)} = \gamma_1 + o\left(h_n^{p/2}\right) \quad and \quad n^{-1} h_n^{-\frac{3p}{2}} \left(\sum_{j=1}^{n} h_j^{p+2} \right)^s \to 0, \ \ s = 1,2 \quad as \ \ n \to \infty, \ \ 0 < p \le 3,$$

then

$$h_n^{-p/2} \sigma_n^{-1} \left(U_n - \Theta \right) \xrightarrow{d} N(0,1),$$
$$\Theta = \gamma_1 \int f(x) r(x) dx \int K^2(u) du.$$

Corollary. *If* $h_k = h_n$, $k = 1, \ldots, n$, *then from Theorem 2 we obtain the well-known result on the limit distribution of an integral quadratic deviation of the estimation* $f_n^{(RP)}(x)$ [6-8].

The conditions c) are fulfilled, for instance, if we assume $h_k = k^{-\alpha}$ for $\frac{2}{p+8} < \alpha < \frac{1}{p+2}$, $\frac{1}{p+2} < \alpha < \frac{2}{3p}$, $p < 4$, noting that $\gamma_1 = 1 - \alpha p > 0$.

Remark. Theorems 1 and 2 define more precisely the results obtained in [9]. A more precise definition of these results consists in the following: in the present paper there is no integrability of the weight function $r(x)$; the conditions on h_k, $k = 1, 2, \ldots$, used in [10] are replaced by less restrictive ones.

მათემატიკა

განაწილების სიმკვრივის დეველსის არაპარამეტრული შეფასების შესახებ

ე. ნადარაია*, პ. ბაბილუა**

* *აკადემიის წევრი, ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*
** *ი. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი*

ნაშრომში დადგენილია მრავალგანზომილებიანი სიმკვრივის დეველსის არაპარამეტრული შეფასების ინტეგრალური კვადრატული გადახრის ზღვარითი განაწილება.

REFERENCES

1. *M. Rosenblatt*. (1956), Ann. Stat., **27**, 3: 832-837.
2. *E. Parzen*. (1962), Ann. Stat., **33,** 3: 1065-1076.
3. *P. Deheuvels*. (1973), Comptes Rendus de l'Académie des Sciences de Paris, 276, 1013-1015.
4. *R. Sh. Liptser, A. N. Shiryaev.* (1980), Teoriya veroyatnosti i ee primenenie, **XXV**, 4, 683-703 (Russian).
5. *E. A. Nadaraya*. (1984), Soobshch. AN GSSR, **113,** 2: 253-257 (Russian).
6. *P. Hall.* (1984), Journal of Multivariate Analysis, **14**: 1-16.
7. *M. Rosenblatt*. (1975), Ann. Math. Stat., **3**: 1-14.
8. *E. A. Nadaraya*. (1989), Nonparametric estimation of probability densities and regression curves. Translated from the Russian. Dordrecht.
9. *I. Kikvadze*. (2005), Bull. Georg. Acad. Sci., **172**, 3: 378-380.

Received April, 2007