საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, 175, №1, 2007
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, 175, №1, 2007

Mathematics

On the Wolverton-Wagner Estimate of a Distribution Density

Elizbar Nadaraya*, Petre Babilua**

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

ABSTRACT. The result of the work consists mainly in obtaining the limit distribution of an integral quadratic deviation of the Wolverton-Wagner nonparametric estimate of a multidimensional distribution density. *© 2007 Bull. Georg. Natl. Acad. Sci.*

Key words: *nonparametric estimate, recursive estimate, convergence in distribution.*

1. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, equally distributed random variables with values in a Euclidean p-dimensional space R_p, $p \ge 1$, whose distribution density is $f(x)$, $x = (x_1, \ldots, x_p)$. As is known, Rosenblatt [1] and Parzen [2] gave the following definition of an empirical kernel density $f_n^{(RP)}(x)$, which is based on the sampling $X_1, X_2, \ldots, X_n$:

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

where $K(x)$ is a given kernel and $\{a_n\}$ is a sequence of positive integers converging monotonically to infinity.

Wolverton and Wagner [3] introduce the following definition of an empirical density $f_n^{(W)}(x)$ that differs little from $f_n^{(RP)}(x)$ but is recurrent:

$$f_n^{(W)}(x) = \frac{1}{n} \sum_{j=1}^{n} a_j^p K(a_i(x - X_i)) = \frac{n-1}{n} f_{n-1}^{(W)}(x) + \frac{a_n^p}{n} K(a_n(x - X_n)).$$

The recurrent definition of probability density estimates $f_n^{(W)}(x)$ has two obvious advantages: 1) there is no need to memorize data, i.e. if the estimate $f_{n-1}^{(W)}(x)$ is known, then $f_n^{(W)}(x)$ can be calculated by means of the last observation X_n only, without using the sampling $X_1, X_2, \ldots, X_{n-1}$; 2) the asymptotic dispersion of the estimate $f_n^{(W)}(x)$ does not exceed the dispersion of the estimate $f_n^{(RP)}(x)$.

The aim of this work is to study the asymptotics of a mean value of the integral of the squared error

© 2007 Bull. Georg. Natl. Acad. Sci.

$$E\int\left[f_n^{(W)}(x)-f(x)\right]^2 r(x)dx$$

and the limit distribution of the functional

$$\int\left[f_n^{(W)}(x)-f(x)\right]^2 r(x)dx\,.$$

Theorem 1. *Let the kernel*

$$K(x)=\prod_{j=1}^{p} K_j(x_j)$$

and each of the kernels $K_j(x)$ *possess the following properties:*

$$0\le K_j(x)\le c<\infty,\quad K_j(-u)=K_j(u),\quad u^2K_j(u)\in L_1(-\infty,\infty)$$

and

$$\int K_j(u)du=1\,.$$

Assume that the density $f(x)$ *and its partial derivatives up to second order are bounded and belong to* $L_2(R_p)$. *If*

$$\frac{a_n^p}{n}\to 0,\quad \frac{\gamma_n}{a_n^p}\to\gamma>0\quad and\quad \sum_{k=1}^{\infty}a_k^{-2}=\infty\,,$$

then we have

$$E\int\left[f_n^{(W)}(x)-f(x)\right]^2 r(x)dx=\frac{\gamma_n}{n}\int f(x)r(x)dx\int K^2(x)du+\left(\frac{1}{n}\sum_{k=1}^{n}a_k^{-2}\right)\frac{1}{4}\int\left[\sum_{j=1}^{p}\alpha_j\frac{\partial^{(2)}f(x)}{\partial x_j^2}\right]^2 r(x)dx+$$

$$+o\left(\frac{\gamma_n}{n}+\left(\frac{1}{n}\sum_{k=1}^{n}a_k^{-2}\right)^2\right),\quad \gamma_n=\frac{1}{n}\sum_{k=1}^{n}a_k^p,\quad \alpha_j=\int x_j^2K_j(x_j)dx_j\,,$$

where $r(x)$ *is a bounded and piecewise-continuous function.*

2. Let us formulate the assumptions with regard to $K(x)$ and $f(x)$, $x=(x_1,\dots,x_p)$, used in studying the limit distribution of an integral standard deviation $f_n^{(W)}(x)$ from $f(x)$.

1⁰. The kernel $K(x)$ satisfies the conditions of Theorem 1 and, moreover, $K_j^0(ux)\ge K_j^0(x)$ for all $u\in[0,1]$ and for all $x\in R_1$, where $K_j^0=K_j*K_j$, * being the convolution operator.

2⁰. The density $f(x)$ is bounded and has bounded partial derivatives up to second order.

Notation.

$$L_n=n\int\left(f_n^{(W)}(x)-Ef_n^{(W)}(x)\right)^2 r(x)dx,$$

$$U_n=\frac{n}{a_n^p}\int\left(f_n^{(W)}(x)-f(x)\right)^2 r(x)dx,$$

$$\alpha_i(x,y)=a_i^p\left[K(a_i(x-y))-EK(a_i(x-X_1))\right],$$

$$b_n^2=\frac{4}{n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}E\left(\int\alpha_i(x,X_i)\alpha_j(x,X_j)r(x)dx\right)^2,$$

$$d_n^2=\frac{2}{n^2}\iint f^2(x)\left(\sum_{i=1}^n a_i^p K_0(a_i(x-y))\right)^2 r(x)r(y)\,dx\,dy,$$

$$K_0=K*K,\quad \gamma_s(n)=\frac{1}{n}\sum_{i=1}^n a_i^{ps},\quad s=1,2,\ldots,$$

$$\Theta_n^{(1)}=na_n^{-p}\int\left(Ef_n^{(W)}(x)-f(x)\right)^2 r(x)dx,\quad \widetilde{L}_n=a^{-p}L_n,$$

$$\Theta_n^{(2)}=n^{-1}a_n^{-p}\sum_{k=1}^n a_k^{2p}\iint K^2(a_k(x-u))f(u)r(x)\,du\,dx.$$

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

$$2\gamma_1^2\le\int f^2(x)r^2(x)dx\int K_0^2(u)du\le\varliminf_{n\to\infty}\frac{d_n^2}{a_n^p}\le\varlimsup_{n\to\infty}\frac{d_n^2}{a_n^p}\le 2\gamma_1\int f^2(x)r^2(x)dx\int K_0^2(u)du.$$

Theorem 2. *Let all conditions of Lemma* 1 *be fulfilled. Then*

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

where d denotes convergence in distribution, $N(0,1)$ *is a random value having normal distribution with mean 0 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 bounded and piecewise-continuous. If*

$$n^{-1}a_n^{-p}\left(\sum_{k=1}^n a_k^{-2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

$$a_n^{p/2}\left(U_n-\widetilde{L}_n-\Theta_n^{(1)}\right)=o_p(1).$$

Lemma 3. $E\widetilde{L}_n=\Theta_n^{(2)}+O\left(a_n^{-p}\right)$.

Theorem 3. a) *Let* $K(x)$, $f(x)$ *and* $r(x)$ *satisfy the conditions of Lemma* 2. *If*

$$a_n\uparrow\infty,\quad \frac{a_n^p}{n}\to 0,\quad \frac{\gamma_s(n)}{a_n^{sp}}\to\gamma_s,\quad s=1,2,$$

$$n^{-1}a_n^{-p}\left(\sum_{k=1}^n a_k^2\right)^2\to 0\quad as\quad n\to\infty,$$

then

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

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

b) To the conditions with regard to $f(x)$ *we add the assumption: all partial derivatives of second order of the function* $f(x)$ *belong to* $L_2(R_p)$. *If*

$$a_n\uparrow\infty,\quad \frac{a_n^p}{n}\to 0,\quad \frac{\gamma_s(n)}{a_n^{sp}}\to\gamma_s,\quad s=1,2,$$

$$\frac{1}{n\sqrt{a_n^p}}\left(\sum_{j=1}^n a_j^{-2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

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

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

$$\frac{a_n^p}{n}\to 0,\quad \frac{\gamma_s(n)}{a_n^{sp}}\to\gamma_s,\quad s=1,2\ ,$$

and also

$$\frac{\gamma_1(n)}{a_n^p}=\gamma_1+o\left(a_n^{-p/2}\right),\quad \left(na_n^{p/2}\right)^{-1}\sum_{j=1}^n a_j^{p-2}\to 0$$

and

$$\left(na_n^{p/2}\right)^{-1}\left(\sum_{j=1}^n a_j^{-2}\right)^2\to 0\quad as\quad n\to\infty,$$

then

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

Corollary. *If* $a_k=a_n$, $k=1,2,\ldots$, *then from Theorem* 3 *we obtain the well-known result on the limit distribution of a standard deviation of the estimate* $f_n^{(RP)}(x)$ ([4, 5]).

Remark. Theorems 2 and 3 generalize and refine the results obtained by one of the authors in [6]. The refinement of these results is done by a method differing from the method used in [6] and consists in the following: in this paper there is no integrability of the weight function $r(x)$; the conditions on a_k, $k=1,2,\ldots$ used in [6] are replaced by less restrictive requirements.

მათემატიკა

განაწილების სიმკვრივის ვოლვერტონ-ვაგნერის არაპარამეტრული შეფასების შესახებ

ე. ნადარაია*, პ. ბაბილუა**

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

ნაშრომში დადგენილია მრავალგანზომილებიანი განაწილების სიმკვრივის ვოლვერტონ-ვაგნერის არაპარამეტრული შეფასების ინტეგრალური კვადრატული გადახრის ზღვარითი განაწილების სახე.

REFERENCES

1. *M. Rosenblatt*. Ann. Stat., **27**, 3, 832-835, 1956.
2. *E. Parzen*. Ann. Stat., **3**, 33, 1065-1076, 1962.
3. *C. Wolverton, T. Wagner*. IEEE Trans. Systems Sci. Cybernet. **5**, 307, 1967.
4. *E. Nadaraya*. Nonparametric estimation of probability densities and regression curves. Netherlands, Kluwer Acad. Publishers, 1989.
5. *M. Rosenblatt*. Ann. Stat., **3**, 1, 1-14, 1975.
6. *E. Nadaraya*. Teoriya. Veroyatnosti i ee Primeneniya, 37, 413-421, 1992.

Received January, 2007