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


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

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

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

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 $f_n^{(KP)}(x)$ but is recurrent:

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

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 $f_n^{(W)}(x)$ is known, then $f_n^{(W)}(x)$ can be calculated by means of the last observation $X_n$ only, without using the sampling $X_1, X_2, ..., X_{n-1}$; 2) the asymptotic dispersion of the estimate $f_n^{(W)}(x)$ does not exceed the dispersion of the estimate $f_n^{(KP)}(x)$.

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

\[ E \int \left[ f^{(w)}_n(x) - f(x) \right]^2 r(x) \, dx \]

and the limit distribution of the functional
\[ \int \left[ f^{(w)}_n(x) - f(x) \right]^2 r(x) \, dx. \]

**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:

\[ 0 \leq K_j(x) \leq c < \infty, \quad K_j(-u) = K_j(u), \quad u^2 K_j(u) \in L_1(-\infty, \infty) \]

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 \( L_2(\mathbb{R}_p) \). If

\[ \frac{\alpha_n^p}{n} \to 0, \quad \frac{\gamma_n}{\alpha_n^p} \to \gamma > 0 \quad \text{and} \quad \sum_{k=1}^{\infty} a_k^2 = \infty, \]

then we have

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

\[ + \left( \frac{\gamma_n}{n} + \left( \frac{1}{n} \sum_{k=1}^{n} a_k^2 \right)^2 \right)^2 \]

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

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

1. The kernel \( K(x) \) satisfies the conditions of Theorem 1 and, moreover, \( K_j^0(u) \geq K_j^0(x) \) for all \( u \in [0, 1] \) and for all \( x \in \mathbb{R}_1 \), where \( K_j^0 = K_j \ast K_j \), \( \ast \) being the convolution operator.

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

Notation.

\[ L_n = n \int \left[ f^{(w)}_n(x) - Ef^{(w)}_n(x) \right]^2 r(x) \, dx, \]

\[ U_n = n \int \left[ f^{(w)}_n(x) - f(x) \right]^2 r(x) \, dx, \]

\[ \alpha_j(x, y) = a_j^n \left[ K(\alpha_j(x - y)) - EK(\alpha_j(x - X_1)) \right], \]

\[ b_i^n = \frac{4}{n^2} \sum_{i=2}^{n} \sum_{j=1}^{i-1} E \left( \int \alpha_j(x, X_i) \alpha_j(x, X_j) r(x) \, dx \right)^2. \]
\[
d^2_n = \frac{2}{n^2} \int \int f^2(x) \left( \sum_{i=1}^{n} a_i^p K_0(a_i(x-y)) \right)^2 r(x) r(y) \, dx \, dy,
\]

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

\[
\Theta^{(1)}_n = n a_n^{-p} \int \left( \frac{Ef_n(y)}{n} - f(x) \right)^2 r(x) \, dx, \quad \overline{L}_n = a_n^{-p} L_n,
\]

\[
\Theta^{(2)}_n = n^{-1} a_n^{-p} \sum_{k=1}^{n} a_k^{p_2} \int \int K^2(a_k(x-u)) f(u) r(x) \, du \, dx.
\]

**Lemma 1.** Let conditions $1^0$ be fulfilled, $f(x)$ and $r(x) \geq 0$ be piecewise-continuous and bounded functions. If

\[
a_n \to \infty, \quad \frac{a_n^p}{n} \to 0 \quad \text{and} \quad \frac{\gamma_s(n)}{a_n^{p_s}} \to \gamma_s, \quad s = 1, 2, \quad 0 < \gamma_2 \leq \gamma_1 \leq 1 \quad \text{as} \quad n \to \infty,
\]

then

\[
b_n^2 = a_n^2 + \gamma_1(n) o(1) + O(1) + O\left(\frac{\gamma_1(n)}{n}\right)
\]

and also

\[
2 \gamma_1^2 \leq \int f^2(x) r^2(x) \, dx \int K_0^2 (u) \, du \leq \lim_{n \to \infty} \frac{d_n^2}{a_n^p} \leq \lim_{n \to \infty} \frac{d_n^2}{a_n^p} \leq 2 \gamma_1 \int f^2(x) r^2(x) \, dx \int K_0^2 (u) \, du.
\]

**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, $r(x) \geq 0$ be bounded and piecewise-continuous. If

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

then

\[
a_n^{p/2} (U_n - \overline{L}_n - \Theta^{(1)}_n) = o_p(1).
\]

**Lemma 3.** $EL_n = \Theta^{(0)}_n + O(a_n^{-p})$.

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

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

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

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

\[
\begin{align*}
& a_n \uparrow \infty, \quad \frac{a_n^p}{n} \to 0, \quad \gamma_s(n) \to \gamma_s, \quad s = 1,2, \\
& \frac{1}{n\sqrt{a_n}} \left( \sum_{j=1}^{n} a_j^{-p} \right)^{2} \to 0 \quad \text{as} \quad n \to \infty,
\end{align*}
\]

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

\[
\begin{align*}
& \frac{a_n^p}{n} \to 0, \quad \frac{\gamma_s(n)}{a_n^{sp}} \to \gamma_s, \quad s = 1,2, \\
& \gamma_1(n) \to \gamma_1 + o(a_n^{-p/2}), \quad \left( na_n^{p/2} \right)^{-1} \sum_{j=1}^{n} a_j^{-p} \to 0
\end{align*}
\]

and also

\[ \gamma_1(n) \to \gamma_1 + o(a_n^{-p/2}), \quad \left( na_n^{p/2} \right)^{-1} \sum_{j=1}^{n} a_j^{-p} \to 0 \quad \text{as} \quad n \to \infty, \]

then

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

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.
ხანგრძლივობის სიმკვრივის ვოლ्वერ-ვაგერის
არაპარამეტრული განაწილების შეფასება

ე. ნადარაია*, ბ. ბაბიუა**

* ა. ჯოხოსიას ბრუნი, ა. ჯოხოსიას ნინოწმინდის სახელობის უნივერსიტეტი
** ა. ჯოხოსიას ნინოწმინდის სახელობის უნივერსიტეტი

ნაწილოვნიდან დაგრძელების პროცესის ახასიათებელი გზების დამტკიცების მეთოდებიდან ვოლ्वერ-ვაგერის
არაპარამეტრული განაწილების შეფასების შემდგომში უზრუნველყოფილი განაწილების ხელშეწყობის სახე.

REFERENCES


Received January, 2007