

ON A STATISTIC FOR TESTING THE HYPOTHESIS OF THE EQUALITY OF SEVERAL DISTRIBUTION DENSITIES

R. ABSAVA AND E. NADARAYA

ABSTRACT. A statistic is considered, which describes a mutual distinction of several kernel type estimates of the distribution density. The limit distribution of the considered statistic is found.

რეზიუმე. ნაშრომში განხილულია სტატისტიკა, რომელიც აღწერს განსხვავების სიძვერის რამდენიმე გულოვანი შეფასების ურთიერთგადახრას. მოძებნილია ამ სტატისტიკის ზღვარიოი განაწილების კანონი.

0. Let $X^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$, $i = \overline{1, p}$, be independent samples of sizes $n_1, n_2, \dots, n_p$ from p (≥ 2) populations with distribution densities $f_1(x), \dots, f_p(x)$, respectively. Based on samples $X^{(i)}$, $i = \overline{1, p}$, it is required to test the hypothesis

$$H_0 : f_1(x) = \dots = f_p(x) = f_0(x)$$

against the sequence of “close” alternatives ([1], [2])

$$H_1 : f_i(x) = f_0(x) + \alpha_{n_0} \sum_{k=1}^m \varphi_{ik} \left(\frac{x - l_k}{\gamma_{n_0}} \right) + o(\alpha_{n_0} \cdot \gamma_{n_0}) \quad (\alpha_{n_0}, \gamma_{n_0} \rightarrow 0), \quad (1)$$

$$\int \varphi_{ik}(x) dx = 0, \quad n_0 = \min(n_1, \dots, n_p) \rightarrow \infty.$$

Alternatives of this kind mean that densities $f_i(x)$, $i = \overline{1, p}$, have m (≥ 1) sharp peaks at the fixed points l_j , $j = \overline{1, m}$.

We propose to test the hypothesis H_0 using the statistic

$$T(n_1, n_2, \dots, n_p) = \sum_{i=1}^p N_i \int \left[\widehat{f}_i(x) - \frac{1}{N} \sum_{j=1}^p N_j \widehat{f}_j(x) \right]^2 r(x) dx, \quad (2)$$

2000 *Mathematics Subject Classification.* 62G05.

Key words and phrases. Testing homogeneity, distribution density, kernel estimate of a density.

where $\widehat{f}_i(x)$ are Rosenblatt-Parzen kernel estimates of the distribution density $f_i(x)$:

$$\widehat{f}_i(x) = \frac{n_i}{a_i} \sum_{j=1}^{n_i} K(a_i(x - X_j^{(i)})),$$

$$N_i = \frac{n_i}{a_i}, \quad N = N_1 + \dots + N_p.$$

The particular case $p = 2$ was considered in [3] and [4]. In that case the statistic T takes a more convenient form

$$T(n_1, n_2) = \frac{N_1 N_2}{N_1 + N_2} \int (\widehat{f}_1(x) - \widehat{f}_2(x))^2 r(x) dx.$$

1. In this paper, a limit distribution of statistic (2) is found under the hypothesis H_1 when n_i increases infinitely so that $n_i = nk_i$, where $n \rightarrow \infty$ and k_i are constants $i = \overline{1, p}$. Let $a_1 = a_2 = \dots = a_p = a_n$, where $a_n \rightarrow \infty$ as $n \rightarrow \infty$.

To obtain the limit law on the distribution of the functional $T_n = T(n_1, \dots, n_p)$ we introduce the following conditions for the functions $K(x)$, $f_0(x)$, $\varphi_{ik}(x)$, $k = \overline{1, m}$, $i = \overline{1, p}$, and $r(x)$:

1. $K(x)$ is the function with bounded variation,

$$\int K(x) dx = 1, \quad x K(x) \in L_1(-\infty, \infty).$$

2. The functions $f_0(x)$ and $\varphi_{ik}(x)$, $i = \overline{1, p}$, $k = \overline{1, m}$, are bounded and have a bounded derivative of first order; moreover, $\varphi_{ik}(x)$ and $\varphi_{ik}^{(1)}(x) \in L_1(-\infty, \infty)$ and $\lim_{|x| \rightarrow \infty} \varphi_{ik}(x) = 0$.
3. The weight function $r(x)$ is piecewise-continuous, bounded and integrable; moreover, the above-mentioned l_k are points of continuity of $r(x)$ and $r(l_k) \neq 0$, $k = \overline{1, m}$.

The following statement is true.

Theorem 1. *Let conditions 1, 2 and 3 be fulfilled. If $n a_n^{-1/2} \alpha_n^2 \gamma_n \rightarrow c_0 \neq 0$, $a_n \gamma_n \rightarrow \infty$, $\alpha_n \cdot \gamma_n = o(n^{-1/2})$ and $a_n^{-1} \cdot \sqrt{n} \geq c \neq 0$ as $n \rightarrow \infty$, then under the hypothesis H_1 the limiting distribution of the random variable*

$$a_n^{1/2} (T_n - \mu)$$

is normal $(A(\varphi), \sigma^2)$, where

$$A(\varphi) = c_0 \sum_{i=1}^p k_i \sum_{k=1}^m r(l_k) \int \left[\varphi_{ik}(x) - \frac{1}{k} \sum_{j=1}^p k_j \varphi_{jk}(x) \right]^2 dx,$$

$$\sigma^2 = 2(p-1) \int f_0^2(x) r^2(x) dx \cdot R(K_0), \quad K_0 = K * K,$$

$$\mu = (p-1) \int f_0(x) r(x) dx \cdot R(K), \quad R(g) = \int g^2(x) dx,$$

$$\bar{k} = k_1 + \dots + k_p.$$

Proof. Let us represent T_n as a sum

$$T_n = T_n^{(1)} + A_{1n} + A_{2n},$$

where

$$T_n^{(1)} = \frac{n}{a_n} \sum_{i=1}^p k_i \int \left[\hat{f}_i(x) - E\hat{f}_i(x) - \frac{1}{\bar{k}} \sum_{j=1}^p k_j (\hat{f}_j(x) - E\hat{f}_j(x)) \right]^2 r(x) dx,$$

$$A_{1n} = 2 \frac{n}{a_n} \sum_{i=1}^p k_i \int [\hat{f}_i(x) - E\hat{f}_i(x)] \left[E\hat{f}_i(x) - \frac{1}{\bar{k}} \sum_{j=1}^p k_j E\hat{f}_j(x) \right] r(x) dx,$$

$$A_{2n} = \frac{n}{a_n} \sum_{i=1}^p k_i \int \left[E\hat{f}_i(x) - \frac{1}{\bar{k}} \sum_{j=1}^p k_j E\hat{f}_j(x) \right]^2 r(x) dx.$$

Here and in what follows $E(\cdot)$ is a mathematical expectation under the hypothesis H_1 . It is not difficult to see that

$$E\hat{f}_i(x) = \bar{f}_0(x) + \alpha_n \sum_{k=1}^m \varphi_{ik} \left(\frac{x - l_k}{\gamma_k} \right) -$$

$$- \frac{\alpha_n}{a_n \gamma_n} \int t K(t) \sum_{k=1}^m \int_0^1 \varphi_{ik}^{(1)} \left(\frac{x - l_k}{\gamma_n} - \frac{tz}{a_n \gamma_n} \right) dz dt,$$

where

$$\bar{f}_0(x) = a_n \int K(a_n(x - u)) f_0(u) du.$$

Therefore

$$A_{2n} = \frac{n \alpha_n^2}{a_n} \left\{ \sum_{i=1}^p k_i \int \left[\sum_{k=1}^m \varphi_{ik} \left(\frac{x - l_k}{\gamma_n} \right) - \right. \right.$$

$$\left. - \frac{1}{\bar{k}} \sum_{j=1}^p k_j \sum_{k=1}^m \varphi_{ik} \left(\frac{x - l_k}{\gamma_n} \right) \right]^2 r(x) dx + O(a_n^{-1}) +$$

$$\left. + O \left(\frac{1}{a_n \gamma_n} \int \left[\int t K(t) \sum_{k=1}^m \int_0^1 \varphi_{ik}^{(1)} \left(\frac{x - l_k}{\gamma_n} - \frac{tz}{a_n \gamma_n} \right) dt dz \right]^2 r(x) dx \right) \right\}.$$

Hence, by virtue of the generalized Minkowski inequality [5], we obtain

$$a_n^{1/2} A_{2n} = A_n(\varphi) + O \left(\frac{n \alpha_n^2}{a_n^{3/2}} \right) + O \left(\frac{n \alpha_n^2}{a_n^{5/2} \gamma_n} \right), \quad (3)$$

where

$$A_n(\varphi) = \frac{n\alpha_n^2}{\sqrt{a_n}} \sum_{i=1}^p k_i \int \left[\sum_{k=1}^m \varphi_{ik} \left(\frac{x-l_k}{\gamma_n} \right) - \frac{1}{k} \sum_{j=1}^p k_j \sum_{k=1}^m \varphi_{ik} \left(\frac{x-l_k}{\gamma_n} \right) \right]^2 r(x) dx.$$

Since

$$O(n\alpha_n^2 a_n^{-3/2}) = O\left(\frac{n\alpha_n^2 a_n^{-1/2} \gamma_n}{a_n \gamma_n}\right) = O\left(\frac{1}{a_n \gamma_n}\right)$$

and

$$O\left(\frac{n\alpha_n^2}{a_n^{5/2} \gamma_n}\right) = O\left(\frac{1}{(a_n \gamma_n)^2}\right),$$

from (3) we have

$$a_n^{1/2} A_{2n} = A_n(\varphi) + O\left(\frac{1}{a_n \gamma_n}\right). \quad (4)$$

Let us now discuss the asymptotic form of $A_n(\varphi)$. By Lebesgue's theorem on dominated convergence and the condition

$$\lim_{|x| \rightarrow \infty} \varphi_{ik}(x) = 0,$$

it is not difficult to establish that

$$\begin{aligned} \frac{1}{\gamma_n} \int \varphi_{ik} \left(\frac{x-l_k}{\gamma_n} \right) \varphi_{jk} \left(\frac{x-l_k}{\gamma_n} \right) r(x) dx &\rightarrow r(l_k) \int \varphi_{ik}(t) \varphi_{jk}(t) dt, \\ \frac{1}{\gamma_n} \int \varphi_{ir_1} \left(\frac{x-l_{r_1}}{\gamma_n} \right) \varphi_{ir_2} \left(\frac{x-l_{r_2}}{\gamma_n} \right) r(x) dx &\rightarrow 0, \quad r_1 \neq r_2, \end{aligned} \quad (5)$$

as $n \rightarrow \infty$.

Further, taking (5) into account, we find that

$$\begin{aligned} \int \left(\sum_{k=1}^m \varphi_{ik} \left(\frac{x-l_k}{\gamma_n} \right) \right)^2 r(x) dx &\rightarrow \sum_{k=1}^m r(l_k) \int \varphi_{ik}^2(x) dx, \\ \sum_{j=1}^p k_j \int \sum_{k=1}^m \varphi_{ik} \left(\frac{x-l_k}{\gamma_n} \right) \sum_{r=1}^m \varphi_{jr} \left(\frac{x-l_r}{\gamma_n} \right) r(x) dx &\rightarrow \\ &\rightarrow \sum_{j=1}^p k_j \sum_{k=1}^m r(l_k) \int \varphi_{ik}(t) \varphi_{jk}(t) dt, \\ \int \left(\sum_{j=1}^p k_j \sum_{k=1}^m \varphi_{jk} \left(\frac{x-l_k}{\gamma_n} \right) \right)^2 r(x) dx &\rightarrow \sum_{k=1}^m r(l_k) \int \left(\sum_{j=1}^p k_j \varphi_{jk}(t) \right)^2 dt. \end{aligned} \quad (6)$$

From the definition of $A_n(\varphi)$ and (6) it follows that

$$A_n(\varphi) \longrightarrow A(\varphi)$$

as $n \rightarrow \infty$. Therefore

$$a_n^{1/2} A_{2n} \longrightarrow A(\varphi). \quad (7)$$

Now let us show that $a_n^{1/2} A_{1n} \rightarrow 0$ in probability. For this it is sufficient to show that $a_n^{1/2} E|A_{1n}| \rightarrow 0$ as $n \rightarrow \infty$.

We have

$$\begin{aligned}
E|A_{1n}| &\leq 2\sqrt{n} \alpha_n \sum_{i=1}^p k_i \left\{ \int f_i(u) du \left[\int K(a_n(x-u)) \times \right. \right. \\
&\quad \times \left(\sum_{k=1}^m a_n \int K(a_n(x-v)) \varphi_{ik} \left(\frac{v-l_k}{\gamma_n} \right) dv - \right. \\
&\quad \left. \left. - \frac{1}{k} \sum_{j=1}^p k_j \sum_{k=1}^m a_n \int K(a_n(x-v)) \varphi_{jk} \left(\frac{v-l_k}{\gamma_n} \right) dv \right) r(x) dx \right]^2 \right\}^{1/2} \leq \\
&\leq c_1 \sqrt{n} \cdot \alpha_n \sum_{i=1}^p k_i \left\{ \int f_i(u) du \left[K_1(a_n(u-v)) \sum_{k=1}^m \left| \varphi_{ik} \left(\frac{v-l_k}{\gamma_n} \right) \right| dv + \right. \right. \\
&\quad \left. \left. + \frac{1}{k} \sum_{j=1}^p k_j \int K_1(a_n(u-v)) \sum_{k=1}^m \left| \varphi_{jk} \left(\frac{v-l_k}{\gamma_n} \right) \right| dv \right]^2 \right\}^{1/2} \leq c_2 \sqrt{n} \times \\
&\times \alpha_n \max_{1 \leq i \leq p} \left\{ \int f_i(u) du \left[\int K_1(a_n(u-v)) \sum_{k=1}^m \left| \varphi_{ik} \left(\frac{v-l_k}{\gamma_n} \right) \right| dv \right]^2 \right\}^{1/2} \leq \\
&\leq c_3 \frac{\sqrt{n} \cdot \alpha_n}{a_n} \left[\max_{1 \leq i \leq p} \left\{ \int f_i(u) \sum_{k=1}^m \varphi_{ik}^2 \left(\frac{v-l_k}{\gamma_n} \right) du \right\}^{1/2} + \frac{1}{a_n \gamma_n} \times \right. \\
&\times \max_{1 \leq i \leq p} \left\{ \int f_i(u) du \left(\int_0^1 t K(t) \sum_{k=1}^m \varphi_{ik}^{(1)} \left(\frac{u-l_k}{\gamma_n} - \frac{zt}{a_n \gamma_n} \right) dz dt \right)^2 \right\}^{1/2} \Big] \leq \\
&\leq c_4 \left(\frac{\sqrt{n} \alpha_n \gamma_n^{1/2}}{a_n} + \frac{\sqrt{n} \alpha_n \gamma_n^{1/2}}{a_n^2 \gamma_n} \right),
\end{aligned}$$

where $K_1(x) = \int |k(t)| |K(t-x)| dt$.

Hence

$$\begin{aligned}
a_n^{1/2} E|A_{1n}| &\leq \\
&\leq c_4 \left(\frac{\sqrt{n} \alpha_n \gamma_n^{1/2} a^{-1/4}}{a_n^{1/4}} + \frac{\sqrt{n} \alpha_n a_n^{-1/4} \gamma_n^{1/2}}{(a_n \gamma_n) a_n^{1/4}} \right) = O(a_n^{-1/4}). \quad (8)
\end{aligned}$$

Now consider the functional

$$T_n^{(1)} = \frac{n}{a_n} \sum_{i=1}^p k_i \int \left[\hat{f}_i(x) - E \hat{f}_i(x) - \frac{1}{k} \sum k_i (\hat{f}_j(x) - E \hat{f}_j(x)) \right]^2 r(x) dx.$$

After some simple transformations we get

$$T_n^{(1)} = \int \left[\sum_{i=1}^p \left(\sqrt{\frac{n_i}{a_n}} (\hat{f}_i(x) - E \hat{f}_i(x)) \right)^2 - \left(\sum_{j=1}^p \alpha_j \sqrt{\frac{n_j}{a_n}} (\hat{f}_j(x) - E \hat{f}_j(x)) \right)^2 \right] r(x) dx, \quad (9)$$

where

$$\alpha_i = \sqrt{\frac{k_i}{k_1 + \dots + k_p}}, \quad i = \overline{1, p}.$$

Let

$$\mathbb{Z}(x) = (Z_1(x), \dots, Z_p(x))$$

be the vector with components

$$Z_i(x) = \sqrt{\frac{n_i}{a_n}} (\hat{f}_i(x) - E \hat{f}_i(x)), \quad i = \overline{1, p}.$$

Then

$$T_n^{(1)} = \int \left[\sum_{i=1}^p Z_i^2(x) - \left(\sum_{j=1}^p \alpha_j Z_j(x) \right)^2 \right] r(x) dx.$$

As is known, there exists an orthogonal matrix $\mathbb{C} = \|c_{ij}\|$, $i, j = \overline{1, p}$, which depends only on $k_1, k_2, \dots, k_p$ and for which

$$c_{pi} = \alpha_i, \quad i = \overline{1, p}.$$

Let us transform the vector $\mathbb{Z}(x)$ by the orthogonal transformation $u = \mathbb{C}\mathbb{Z}$. As a result, we obtain

$$\begin{aligned} T_n^{(1)} &= \int \left[|\mathbb{C}\mathbb{Z}|^2 - \left(\sum_{j=1}^p \alpha_j Z_j(x) \right)^2 \right] r(x) dx = \\ &= \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^p c_{ij} Z_j(x) \right)^2 r(x) dx, \end{aligned} \quad (10)$$

where $|a|$ is the length of the vector $a = (a_1, \dots, a_p)$.

As is known ([2], [6]),

$$Z_i(x) = \sqrt{\frac{n_i}{a_n}} (\hat{f}_i(x) - E \hat{f}_i(x)) = \xi_i(x) + O_p\left(\frac{\ln n}{\sqrt{n a_n^{-1}}}\right), \quad (11)$$

uniformly with respect to x , where

$$\xi_i(x) = a_n^{1/2} \int K(a_n(x - u)) dW_i^0(F_i(u)), \quad i = \overline{1, p},$$

while $W_i^0(t)$ are independent Brownian bridges depending only on $X^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$ and $F_i(u)$ is the distribution function of the random variable $X_1^{(i)}$.

Now by virtue of (10) we can write

$$\sum_{j=1}^p c_{ij} Z_j(x) = \sum_{j=1}^p c_{ij} \xi_j(x) + O\left(\frac{\ln n}{\sqrt{n a_n^{-1}}}\right). \quad (12)$$

Next, let us estimate the variance of the variable

$$Y_i = \int \sum_{j=1}^p c_{ij} \xi_j(x) r(x) dx.$$

From the independence of W_i^0 , $i = \overline{1, p}$, we obtain

$$\text{Var } Y_i = \sum_{j=1}^p c_{ij}^2 \iint E \xi_j(x) \xi_j(y) r(x) r(y) dx dy.$$

$\xi_j(x)$ can be represented as

$$\xi_j(x) = a_n^{1/2} \int \left[K(a_n(x-t)) - \int K(a_n(x-u)) dF_j(u) \right] dW_j(F_j),$$

where $W_j(t)$, $j = \overline{1, p}$, are independent standard Wiener processes on $[0, 1]$.

Therefore

$$\begin{aligned} E \xi_j(x) \xi_j(y) &= a_n \left[\int K(a_n(x-u)) K(a_n(y-u)) f_j(u) du - \right. \\ &\quad \left. - \int K(a_n(x-u)) f_j(u) du \int K(a_n(y-u)) f_j(u) du \right]. \end{aligned}$$

Hence, after some simple transformations, we obtain

$$\iint E \xi_j(x) \xi_j(y) r(x) r(y) dx dy = O(a_n^{-1}).$$

Thus

$$\text{Var } Y_j = O(a_n^{-1}), \quad j = \overline{1, p}. \quad (13)$$

From representations (10) and (12), and also from relation (13) we have

$$\begin{aligned} T_n^{(1)} &= \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^p c_{ij} \xi_j(t) \right)^2 r(t) dt + O_p\left(\frac{\ln^2 n}{n a_n^{-1}}\right) = \\ &= T_n^{(2)} + O_p\left(\frac{\ln^2 n}{n a_n^{-1}}\right). \end{aligned} \quad (14)$$

Denote

$$\eta_i(t) = a_n^{1/2} \int K(a_n(t-u)) dW_i(F_i(u)), \quad i = \overline{1, p},$$

$$T_n^{(3)} = \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^p c_{ij} \eta_j(t) \right)^2 r(t) dt,$$

$$\mathcal{E}_i(t) = a_n^{1/2} W_i(1) \int K(a_n(t-u)) f_i(u) du, \quad i = \overline{1, p}.$$

Then

$$a_n^{1/2} (T_n^{(2)} - T_n^{(3)}) = o_p(1). \quad (15)$$

Indeed,

$$E|T_n^{(2)} - T_n^{(3)}| \leq 2 \sum_{i=1}^{p-1} E \left| \int \sum_{j=1}^p c_{ij} \eta_j(t) \sum_{r=1}^p c_{ir} \mathcal{E}_r(t) \right| r(t) dt +$$

$$+ \sum_{i=1}^{p-1} E \int \left(\sum_{j=1}^p c_{ij} \mathcal{E}_j(t) \right)^2 r(t) dt = B_n^{(1)} + B_n^{(2)}. \quad (16)$$

It is not difficult to see that

$$B_n^{(2)} = \sum_{i=1}^{p-1} E \int \sum_{j=1}^p c_{ij}^2 \mathcal{E}_j^2(t) r(t) dt =$$

$$= a_n \sum_{i=1}^{p-1} \sum_{j=1}^p c_{ij}^2 E W_j^2(1) \int \left[\int K(a_n(t-u)) f_j(u) du \right]^2 r(t) dt \leq c_5 a_n^{-1}.$$

Now let us estimate $B_n^{(1)}$. We have

$$B_n^{(1)} \leq 2 \sum_{i=1}^{p-1} \left[\sum_{j,r}^p |c_{ij} c_{ir}| E|W_r(1)| \times \right.$$

$$\left. \times \left| \int \left[\int \Psi_r(t) \cdot K(a_n(t-u)) r(t) dt \right] dW_j(F_j) \right| \leq \right.$$

$$\leq 2 \sum_{i=1}^{p-1} \left[\sum_{j=1}^p \sum_{r=1}^p |c_{ij} c_{ir}| E^{1/2} W_r^2(1) \times \right.$$

$$\left. \times E^{1/2} \left\{ \int \left[\int \Psi_r(t) K(a_n(t-u)) r(t) dt \right] dW_j(F_j) \right\}^2 = \right.$$

$$= 2 \sum_{i=1}^{p-1} \sum_{j=1}^p \sum_{r=1}^p |c_{ij} c_{ir}| \left\{ \int \left(\int \Psi_r(t) K(a_n(t-u)) r(t) dt \right)^2 \cdot dF_j(u) \right\}^{1/2} \leq$$

$$\leq c_6 a_n^{-1},$$

where

$$\Psi_r(t) = \int K(z) f_r(t - z a_n^{-1}) dz.$$

Thus, after substituting the estimates of the expressions $B_k^{(1)}$ and $B_k^{(2)}$ into (16), we get

$$a_n^{1/2} (T_n^{(1)} - T_n^{(3)})^{1/2} = o_p(1) + O_p\left(\frac{a_n^{3/2} \ln n}{n}\right). \quad (17)$$

Relation (17) provides the proof of (15).

Let us now investigate the limit distribution of the functional

$$T_n^{(3)} = \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^p c_{ij} \eta_j(t) \right)^2 r(t) dt.$$

Clearly, the processes $\eta_i(t)$, $i = \overline{1, p}$, are independent and Gaussian so that the new processes $\sum_{j=1}^p c_{ij} \eta_j(t)$, $i = \overline{1, p}$, are also independent and Gaussian since the matrix $\|c_{ij}\|$ is orthogonal. Thus, to find the limit distribution of the functional $T_n^{(3)}$, it remains to establish the limit distribution of the functional

$$U_n^{(i)} = \int \left(\sum_{j=1}^p c_{ij} \eta_j(t) \right)^2 r(t) dt$$

for each fixed i , $i = \overline{1, p-1}$.

The covariance function $R_n^{(i)}(t_1, t_2)$ of the Gaussian process $\sum_{j=1}^p c_{ij} \eta_j(t)$ is equal to

$$R_n^{(i)}(t_1, t_2) = \sum_{j=1}^p c_{ij}^2 E \eta_j(t_1) \eta_j(t_2).$$

However

$$\begin{aligned} E \eta_j(t_1) \eta_j(t_2) &= \int K(u) K(a_n(t_1 - t_2) + u) f_0(t_1 - a_n^{-1}u) du + \\ &+ \alpha_n \int K(a_n(t_1 - u)) K(a_n(t_2 - u)) \sum_{k=1}^m \varphi_{jk} \left(\frac{u - l_k}{\gamma_n} \right) du = \\ &= f_0(t_1) K_0(a_n(t_1 - t_2)) + O(a_n^{-1}) + O(\alpha_n \cdot \gamma_n), \end{aligned} \quad (18)$$

where the estimate $O(\cdot)$ is uniform with respect to t_1, t_2 and $K_0 = K * K$.

Since

$$a_n \alpha_n \gamma_n = (\sqrt{n} \cdot \alpha_n \cdot \gamma_n a_n) n^{-1/2} = o(1) \frac{a_n}{\sqrt{n}},$$

and $a_n^{-1} \sqrt{n} \geq c$, it follows from (18) that

$$R_n^{(i)}(t_1, t_2) = f_0(t_1) K_0(a_n(t_1 - t_2)) + O(a_n^{-1}).$$

The semiinvariant $\chi_n^{(i)}(m)$ of order m of the random variable $U_n^{(i)}$ is defined by the formula (see [7]; it also follows from the formula given in [8], p. 417)

$$\begin{aligned} \chi_n^{(i)}(m) &= \\ &= (m-1)2^{m-1} \int \cdots \int \prod_{j=1}^m R_n^{(i)}(u_j, u_{j+1}) r(u_j) du_j, \quad u_{m+1} = u_1. \end{aligned} \quad (19)$$

We obtain

$$\begin{aligned} EU_n^{(i)} &= \chi_n^{(i)}(1) = R(K) \int f_0(x) r(x) dx + O(a_n^{-1}), \\ \text{Var } U_n^{(i)} &= \chi_n^{(i)}(2) = 2R(K_0) \int f_0^2(x) r^2(x) dx + O(a_n^{-2}). \end{aligned} \quad (20)$$

Let us introduce the transformation

$$u_s = \tau_m + a_n^{-1} \sum_{i=s}^{m-1} \tau_i, \quad s = \overline{1, m-1},$$

and $u_m = \tau_m$. The Jacobian of the transformation is equal to $(a_n^{-1})^{m-1}$. Using this transformation in (19), it is not difficult to show that the m th semiinvariant $\chi^{(i)}(m)$ can be written to within terms of higher order of smallness as

$$(m-1)! 2^{m-1} (a_n^{-1})^{m-1} [K * K]^{(m)}(0) \int f_0^m(x) r^m(x) dx, \quad (21)$$

where $[K * K]^{(m)}(u)$ denotes the m th order convolution of $K_0(x)$ with itself.

From relations (20) and (21) it follows that

$$a_n^{1/2} \left(U_n^{(i)} - R(K) \int f_0(x) r(x) dx \right)$$

is distributed in the limit normally with mathematical expectation 0 and variance

$$2R(K_0) \int f_0^2(u) r^2(u) du,$$

and therefore

$$\sqrt{a_n} (T_n^{(3)} - \mu)$$

is distributed in the limit normally $(0, \sigma^2)$.

Eventually, taking into account (4), (7), (8), (14), (15) and the representation

$$\begin{aligned} a_n^{1/2} (T_n - \mu) &= a_n^{1/2} (T_n^{(3)} - \mu) + a_n^{1/2} (T_n^{(1)} - T_n^{(3)}) + \\ &+ A_n(\varphi) + O\left(\frac{1}{a_n \gamma_n}\right) + o_p(1) + O_p\left(\frac{a_n^{3/2} \ln^2 n}{n}\right), \end{aligned}$$

we conclude that

$$a_n^{1/2} (T_n - \mu)$$

is distributed in the limit normally with mathematical expectation $A(\varphi)$ and variance σ^2 . $\square$

The conditions of Theorem 1 imposed on a_n, α_n and γ_n are fulfilled if, for instance, $a_n = n^\delta$, $\alpha_n = n^{-\alpha}$, and $\gamma_n = n^{-\beta}$ for $\frac{\delta}{2} = 1 - 2\alpha - \beta$, $\alpha + \beta > \frac{1}{2}$, $0 < \delta \leq \frac{1}{2}$, $0 < \beta < \delta$, while the conditions imposed on α, β and δ are fulfilled if, for instance,

$$\begin{aligned} \delta = \frac{1}{2}, \quad \beta = \frac{5}{12}, \quad \alpha = \frac{1}{6}; \quad \delta = \frac{2}{5}, \quad \beta = \frac{2}{7}, \quad \alpha = \frac{18}{35}; \\ \delta = \frac{1}{4}, \quad \beta = \frac{1}{5}, \quad \alpha = \frac{27}{80}; \quad \delta = \frac{2}{9}, \quad \beta = \frac{1}{6}, \quad \alpha = \frac{13}{36}; \\ \delta = \frac{1}{5}, \quad \beta = \frac{1}{6}, \quad \alpha = \frac{11}{30} \end{aligned}$$

and so on.

Theorem 1 gives rise to two corollaries.

Corollary 1. *Let conditions 1, 2 and 3 be fulfilled on $K(x)$, $f_0(x)$ and $r(x)$. If $a_n^{-1} \cdot \sqrt{n} \geq c > 0$, then under the hypothesis H_0 the random variable*

$$a_n^{1/2} (T_n - \mu)$$

is distributed in the limit normally $(0, \sigma^2)$.

Using this corollary, we can construct the test for testing the hypothesis H_0 ; the critical region is defined by the inequality

$$T_n \geq d_n(\alpha),$$

where

$$d_n(\alpha) = \mu + a_n^{-1/2} \cdot \sigma \cdot \lambda_\alpha,$$

λ_α is a quantile of level α ($0 < \alpha < 1$) of the standard normal distribution $\Phi(x)$.

Corollary 2. *In the conditions of Theorem 1 the local behavior of the power $P_{H_1}(T_n \geq d_n(\alpha))$ is as follows:*

$$P_{H_1}(T_n \geq d_n(\alpha)) \longrightarrow 1 - \Phi\left(\lambda_\alpha - \frac{A(\varphi)}{\sigma}\right),$$

where

$$A(\varphi) = c_0 \sum_{i=1}^p k_i \sum_{k=1}^m r(l_k) \int \left[\varphi_{ik}(x) - \frac{1}{\bar{k}} \sum_{j=1}^p k_j \varphi_{jk}(x) \right]^2 dx,$$

$$\sigma^2 = 2(p-1) R(K_0) \int f_0^2(x) r^2(x) dx,$$

$$\mu = (p-1) R(K) \int f_0(x) r(x) dx, \quad R(g) = \int g^2(x) dx,$$

$$\bar{k} = k_1 + \dots + k_p, \quad c_0 = \lim_{n \rightarrow +\infty} n \cdot a_n^{-1/2} \cdot \alpha_n^2 \cdot \gamma_n.$$

It should be noted that for positive weight functions $r(x)$, the test of the hypothesis H_0 against alternatives of form (1) is asymptotically strictly unbiased since the mathematical expectation $A(\varphi) > 0$ is equal to 0 if and only if $\varphi_{ik}(x) = \varphi_{jr}(x)$, $i, j = \overline{1, p}$, $k, r = \overline{1, m}$.

By integrating $f_i(x)$, $i = \overline{1, p}$, we establish that the alternatives differ from the zero hypothesis in a value of order $\alpha_n \cdot \gamma_n = o(n^{-1/2})$. This means that tests based on a deviation between random distribution functions like Kolmogorov-Smirnov type tests cannot distinguish such hypotheses from the basic hypothesis H_0 .

Remark. If it is assumed in (1) that $m = 1$, $l_1 = 0$ and $\gamma_n = 1$, then from (1) we obtain the sequence of Pitman type alternatives

$$H_2 : f_i(x) = f_0(x) + \alpha_n \varphi_i(x) + o(\alpha_n).$$

In that case the following theorem is true.

Theorem 2. *Let conditions 1, 2 and 3 be fulfilled. If $na_n^{-1/2}\alpha_n^2 \rightarrow c_0$ and $a_n^{-1}\sqrt{n} \geq c$, then under the hypothesis H_2 the random variable*

$$a_n^{1/\alpha} (T_n - \mu)$$

is distributed in the limit normally (A_p, σ^2) ,

$$A_p = A(\varphi_1, \dots, \varphi_p) = c_0 \sum_{i=1}^p k_i \int \left[\varphi_i(x) - \frac{1}{k} \sum_{j=1}^p k_j \varphi_j(x) \right]^2 r(x) dx.$$

The proof of this theorem is similar to that of Theorem 1.

The conditions of Theorem 2 are fulfilled on a_n and α_n if, for instance, it is assumed that

$$a_n = n^\delta, \quad \alpha_n = n^{-1/2+\delta/4}, \quad 0 < \delta \leq \frac{1}{2}.$$

2. The statistic T_n is normalized by values depending on the distribution density $f_0(x)$. If however $f_0(x)$ is not defined by the hypothesis, then

$$\bar{\mu} = \int f_0(x) r(x) dx, \quad \Delta^2 = \int f_0^2(x) r^2(x) dx$$

in the normalizing constants have to be replaced by

$$\bar{\mu}_n = \int f_n^*(x) r(x) dx, \quad \Delta_n^2 = \frac{1}{k} \sum_{i=1}^p k_i \Delta_{in}^2,$$

respectively, where

$$f_n^*(x) = \frac{1}{k} \sum_{j=1}^p k_j \hat{f}_j(x),$$

$$\Delta_{in}^2 = \int \hat{f}_i(x) r^2(x) d\hat{F}_{n_i}(x) = \frac{a_n}{n_i^2} \sum_{j_1, j_2}^{n_i} U_{j_1, j_2}^{(i)},$$

$$U_{j_1, j_2}^{(i)} = K(a_n(X_{j_1}^{(i)} - X_{j_2}^{(i)})) r^2(X_{j_1}^{(i)}),$$

$\hat{F}_{n_i}(x)$ is the empirical distribution function of the sample $X^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$, $i = \overline{1, p}$.

The estimate Δ_{in}^2 for $r(x) \equiv 1$ was for the first time proposed by E. Schuster [9] and later by T. Schweder [10].

Let us in the first place establish the asymptotics of the mathematical expectation (under the hypothesis H_1) of the random variable Δ_{in}^2 .

We obtain

$$\begin{aligned} E \Delta_{in}^2 &= a_n \iint K(a_n(u-v)) r^2(u) f_i(v) f_i(u) du dv + O\left(\frac{a_n}{n}\right) = \\ &= a_n \iint K(a_n(u-v)) r^2(u) \left[f_0(u) + \alpha_n \sum_{k=1}^m \varphi_{ik}\left(\frac{u-l_k}{\gamma_n}\right) \right] \times \\ &\times \left[f_0(v) + \alpha_n \sum_{k=1}^m \varphi_{ik}\left(\frac{v-l_k}{\gamma_n}\right) \right] du dv + O\left(\frac{a_n}{n}\right) = \\ &= \int f_0(v) \left[a_n \int K(a_n(u-v)) r^2(u) f_0(u) du \right] dv + \\ &+ O(\alpha_n \cdot \gamma_n) = \Delta^2 + o(1) + O(\alpha_n \cdot \gamma_n). \end{aligned}$$

Therefore

$$E \Delta_n^2 = \Delta^2 + o(1).$$

Now

$$\text{Var } \Delta_n^2 = \frac{1}{k^2} \sum_{i=1}^p k_i^2 \text{Var } \Delta_{in}^2 \leq c_7 \max_{1 \leq k \leq p} \text{Var } \Delta_{kn}^2.$$

But it is not difficult to check that

$$\begin{aligned} E(\Delta_{kn}^2)^2 &= \\ &= E \left[\sum_{i,j} (U_{ij}^{(k)})^2 + \sum_i \sum_{j \neq t} U_{ij}^{(k)} U_{it}^{(k)} + \sum_j \sum_{i \neq r} U_{ij}^{(k)} U_{rj}^{(k)} + \sum_{i \neq r} \sum_{j \neq t} U_{ij}^{(k)} U_{rt}^{(k)} \right] = \\ &= \frac{a_n^2}{n_k^4} \left[n_k E U_{11}^{(k)} + 2n_k(n_k-1) E (U_{12}^{(k)})^2 + \right. \\ &+ 4n_k(n_k-1) E U_{11}^{(k)} U_{12}^{(k)} + 6n_k(n_k-1)(n_k-2) E U_{12}^{(k)} \cdot U_{13}^{(k)} + \\ &\left. + n_k(n_k-1) E U_{11}^{(k)} U_{22}^{(k)} + n_k(n_k-1)(n_k-2)(n_k-3) (E U_{12}^{(k)})^2 \right]. \end{aligned}$$

Further, likewise easily we find that

$$\begin{aligned} |E U_{12}^{(k)} U_{31}^{(k)}| &= \left| \iiint K(a_n(x-y)) K(a_n(z-x)) f_k(x) \times \right. \\ &\quad \left. \times f_k(y) f_k(z) r(x) r(z) dx dy dz \right| \leq \frac{c_8}{a_n^2}, \\ E(U_{12}^{(k)})^2 &= \iint K^2(a_n(u-v)) f_k(u) f_k(v) r^2(u) du dv \leq \frac{c_9}{a_n}. \end{aligned}$$

Other mathematical expectations are estimated in a similar manner. Therefore

$$E(\Delta_{kn}^2)^2 = \frac{n_k(n_k-1)(n_k-2)(n_k-3)}{n_k^4} [a_n E U_{12}^{(k)}]^2 + O\left(\frac{a_n}{n}\right).$$

After simplifying transformations and applying the formula $\text{Var } \Delta_{kn}^2 = E(\Delta_{kn}^2)^2 - (E \Delta_{kn}^2)^2$, we obtain $\max_{1 \leq k \leq p} \text{Var } \Delta_{kn}^2 \rightarrow 0$ as $n \rightarrow \infty$.

Thus we have shown that Δ_n^2 is an asymptotically unbiased and consistent estimate of the value

$$\Delta^2 = \int f_0^2(x) r^2(x) dx.$$

Further, we readily obtain

$$\begin{aligned} a_n^{1/2} E \int |f_n^*(x) r(x) dx - \int f_0(x) r(x) dx| &\leq \\ &\leq a_n^{1/2} E \int |f_n^*(x) - E f_n^*(x)| r(x) dx + a_n^{1/2} \int |E f_n^*(x) - \\ &\quad - f_0(x)| r(x) dx \leq c_{10} \left(a_n^{-1/2} + \sqrt{a_n} \alpha_n \gamma_n + \frac{\alpha_n}{\sqrt{a_n}} \right) + \\ &\quad + c_{11} a_n^{1/2} \max_{1 \leq j \leq p} \left\{ \frac{1}{n} \int f_j(u) du \left(\int k(t) r\left(u - \frac{t}{a_n}\right) \right)^2 - \right. \\ &\quad \left. - \frac{a_n^2}{n} \left(\iint K(a_n(x-t)) f_j(t) r(x) dx dt \right)^2 \right\}^{1/2} \leq \\ &\leq c_{10} \left(a_n^{-1/2} + \sqrt{a_n} \alpha_n \gamma_n + \frac{\alpha_n}{\sqrt{a_n}} \right) + c_{12} \left(\frac{a_n}{n} \right)^{1/2} \rightarrow 0, \end{aligned}$$

since $\sqrt{a_n} \alpha_n \cdot \gamma_n = \sqrt{n} \alpha_n \gamma_n \cdot \left(\frac{a_n}{n} \right)^{1/2} \rightarrow 0$.

Thus

$$a_n^{1/2} \left(\int f_n^*(x) r(x) dx - \int f_0(x) r(x) dx \right)$$

converges to zero in probability.

Finally, we summarize the obtained results in the following theorem.

Theorem 3. *Let the conditions of Theorem 1 be fulfilled. Then under the hypothesis H_1 the random variable*

$$a_n^{1/2} (T_n - \mu_n) \sigma_n^{-1}$$

is asymptotically normal with the mean $A(\varphi)$ and variance 1. Here

$$\mu_n = \bar{\mu}_n \cdot (p-1) \cdot R(K), \quad \sigma_n^2 = \Delta_n^2 \cdot (p-1) \cdot R(K_0).$$

Corollary 1. *Under the hypothesis $H_0 : f_1(x) = \dots = f_p(x)$ the random variable*

$$a_n^{1/2} (T_n - \mu_n) \sigma_n^{-1}$$

is asymptotically normal with mean 0 and variance 1.

This corollary makes it possible to construct an asymptotic test of level α ($0 < \alpha < 1$) for testing the hypothesis $H_0 : f_1(x) = \dots = f_p(x)$ (homogeneity hypothesis); the critical domain for testing this hypothesis is defined by the inequality

$$T_n \geq \tilde{d}_n(\alpha) = \mu_n + a_n^{-1/2} \sigma_n \cdot \lambda_\alpha,$$

where λ_α is a quantile of level α of the distribution function of the standard normal law.

Corollary 2. *The local behavior of the power $P_{H_1}(T_n \geq \tilde{d}_n(\alpha))$ is as follows:*

$$\Phi_{H_1}(T_n \geq \tilde{d}_n(\alpha)) \longrightarrow 1 - \Phi\left(\lambda_\alpha - \frac{A(\varphi)}{\sigma}\right).$$

Note that a statistic similar to (2) was studied in [11] for several independent parametric estimates of a density when an estimate of a density parameter is asymptotically normal (see also [12]).

REFERENCES

1. M. Rosenblatt, A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Math. Statist.* **3**(1975), 1–14.
2. E. A. Nadaraya, Nonparametric estimation of probability densities and regression curves. *Kluwer Academic Publishers, Dordrecht*, 1989.
3. E. A. Nadaraya, Limit distribution of the quadratic distribution of two nonparametric estimates of distribution density. (Russian) *Soobshch. Akad. Nauk Gruz. SSR* **78**(1975), No. 1, 25–28.
4. N. H. Anderson, P. Hall, and D. M. Titterton, Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50**(1994), 41–54.

5. S. N. Nikolskii, Approximation of functions of many variables and embedding theorems. (Russian) *Nauka, Moscow*, 1969.
6. P. Hall, Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process. Appl.* **13**(1982), 11–25.
7. P. Bickel and M. Rosenblatt, On some global measures of the deviations of density function estimates. *Ann. Math. Statist.* **1**(1973), No. 6, 1071–1095.
8. I. I. Gihman and A. V. Skorohod, Theory of random processes. (Russian) *Nauka, Moscow*, 1971.
9. E. Schuster, On the rate of convergence of an estimate of a functional of a probability density. *Scand. Actuar. J.* **1**(1974), 103–107.
10. T. Schweder, Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist.* **2**(1975), 113–126.
11. T. L. Shervashidze, Mean square deviation of parametric estimates of the distribution density. (Russian) *Trudy Inst. Prikl. Mat. Tbiliss. Gos. Univ.* **4**(1975), 242–255.
12. T. Shervashidze, Some properties of a parametric estimate of the distribution density. *Seminari di probabilità e statistica matematica aa. aa. 1995/96 e 1996/97*, 99–146, *Edizioni dell'Università degli Studi di Cassino, Cassino*, 1998.

(Received 10.01.2003)

Authors' addresses:

R. Absava
 Sukhumi Branch of I. Javakhishvili Tbilisi State University
 12, Jikia St., Tbilisi 380043
 Georgia

E. Nadaraya
 Faculty of Mechanics and Mathematics
 I. Javakhishvili Tbilisi State University
 2, University St., Tbilisi 380043
 Georgia