УДК 519.21

П. К. Бабилуа, Э. А. Надарая, Г. А. Сохадзе (Тбил. гос. ун-т им. Ив. Джавахишвили, Грузия)

ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЯ

We construct new criteria for testing the hypotheses that $p \geq 2$ the independent samplings have identical distribution densities (homogeneity hypothesis) or the same well-defined distribution density (goodness-of-fit test). The limiting power of the constructed criteria for some local "close" alternatives are obtained.

Побудовано критерії перевірки гіпотез про те, що $p \geq 2$ незалежних вибірок мають однакові щільності розподілу (гіпотеза однорідності) або однакову визначену щільність розподілу (гіпотеза згоди). Знайдено граничну потужність побудованих критеріїв при деяких локальних „близьких" альтернативах.

Пусть $X^{(i)} = (X_1^{(i)}, \ldots, X_{n_i}^{(i)})$, $i = 1, \ldots, p$, — независимые выборки объемов $n_1, n_2, \ldots, n_p$ из $p \geq 2$ генеральных совокупностей с плотностями распределения $f_1(x), \ldots, f_p(x)$. Требуется, основываясь на выборках $X^{(i)}$, $i = 1, \ldots, p$, проверить две гипотезы: гипотезу однородности

$$H_0: \; f_1(x) = \ldots = f_p(x) \tag{1}$$

и гипотезу согласия

$$H_0': \; f_1(x) = \ldots = f_p(x) = f_0(x), \tag{2}$$

где $f_0(x)$ — вполне определенная функция плотности. В случае гипотезы H_0 общая плотность распределения $f_0(x)$ неизвестна.

В работе строятся критерии для проверки гипотез H_0 и H_0' для последовательности „близких" альтернатив [1, 2]:

$$H_1: \; f_i(x) = f_0(x) + \alpha(n_0)\varphi_i\left(\frac{x - \ell_i}{\gamma(n_0)}\right) \qquad (\,\alpha(n_0), \gamma(n_0) \longrightarrow 0\,),$$

$$\int \varphi_i(x)\,dx = 0, \qquad n_0 = \min(n_1, \ldots, n_p) \longrightarrow \infty.$$

Мы рассматриваем критерий проверки гипотез H_0 и H_0', основанный на статистике

$$T(n_1, n_2, \ldots, n_p) = \sum_{i=1}^{p} N_i \int \left[\widehat{f}_i(x) - \frac{1}{N}\sum_{j=1}^{p} N_j \widehat{f}_j(x)\right]^2 r(x)\,dx, \tag{3}$$

где $\widehat{f}_i(x)$ — ядерная оценка Розенблатта – Парзена плотности распределения $f_i(x)$:

$$\widehat{f}_i(x) = \frac{a_i}{n_i}\sum_{j=1}^{n_i} K\left(a_i(x - X_j^{(i)})\right), \qquad N_i = \frac{n_i}{a_i}, \qquad N = N_1 + \ldots + N_p.$$

Частный случай $p = 2$ рассматривался в работах [3, 4]. В этом случае статистика T принимает наиболее наглядный вид

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

1. В настоящем пункте найдено предельное распределение статистики (3) при гипотезе H_1 в случае, когда n_i неограниченно возрастают, так что $n_i = nk_i$, где $n \to \infty$, а k_i — постоянные. Пусть $a_1 = a_2 = \ldots = a_p = a_n$, причем $a_n \to \infty$ при $n \to \infty$.

Для получения предельного закона распределения функционала $T_n = T(n_1, \ldots, n_p)$ введем предположения относительно функций $K(x)$, $f_0(x)$, $\varphi_i(x)$, $i = 1, \ldots, p$, и $r(x)$:

(i) $K(x) \geq 0$ — функция с ограниченным изменением,

$$\int K(x)\,dx = 1, \qquad \int x^2 K(x)\,dx < \infty.$$

(ii) Функция плотности $f_0(x)$ ограничена и положительна на $(-\infty, \infty)$ либо ограничена и положительна в некотором конечном интервале $[c, d]$. Кроме того, она имеет в области положительности ограниченную производную.

(iii) Функции $\varphi_j(x)$, $j = 1, \ldots, p$, ограничены и имеют ограниченные производные первого порядка, причем $\varphi_i(x)$ и $\varphi_i^{(1)}(x) \in L_1(-\infty, \infty)$.

(iv) Весовая функция $r(x)$ кусочно-непрерывна, ограничена и интегрируема, причем $r(\ell_k) \neq \neq 0$, $k = 1, \ldots, p$, где ℓ_k — некоторые фиксированные точки непрерывности $r(x)$.

Справедлива следующая теорема.

Теорема 1. *Пусть выполняются условия* (i)–(iv), *причем* $f_i(x) \geq 0$, $x \in (-\infty, \infty)$. *Если* $na_n^{-1/2}\alpha_n^2\gamma_n \to c_0 \neq 0$, $a_n\gamma_n \to \infty$, $\alpha_n\gamma_n = o(n^{-1/2})$ $(\alpha_n = \alpha(n_0),\ \gamma_n = \gamma(n_0))$, $na_n^{-2} \to \infty$ *и* $a_n^2\alpha_n\gamma_n \to 0$, *то случайная величина* $a_n^{1/2}(T_n - \mu)$ *при гипотезе* H_1 *распределена в пределе нормально с математическим ожиданием* $A(\varphi)$ *и дисперсией* σ^2, *где*

$$A(\varphi) = c_0 \sum_{i=1}^{p} \left(k_i - \frac{k_i^2}{\overline{k}}\right) r(\ell_i) \int \varphi_i^2(x)\,dx,$$

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

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

$$\overline{k} = k_1 + \ldots + k_p, \qquad p \geq 2.$$

Доказательство. Представим T_n в виде суммы

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

где

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

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

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

Здесь и далее $E(\,\cdot\,)$ — математическое ожидание относительно гипотезы H_1.

Нетрудно видеть, что

$$E\widehat{f}_i(x) = \overline{f}_0(x) + \alpha_n \varphi_i \left(\frac{x-\ell_i}{\gamma_n} \right) -$$

$$-\frac{\alpha_n}{a_n\gamma_n} \int tK(t) \int_0^1 \varphi_i^{(1)} \left(\frac{x-\ell_i}{\gamma_n} - \frac{tz}{a_n\gamma_n} \right)\, dz\, dt,$$

где

$$\overline{f}_0(x) = a_n \int K\left(a_n(x-u)\right) f_0(u)\, du.$$

Поэтому

$$\sqrt{a_n}\, A_{2n} = A_n(\varphi) + 2\, \frac{n\alpha_n^2}{\sqrt{a_n}} \sum_{i=1}^{p} k_i \int A_i(x) B_i(x) r(x)\, dx +$$

$$+ \frac{n\alpha_n^2}{\sqrt{a_n}} \sum_{i=1}^{p} k_i \int B_i^2(x) r(x)\, dx = A_n(\varphi) + L_{1n} + L_{2n}.$$

Здесь

$$A_n(\varphi) = \frac{n\alpha_n^2}{\sqrt{a_n}} \sum_{i=1}^{p} k_i \int \left[\varphi_i \left(\frac{x-\ell_i}{\gamma_n} \right) - \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \varphi_j \left(\frac{x-\ell_j}{\gamma_n} \right) \right]^2 r(x)\, dx,$$

$$A_i(x) = \left[\varphi_i \left(\frac{x-\ell_i}{\gamma_n} \right) - \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \varphi_j \left(\frac{x-\ell_j}{\gamma_n} \right) \right],$$

$$B_i(x) = \frac{1}{a_n\gamma_n} \left[\frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \int tK(t) \int_0^1 \varphi^{(1)} \left(\frac{x-\ell_j}{\gamma_n} - \frac{tz}{a_n\gamma_n} \right)\, dt\, dz - \right.$$

$$\left. - \int tK(t) \int_0^1 \varphi_i^{(1)} \left(\frac{x-\ell_i}{\gamma_n} - \frac{tz}{a_n\gamma_n} \right)\, dt\, dz \right].$$

В силу условия (iii) имеем

$$L_{1n} \le c_1\, \frac{n\alpha_n^2}{a_n^{3/2}},$$

а в силу обобщенного неравенства Минковского [5] и условия (iii) получаем

$$L_{2n} \le c_2 \, \frac{n\alpha_n^2}{a_n^{5/2}\gamma_n} \, .$$

Таким образом,

$$\sqrt{a_n}\, A_{2n} = A_n(\varphi) + O\left(\frac{n\alpha_n^2}{a_n^{3/2}}\right) + O\left(\frac{n\alpha_n^2}{a_n^{5/2}\gamma_n}\right). \tag{4}$$

Поскольку

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

и

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

то из (4) находим

$$a_n^{1/2} A_{2n} = A_n(\varphi) + O\left(\frac{1}{a_n\gamma_n}\right). \tag{5}$$

Выясним теперь асимптотический вид $A_n(\varphi)$. Согласно теореме Лебега о мажорируемой сходимости с учетом того, что $\lim_{|x|\to\infty}\varphi_i(x) = 0$ [6, с. 429], легко установить, что при $n \to \infty$

$$\begin{gathered} \frac{1}{\gamma_n}\int \varphi_i^2\left(\frac{x-\ell_i}{\gamma_n}\right) r(x)\,dx \longrightarrow r(\ell_i)\int \varphi_i^2(x)\,dx, \\ \frac{1}{\gamma_n}\int \varphi_i\left(\frac{x-\ell_i}{\gamma_n}\right)\varphi_j\left(\frac{x-\ell_j}{\gamma_n}\right) r(x)\,dx \longrightarrow 0, \quad i \ne j. \end{gathered} \tag{6}$$

Далее, принимая во внимание (6), имеем

$$\begin{gathered} \frac{1}{\gamma_n}\sum_{j=1}^{p} k_j \int \varphi_i\left(\frac{x-\ell_i}{\gamma_n}\right)\varphi_j\left(\frac{x-\ell_j}{\gamma_n}\right) r(x)\,dx \longrightarrow k_i r(\ell_i)\int \varphi_i^2(u)\,du, \\ \frac{1}{\gamma_n}\int\left(\sum_{j=1}^{p} k_j\varphi_j\left(\frac{x-\ell_j}{\gamma_n}\right)\right)^2 r(x)\,dx \longrightarrow \sum_{i=1}^{p} k_i^2 r(\ell_i)\int \varphi_j^2(u)\,du. \end{gathered} \tag{7}$$

Из определения $A_n(\varphi)$ и (7) следует, что при $n \to \infty$

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

Следовательно,

$$a_n^{1/2} A_{2n} \longrightarrow A(\varphi). \tag{8}$$

Теперь покажем, что $a_n^{1/2} A_{1n} \longrightarrow 0$ по вероятности. Для этого достаточно показать, что $a_n^{1/2} E|A_{1n}| \longrightarrow 0$ при $n \to \infty$.

Имеем

$$E|A_{1n}| \le (EA_{1n}^2)^{1/2} = 2\,\frac{n}{a_n}\left\{\sum_{i=1}^{p} k_i^2 E\left[\int \left(\widehat{f}_i(x) - E\widehat{f}_i(x)\right) A_i(x) r(x)\,dx\right]^2\right\}^{1/2},$$

где

$$A_i(x) = E\widehat{f}_i(x) - \frac{1}{\bar{k}}\sum_{j=1}^{p} k_j E\widehat{f}_j(x).$$

Далее, нетрудно видеть, что

$$E\left[\int \left(\widehat{f}_i(x) - E\widehat{f}_i(x)\right) A_i(x) r(x)\,dx\right]^2 =$$

$$= \frac{a_n^2}{k_i n}\,E\left[\int K\left(a_n(x_1 - X_1^{(i)})\right) A_i(x_1) r(x_1)\,dx_1 - \right.$$

$$\left. - E\int k\left(a_n(x_1 - X_1^{(i)})\right) A_i(x_1) r(x_1)\,dx_1\right]^2 \le$$

$$\le \frac{a_n^2}{k_i n}\,E\left[\int K\left(a_n(x - X_1^{(i)})\right) A_i(x) r(x)\,dx\right]^2.$$

Поэтому

$$E|A_{1n}| \le c_3\sqrt{n}\left\{\sum_{i=1}^{p} k_i \int f_i(u)\,du\left[\int K(a_n(x_1 - u)) A_i(x_1) r(x_1)\,dx_1\right]^2\right\}^{1/2} \le$$

$$\le c_4\sqrt{n}\,\alpha_n \max_{1\le i\le p}\left\{\int f_i(u)\,du\left[\int K_1(a_n(u-v)\Big|\varphi_i\left(\frac{v-\ell_i}{\gamma_n}\right)\Big|\,dv\right]^2\right\}^{1/2} \le$$

$$\le c_5\frac{\sqrt{n}\,\alpha_n}{a_n}\left[\max_{1\le i\le p}\left\{\int f_i(u)\varphi_i^2\left(\frac{v-\ell_i}{\gamma_n}\right)du\right\}^{1/2} + \right.$$

$$\left. + \frac{1}{a_n\gamma_n}\max_{1\le i\le p}\left\{\int f_i(u)\,du\left[\int\int_0^1 |t| K_1(t)\Big|\varphi_i^{(1)}\left(\frac{u-\ell_i}{\gamma_n} - \frac{zt}{a_n\gamma_n}\right)\Big|\,dz\,dt\right]^2\right\}^{1/2}\right] \le$$

$$\le c_6\left(\frac{\sqrt{n}\,\alpha_n\gamma_n^{1/2}}{a_n} + \frac{\sqrt{n}\,\alpha_n\gamma_n^{1/2}}{a_n^2\gamma_n}\right),$$

где

$$K_1(x) = \int K(t)K(t-x)\,dt.$$

Таким образом,

$$a_n^{1/2} E|A_{1n}| \le c_7 \left(\frac{\sqrt{n}\,\alpha_n \gamma_n^{1/2} a_n^{-1/4}}{a_n^{1/4}} + \frac{\sqrt{n}\,\alpha_n a_n^{-1/4} \gamma_n^{1/2}}{(a_n\gamma_n) a_n^{1/4}} \right) = O(a_n^{-1/4}). \tag{9}$$

Рассмотрим теперь функционал

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

где $\overline{k} = k_1 + \ldots + k_p$.

После простого преобразования получим

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

где $\alpha_i^2 = \dfrac{k_i}{k_1 + \ldots + k_p}$.

Пусть

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

— вектор с компонентами

$$Z_i(x) = \sqrt{\frac{n_i}{a_n}} \left(\widehat{f}_i(x) - E\widehat{f}_i(x) \right), \quad i = 1, \ldots, p.$$

Тогда

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

где $|a|$ — длина вектора $a = (a_1, \ldots, a_p)$.

Как известно, существует ортогональная матрица $\mathbf{C} = \|c_{ij}\|$, $i, j = 1, \ldots, p$, зависящая только от $k_1, k_2, \ldots, k_p$, для которой

$$c_{pi} = \alpha_i = \sqrt{\frac{k_i}{k_1 + \ldots + k_p}}, \qquad i = 1, \ldots, p.$$

Поскольку при ортогональном преобразовании длина вектора не меняется, то

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

Известно [2, 7], что

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

равномерно по x, где

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

а $W_i^0(t)$ — независимые броуновские мосты, зависящие соответственно только от $X^{(i)} = (X_1^{(i)}, \ldots, X_{n_i}^{(i)})$, $F_i(u)$ — функция распределения случайной величины $X_1^{(i)}$.

Тогда согласно (11) запишем

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

Далее, оценим дисперсию величины

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

Вследствие независимости W_i^0, $i = 1, \ldots, p$, получим

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

$\xi_j(x)$ можно представить так:

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

где $W_j(t)$, $j = 1, \ldots, p$, — независимые стандартные винеровские процессы на $[0,1]$.

Поэтому

$$E\xi_j(x)\xi_j(y) = a_n \left[\int K(a_n(x-u)) K(a_n(y-u)) f_j(u)\, du - \right.$$

$$\left. - \int K(a_n(x-u)) f_j(u)\, du \int K(a_n(y-u) f_j(u)\, du\right].$$

Отсюда после несложных преобразований получаем

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

Следовательно,

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

Из представлений (10) и (12), а также соотношения (13) находим

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

где

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

Обозначим

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

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

$$\varepsilon_i(t) = a_n^{1/2} W_i(1) \int K(a_n(t-u)) f_i(u)\, du.$$

Тогда

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

Действительно,

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

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

Нетрудно видеть, что

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

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

Оценим теперь $B_n^{(1)}$. Имеем

$$B_n^{(1)} \leq 2 \sum_{i=1}^{p-1} \left[\sum_{j,r}^{p} |c_{ij} c_{ir}| E|W_r(1)| \left| \int \left[\int \Psi_r(t) K(a_n(t-u)) r(t)\, dt \right] dW_j(F_j) \right| \right] \leq$$

$$\leq 2 \sum_{i=1}^{p-1} \left[\sum_{j=1}^{p} \sum_{r=1}^{p} |c_{ij} c_{ir}| E^{1/2} W_r^2(1) E^{1/2} \left\{ \int \left[\int \Psi_r(t) K(a_n(t-u)) r(t)\, dt \right] dW_j(F_j) \right\}^2 \right] =$$

$$= 2 \sum_{i=1}^{p-1} \sum_{j=1}^{p} \sum_{r=1}^{p} |c_{ij} c_{ir}| \left\{ \int \left(\int \Psi_r(t) K(a_n(t-u)) r(t)\, dt \right)^2 dF_j(u) \right\}^{1/2} \leq c_6 a_n^{-1},$$

где

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

Итак, подставляя оценки для выражений $B_n^{(1)}$ и $B_n^{(2)}$ в (16), находим

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

откуда следует справедливость (15).

Обозначим

$$\eta_i^0(t) = a_n^{1/2}\int K(a_n(t-x))\, dW_i(F_0),$$

где $F_0(x)$ — функция распределения с плотностью $f_0(x)$. Поскольку $F_i(x) = F_0(x) + O(\alpha_n\gamma_n)$, то

$$E\left|\eta_i(t) - \eta_i^0(t)\right|^2 = O(a_n\alpha_n\gamma_n),$$

а в силу ограниченности $f_0(x)$ имеем

$$E(\eta_i^0(t))^2 = O(1), \quad i = 1, 2, \ldots, p,$$

равномерно по $x \in (-\infty, \infty)$.

Действительно,

$$E\left(\eta_i(t) - \eta_i^0(t)\right)^2 = a_n E\left(\int\left[W_i\left(F_i\left(t - \frac{z}{a_n}\right)\right) - W_i\left(F_0\left(t - \frac{z}{a_n}\right)\right)\right] dK(z)\right)^2 \le$$

$$\le a_n\int E\left[W_i\left(F_i\left(t - \frac{z}{a_n}\right)\right) - W_i\left(F_0\left(t - \frac{z}{a_n}\right)\right)\right]^2 |dK(z)| \int |dK(z)| \le$$

$$\le a_n\int\left|F_i\left(t - \frac{z}{a_n}\right) - F_0\left(t - \frac{z}{a_n}\right)\right| |dK(z)| \int |dK(z)| \le$$

$$\le c_8 a_n\alpha_n\gamma_n.$$

Далее,

$$E|\eta_n^0(t)|^2 = a_n\int K^2(a_n(t-z)) f_0(z)\, dz \le \int K^2(u) f_0\left(t - \frac{u}{a_n}\right) du \le c_9.$$

Обозначим

$$T_n^{(4)} = \sum_{i=1}^{p-1}\int\left(\sum_{j=1}^{p} c_{ij}\eta_j^0(t)\right)^2 r(t)\, dt.$$

Тогда из неравенства Коши – Шварца имеем

$$\sqrt{a_n}\, E|T_n^{(3)} - T_n^{(4)}| \le 2\sqrt{a_n}\sum_{i=1}^{n}\sum_{j_1, j_2=1}^{p}|c_{ij_1}c_{ij_2}|\int E|\eta_{j_1}^0(t)|\left|\eta_{j_2}(t) - \eta_{j_2}^0(t)\right| r(t)\, dt +$$

$$+\sqrt{a_n}\sum_{i=1}^{p-1}\int E\left(\sum_{j=1}^{p} c_{ij}\left(\eta_j(t)-\eta_j^0(t)\right)\right)^2 r(t)\,dt \le$$

$$\le c_{10}a_n\sqrt{\alpha_n\gamma_n}+c_{11}a_n^{3/2}\alpha_n\gamma_n \longrightarrow 0. \tag{18}$$

Перейдем к изучению предельного распределения функционала

$$T_n^{(4)}=\sum_{i=1}^{p-1}\int\left(\sum_{j=1}^{p}c_{ij}\eta_j^0(t)\right)^2 r(t)\,dt.$$

Ясно, что процессы $\eta_i^0(t)$, $i=1,\ldots,p$, независимые и гауссовские, так что новые процессы $\sum_{j=1}^{p}c_{ij}\eta_j^0(t)$, $i=1,\ldots,p$, также независимые и гауссовские в силу ортогональности матрицы $\|c_{ij}\|$. Поэтому для нахождения предельного распределения $T_n^{(4)}$ осталось установить предельное распределение функционала

$$U_n^{(i)}=\int\left(\sum_{j=1}^{p}c_{ij}\eta_j^0(t)\right)^2 r(t)\,dt$$

при каждом фиксированном i, $i=1,\ldots,p-1$.

Ковариационная функция $R_n^{(i)}(t_1,t_2)$ гауссовского процесса $\sum_{j=1}^{p}c_{ij}\eta_j^0(t)$ имеет вид

$$R_n^{(i)}(t_1,t_2)=\sum_{j=1}^{p}c_{ij}^2E\eta_j^0(t_1)\eta_j^0(t_2).$$

Однако

$$E\eta_j^0(t_1)\eta_j^0(t_2)=\int K(u)K(a_n(t_1-t_2)+u)f_0(t_1-a_n^{-1}u)\,du=$$

$$=f_0(t_1)K_0(a_n(t_1-t_2))+O(a_n^{-1}), \tag{19}$$

причем оценка $O(\,\cdot\,)$ равномерна по t_1, t_2 и $K_0=K*K$.

Из (19) следует, что

$$R_n^{(i)}(t_1,t_2)=f_0(t_1)K_0\left(a_n(t_1-t_2)\right)+O(a_n^{-1}). \tag{20}$$

Семиинвариант $\chi_n^{(i)}(s)$ порядка s случайной величины $U_n^{(i)}$ дается формулой [8]

$$\chi_n^{(i)}(s)=(s-1)!\,2^{s-1}\int\ldots\int R_n^{(i)}(x_1,x_2)R_n^{(i)}(x_2,x_3)\ldots R_n^{(i)}(x_s,x_1)\times$$

$$\times r(x_1)r(x_2)\ldots r(x_s)\,dx_1\,dx_2\ldots dx_s. \tag{21}$$

Из (20) и (21) нетрудно установить, что

$$\begin{aligned} EU_n^{(i)}&=\chi_n^{(i)}(1)=R(K)\int f_0(x)r(x)\,dx+O(a_n^{-1}),\\ \operatorname{Var}U_n^{(i)}&=\chi_n^{(i)}(2)=2R(K_0)a_n^{-1}\int f_0^2(x)r^2(x)\,dx+o(a_n^{-1}), \end{aligned} \tag{22}$$

и s-й семиинвариант $\chi_n^{(i)}(s)$ с точностью до членов высшего порядка малости равен [8]

$$(s-1)!\,2^{s-1}(a_n^{-1})^{s-1}[K*K]^{(s)}(0)\int f_0^s(x)r^s(x)\,dx, \tag{23}$$

где (s) обозначает s-кратную свертку $K_0(x)$ с самим собой.

Из соотношений (22), (23) следует, что случайная величина (см. также [2, 8])

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

распределена в пределе нормально с математическим ожиданием 0 и дисперсией

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

и, следовательно, случайная величина $\sqrt{a_n}\,(T_n^{(4)}-\mu)$ распределена в пределе нормально с математическим ожиданием 0 и дисперсией σ^2.

Наконец, принимая во внимание (5), (8), (9), (14), (15), (18) и представление

$$a_n^{1/2}(T_n-\mu) = a_n^{1/2}(T_n^{(4)}-\mu)+$$
$$+A_n(\varphi) + O\left(\frac{1}{a_n\gamma_n}\right) + o_p(1) + O_p\left(\frac{a_n^{3/2}\ln^2 n}{n}\right) + O_p\left((a_n^2\alpha_n\gamma_n)^{1/2}\right),$$

заключаем, что случайная величина $a_n^{1/2}(T_n-\mu)$ распределена в пределе нормально с математическим ожиданием $A(\varphi)$ и дисперсией σ^2.

Условия теоремы 1 относительно a_n, α_n и γ_n выполняются, например, если положить $a_n = n^{\delta}$, $\alpha_n = n^{-\alpha}$, $\gamma_n = n^{-\beta}$ при $\dfrac{\delta}{2} = 1-2\alpha-\beta$, $\alpha+\beta > \dfrac{1}{2}$, $0<\delta<\dfrac{1}{2}$, $0<\beta<\delta$, а условия на α, β и δ выполняются, например, если

$$\delta=\frac{1}{4}\,,\qquad \beta=\frac{1}{5}\,,\qquad \alpha=\frac{27}{80}\,;\qquad\qquad \delta=\frac{2}{9}\,,\qquad \beta=\frac{1}{6}\,,\qquad \alpha=\frac{13}{36}\,;$$
$$\delta=\frac{1}{5}\,,\qquad \beta=\frac{1}{6}\,,\qquad \alpha=\frac{11}{30}\qquad \text{и т. д.}$$

Из теоремы 1 следуют два утверждения.

***Следствие* 1.** *Пусть выполняются условия* (i), (ii) *и* (iv) *относительно* $K(x)$, $f_0(x)$ *и* $r(x)$. *Если* $na_n^{-2}\to\infty$, *то случайная величина* $a_n^{1/2}(T_n-\mu)$ *при гипотезе* H_0' *распределена в пределе нормально с математическим ожиданием* 0 *и дисперсией* σ^2.

С помощью следствия можно построить критерий для проверки гипотезы H_0'; критическая область для проверки этой гипотезы устанавливается неравенством

$$T_n \ge d_n(\alpha), \tag{24}$$

где

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

λ_α — квантиль уровня $1-\alpha$, $0<\alpha<1$, стандартного нормального распределения $\Phi(x)$.

***Следствие* 2.** *При условиях теоремы* 1 *локальное поведение мощности* $P_{H_1}(T_n \geq d_n(\alpha))$ *таково*: *при* $n \to \infty$

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

2. Введем обозначения

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

$$\overline{\mu}_n = \int f_n^*(x) r(x)\, dx,$$

$$\Delta_n^2 = \frac{1}{\overline{k}} \sum_{i=1}^{p} k_i \Delta_{in}^2, \qquad \Delta_{in}^2 = \int \widehat{f}_i^2(x) r^2(x)\, dx.$$

Справедлива следующая теорема.

Теорема 2. *Пусть выполняются все условия теоремы* 1. *Тогда случайная величина*

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

где

$$\mu_n = (p-1)R(K)\overline{\mu}_n, \qquad \sigma_n^2 = 2(p-1)R(K_0)\Delta_n^2,$$

при гипотезе H_1 *распределена в пределе нормально с математическим ожиданием* $A(\varphi)\sigma^{-1}$ *и дисперсией* 1.

Доказательство. Очевидно,

$$a_n^{1/2}(T_n - \mu_n)\sigma_n^{-1} = a_n^{1/2}(T_n - \mu)\sigma^{-1}(\sigma\sigma_n^{-1}) + a_n^{1/2}(\mu - \mu_n)\sigma_n^{-1}.$$

Поэтому достаточно показать, что

$$a_n^{1/2}\left(\overline{\mu}_n - \int f_0(x) r(x)\, dx\right) = o_p(1) \tag{25}$$

и

$$\Delta_n^2 - \int f_0^2(x) r^2(x)\, dx = o_p(1). \tag{26}$$

Но (26) непосредственно следует из теоремы 2.1 [9] (см. также [2, 10]).

Докажем (25). Имеем

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

$$\leq a_n^{1/2} E\left|\int \left(f_n^*(x) - E f_n^*(x)\right) r(x)\, dx\right| + a_n^{1/2} \int \left|E f_n^*(x) - f_0(x)\right| r(x)\, dx = A_{1n} + A_{2n}.$$

Нетрудно убедиться, что

$$E\widehat{f}_i(x) = f_0(x) + O\left(\frac{1}{a_n}\right) + \alpha_n \int K(t)\varphi_i\left(\frac{x-\ell_i}{\gamma_n} - \frac{t}{a_n\gamma_n}\right) dt,$$

где $O(\,\cdot\,)$ равномерно по $x \in (-\infty, \infty)$. Поэтому

$$Ef_n^*(x) = f_0(x) + O\left(\frac{1}{a_n}\right) + \alpha_n \frac{1}{k}\sum_{j=1}^{p} k_j \int K(t)\varphi_j\left(\frac{x-\ell_j}{\gamma_n} - \frac{t}{a_n\gamma_n}\right) dt.$$

Следовательно,

$$A_{2n} \le c_{12}a_n^{-1/2} + c_{13}a_n^{1/2}\alpha_n\gamma_n.$$

Далее, имеем

$$A_{1n} \le a_n^{1/2}E^{1/2}\Biggl(\int \left(f_n^*(x) - Ef_n^*(x)\right) r(x)\, dx\Biggr)^2 \le$$

$$\le c_{14}a_n^{1/2}\max_{1\le j\le p}\left\{\frac{1}{n}\int f_j(u)\, du\left(\int K(t)r\left(u - \frac{t}{a_n}\right) dt\right)^2\right\}^{1/2} \le$$

$$\le c_{15}\left(\frac{a_n}{n}\right)^{1/2}.$$

Таким образом,

$$A_{1n} + A_{2n} \le c_{16}\Biggl(a_n^{-1/2} + \sqrt{a_n}\,\alpha_n\gamma_n + \left(\frac{a_n}{n}\right)^{1/2}\Biggr) \longrightarrow 0,$$

так как $\sqrt{a_n}\,\alpha_n\gamma_n \le a_n^2\alpha_n\gamma_n \to 0$ и $\dfrac{a_n}{n} \to 0$.

Из теоремы 2 вытекают два следствия.

***Следствие* 3.** *Случайная величина*

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

при гипотезе H_0 *распределена в пределе нормально с математическим ожиданием* 0 *и дисперсией* 1.

Этот результат позволяет построить асимптотический критерий проверки гипотезы H_0 : $f_1(x) = \ldots = f_p(x)$ (гипотезы однородности); критическая область устанавливается неравенством

$$T_n \ge \widetilde{d}_n(\alpha) = \mu_n + a_n^{-1/2}\sigma_n\lambda_\alpha, \tag{27}$$

где λ_α — квантиль уровня $1-\alpha$ стандартного нормального распределения $\Phi(x)$.

***Следствие* 4.** *При условиях теоремы* 2 *локальное поведение мощности* $P_{H_1}(T_n \ge \widetilde{d}_n(\alpha))$ *таково*:

$$P_{H_1}(T_n \ge \widetilde{d}_n(\alpha)) \longrightarrow 1 - \Phi\left(\lambda_\alpha - A(\varphi)\sigma^{-1}\right).$$

***Замечание* 1.** При альтернативе H_1 имеем

$$F_i(x) = F_0(x) + \alpha_n \gamma_n U_i\left(\frac{x-\ell_i}{\gamma_n}\right), \qquad U_i(u) = \int\limits_{-\infty}^{u} \varphi_i(x)\,dx,$$

и по условию теоремы 1 $\alpha_n\gamma_n = o\left(\frac{1}{\sqrt{n}}\right)$. Поэтому можно записать

$$\sup_x \left|F_i(x) - F_0(x)\right| = o\left(\frac{1}{\sqrt{n}}\right). \tag{28}$$

Известно, что критерии, основанные на отклонении между выборочными функциями распределения, как, например, критерии типа Колмогорова – Смирнова и критерии Крамера – Мизеса – Смирнова (аналог таких критериев при $p \geq 2$ был построен Кифером [11]), отличают близкие альтернативы от нулевой гипотезы, если $F_i(x) - F_0(x) = O\left(\frac{1}{\sqrt{n}}\right)$ равномерно по $x \in (-\infty, \infty)$, а в случае (28) перечисленные критерии не могут асимптотически отличать такие гипотезы от основной (предельное значение мощности будет совпадать с предельным уровнем критерия). Однако тесты (24) и (27), основанные на оценках плотности распределения, в пределе более мощны (при гипотезе H_1), нежели тесты, основанные на выборочных функциях распределения (аналогичные вопросы для одной выборки рассмотрены в работе Розенблатта [1]).

***Замечание* 2.** Критерии (24) и (27) проверки гипотез H_0' и H_0 соответственно при альтернативе H_1 являются асимптотически строго несмещенными, ибо $A(\varphi) > 0$ и равно 0 тогда и только тогда, когда $\varphi_i(x) = 0$ почти всюду, $i = 1, \ldots, p$.

***Замечание* 3 (примеры функции $\varphi(x)$).** Пусть $\varphi(x)$ — финитная функция с носителем $[0, A]$, $\int \varphi(x)\,dx = 0$, удовлетворяет условиям (iii) и $f_0(x) > 0$, $x \in (-\infty, \infty)$, тогда

$$f(x) = f_0(x) + \alpha_n \varphi\left(\frac{x-\ell}{\gamma_n}\right) \geq 0, \quad \gamma_n \downarrow 0,$$

при подходящем выборе α_n. Действительно, $f(x)$ совпадает с $f_0(x)$ всюду, кроме, возможно, интервала $I = [\ell, \ell + A\gamma_1]$. Пусть $\min_{x\in I} f_0(x) \geq \mu > 0$. Обозначим через L минимальное значение $\varphi(x)$ в этом же интервале. Если мы выберем α_n так, что $\mu + L\alpha_n > 0$, то будем иметь $f(x) \geq 0$.

Приведем конкретный пример. Пусть

$$f_0(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}},$$

$$\psi_1(x) = \begin{cases} \exp\Big\{-\dfrac{1}{1-(x-0,5)^2}\Big\}, & 0 < x < 1, \\ 0, & x \leq 0, \qquad x \geq 1, \end{cases}$$

$$\psi_2(x) = \begin{cases} -\exp\Big\{-\dfrac{1}{1-4(x-1,5)^2}\Big\}, & 1 < x < 2, \\ 0, & x \leq 1, \qquad x \geq 2, \end{cases}$$

и

$$\varphi(x) = \psi_1(x) + \psi_2(x).$$

Пусть, далее, $\ell = 0$, $\gamma_n = n^{-\frac{1}{5}}$. Тогда $A = 2$, $\min\limits_x \varphi(x) = -e^{-1}$ и $\mu = \dfrac{1}{e^2\sqrt{2\pi}}$. В качестве α_n выберем последовательность положительных чисел, сходящихся к нулю, и $\alpha_n < \dfrac{1}{e\sqrt{2\pi}}$. Например, $\alpha_n = 10^{-1} n^{-\frac{27}{80}}$ (см. условия теоремы 1 относительно a_n, α_n и γ_n).

Литература

1. *Rosenblatt M.* A quadratic measure of deviation of two-dimensional density estimates and a test of independence // Ann. Statist. – 1975. – **3**. – P. 1 – 14.
2. *Nadaraya E. A.* Nonparametric estimation of probability densities and regression curves // Math. and its Appl. (Soviet Series). – Dordrecht: Kluwer Acad. Publ. Group, 1989. – **20**.
3. *Nadaraya E. A.* Limit distribution of the quadratic deviation of two nonparametric estimators of the density of a distribution (in Russian) // Soobshch. Akad. Nauk Gruz.SSR. – 1975. – **78**. – P. 25 – 28.
4. *Anderson N. H., Hall P., Titterington D. M.* Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates // J. Multivar. Anal. – 1994. – **50**, № 1. – P. 41 – 54.
5. *Nikol'skii S. M.* Approximation of functions of several variables and imbedding theorems (in Russian). – Moscow: Nauka,1969.
6. *Колмогоров А. Н., Фомин С. В.* Элементы теории функций и функционального анализа. – М.: Наука, 1976.
7. *Hall P.* Limit theorems for stochastic measures of the accuracy of density estimators // Stochast. Process. and Appl. – 1982. – **13**, №. 1. – P. 11 – 25.
8. *Bickel P. J., Rosenblatt M.* On some global measures of the deviations of density function estimates // Ann. Statist. – 1973. – **1**. – P. 1071 – 1095.
9. *Bhattacharyya G. K., Roussas G. G.,* Estimation of a certain functional of a probability density function // Skand. Aktuarietidskr. – 1969. – **1969**. – P. 201 – 206.
10. *Mason D. M., Nadaraya E. A., Sokhadze G. A.* Integral functionals of the density // Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: a Festschrift in honor of Professor Jana Jurečková. – Beachwood, OH: Inst. Math. Statist., 2010. – P. 153–168.
11. *Kiefer J.* K-sample analogues of the Kolmogorov – Smirnov and Cramér – V. Mises tests // Ann. Math. Statist. – 1959. – **30**. – P. 420 – 447.

Получено 03.12.15