УДК 519.21

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

ОБ ОДНОМ КРИТЕРИИ ОДНОРОДНОСТИ, ОСНОВАННОМ НА ЯДЕРНЫХ ОЦЕНКАХ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ

We construct a homogeneity test based on the kernel-type estimators of the distribution density and investigate its consistency.

Побудовано критерій однорідності, що ґрунтується на оцінках щільності розподілу типу ядра, і досліджено його обґрунтованість.

1. 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 $(p \geq 2)$ генеральных совокупностей с плотностями распределения $f_1(x), \ldots, f_p(x)$. Требуется, основываясь на выборках $X^{(i)}$, $i = 1, \ldots, p$, проверить гипотезу однородности

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

В (1) общую плотность распределения будем обозначать через $f_0(x)$, т. е. $H_0: f_0(x) = f_1(x) = \ldots = f_p(x)$. Гипотеза однородности утверждает лишь, что плотности распределения $f_i(x)$, $i = 1, \ldots, p$, совпадают. Однако она не фиксирует вид этой общей плотности распределения $f_0(x)$.

Мы рассматриваем критерий проверки гипотезы H_0, основанный на статистике

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

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 $p = 2$ рассмотрен в работах [1, 2]. В этом случае статистика T принимает наиболее наглядный вид

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

В настоящей работе найдено предельное распределение (теорема 1) статистики $T_{n_1 n_2}$ при справедливости гипотезы H_0 и построен критерий проверки этой гипотезы. Доказано (теорема 2) важное свойство построенного критерия — состоятельность, причем доказательство существенно основано на идее доказательства теоремы 1. Следует отметить, что по схеме доказательства теоремы 1 имеется пересечение с теоремой 1 работы [3]. На основе этой теоремы в работе [3] проверяются гипотезы однородности и согласия против последовательности некоторых „близких” к H_0 альтернатив [4, 5] и исследуются асимптотики мощностей построенных критериев.

2. В данном пункте найдено предельное распределение статистики (2) при гипотезе H_0 в случае, когда n_i неограниченно возрастает, 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 $K(x),\ f_0(x)$ и $r(x)$:

(i) $K(x) \geq 0$, равна нулю вне конечного интервала $(-A, A)$ и вместе со своей производной непрерывна на этом интервале; или же абсолютно непрерывна на $(-\infty, \infty),\ x^2K(x)$ интегрируема и $K^{(1)}(x) \in L_1(-\infty, \infty)$, причем в обоих случаях $\displaystyle\int K(x)\,dx = 1$;

(ii) плотность распределения $f_0(x)$ ограничена и имеет ограниченную производную;

(iii) весовая функция $r(x)$ кусочно-непрерывна, ограничена и интегрируема.

Введем обозначения

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

$$\mu_n^* = \int f_n^*(x) r(x)\,dx, \qquad \overline{k} = k_1 + \ldots + k_p, \tag{3}$$

$$\Delta_n^2 = \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \Delta_{jn}^2, \qquad \Delta_{jn}^2 = \int \widehat{f}_j^2(x) r^2(x)\,dx.$$

Справедлива следующая теорема.

Теорема 1. *Пусть выполняются условия* (i) – (iii). *Если* $n^{-1}a_n^2 \to 0$, *то случайная величина* $a_n^{1/2}(T_n - \mu_n)\sigma_n^{-1}$ *при гипотезе* H_0 *асимптотически нормальна с математическим ожиданием* 0 *и дисперсией* 1, *где*

$$\mu_n = (p-1)R(K)\mu_n^*, \quad \sigma_n^2 = 2(p-1)R(K_0)\Delta_n^2,$$

$$K_0 = K * \tilde{K}, \qquad \tilde{K}(x) = K(-x), \quad R(g) = \int g^2(x)\,dx, \qquad p \geq 2.$$

Доказательство. Представим T_n в виде

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

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

Здесь и далее в этом пункте $\mathrm{E}(\,\cdot\,)$ — математическое ожидание относительно гипотезы H_0.

Выполняя простые преобразования, получаем

$$T_n = \int \Bigg[\sum_{i=1}^{p} \left(\sqrt{\frac{n_i}{a_n}} \, \left(\widehat{f}_i(x) - \mathrm{E}\widehat{f}_i(x) \right) \right)^2 -$$

$$- \left(\sum_{j=1}^{p} \alpha_j \sqrt{\frac{n_i}{a_n}} \, \left(\widehat{f}_j(x) - \mathrm{E}\widehat{f}_j(x) \right) \right)^2 \Bigg] r(x)\, dx,$$

где

$$\alpha_j^2 = \frac{k_j}{k_1 + \ldots + k_p} \, .$$

Пусть

$$Z(x) = (Z_1(x), \ldots, Z_p(x))$$

— вектор с компонентами

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

Тогда

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

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

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

$$= \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^{p} c_{ij} Z_j(x) \right)^2 r(x)\, dx. \tag{4}$$

Пусть $F_0(x)$ — функция распределения случайной величины $X_1^{(i)}$ (при гипотезе H_0) и $\widehat{F}_{n_i}(x)$ — эмпирическая функция распределения выборки $X^{(i)} = (X_1^{(i)}, \ldots, X_{n_i}^{(i)})$.

Используя теорему 3 работы [6] (см. также [5, 7]), можем записать

$$\widehat{F}_{n_i}(x) - F_i(x) = n_i^{-1/2} W_i^0(F_0(x)) + \varepsilon_{n_i}^{(1)}(x), \tag{5}$$

причем

$$\sup_{-\infty<x<\infty} |\varepsilon^{(1)}_{n_i}(x)| = O_p\left(\frac{\ln n_i}{n_i}\right) = O_p\left(\frac{\ln n}{n}\right), \qquad i=\overline{1,p},$$

$W^0_i(t),\ t\in[0,1],\ i=1,\ldots,p,$ — независимые броуновские мосты, зависящие соответственно только от $X^{(i)}$.

Используя (5) и условие (i), можно установить [5, 7, 8], что

$$Z_i(x)=\sqrt{\frac{n_i}{a_n}}\,\left(\widehat{f}_i(x)-\mathrm{E}\widehat{f}_i(x)\right)=\xi_i(x)+\varepsilon^{(2)}_{n_i}(x), \tag{6}$$

где

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

$$\sup_{-\infty<x<\infty} |\varepsilon^{(2)}_{n_i}(x)| = O_p\left(\frac{\ln n}{\sqrt{na_n^{-1}}}\right), \qquad i=\overline{1,p}.$$

Тогда в силу (6) имеем

$$\sum_{j=1}^{p} c_{ij}Z_j(x)=\sum_{j=1}^{p} c_{ij}\xi_j(x)+\varepsilon^{(3)}_{n_i}(x), \tag{7}$$

причем

$$\sup_{-\infty<x<\infty} |\varepsilon^{(3)}_{n_i}(x)| = O_p\left(\frac{\ln n}{\sqrt{na_n^{-1}}}\right), \qquad i=\overline{1,p}.$$

Процессы $\xi_j(x)$ можно представить так:

$$\xi_j(x)=a_n^{1/2}\int\left[K(a_n(x-t))-\int K(a_n(x-u))\,dF_0(u)\right]\,dW_j(F_0(t)),$$

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

Легко можно показать, что

$$\mathrm{E}\xi_j^2(x)=a_n\left[\int K^2(a_n(x-u))f_0(u)\,du-\left(\int K(a_n(x-u))f_0(u)\,du\right)^2\right]\le$$

$$\le a_n\int K^2(a_n(x-u))f_0(u)\,du\le c_1\int K^2(u)\,du,$$

$$c_1=\sup_{-\infty<x<\infty} f_0(x).$$

Следовательно,

$$\sup_{-\infty<x<\infty}\mathrm{E}\xi_j^2(x)\le c_2, \qquad i=\overline{1,p}. \tag{8}$$

(Напомним следующую терминологию: случайную величину X_n, зависящую от n, называют равномерно ограниченной, если $\mathrm{P}(|X_n|>M)\longrightarrow 0$ при $M\longrightarrow\infty$ равномерно по n.)

Следующая элементарная лемма является следствием известных свойств сходимости по вероятности.

Лемма 1. *Если случайная величина Y_n сходится по вероятности к нулю, а случайная величина X_n равномерно ограничена, то $X_n \cdot Y_n \xrightarrow{\mathrm{P}} 0$ при $n \longrightarrow \infty$.*

Обозначим

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

Тогда из (4) и (7) имеем

$$|T_n - T_n^{(1)}| \leq 2 \sum_{i=1}^{p-1} \left[\int \sum_{j=1}^{p} |c_{ij}| |\xi_j(x)| \sum_{r=1}^{p} |c_{ir}| |\varepsilon_{n_r}^{(3)}(x)| r(x)\, dx \right] +$$

$$+ \sum_{i=1}^{p-1} \int \left(\sum_{j=1}^{p} |c_{ij}| |\varepsilon_j^{(3)}(x)| \right)^2 r(x)\, dx = A_n^{(1)} + A_n^{(2)}.$$

Но

$$\sqrt{a_n} A_n^{(1)} \leq c_3 \sqrt{a_n} \max_{1 \leq i \leq p} \sup_{-\infty < x < \infty} |\varepsilon_{n_i}^{(3)}(x)|\, A_n = O_p \left(\frac{a_n \ln n}{\sqrt{n}} \right) A_n,$$

где

$$A_n = \sum_{i=1}^{p-1} \left[\int \sum_{j=1}^{p} |c_{ij}| |\xi_j(x)| r(x)\, dx \right].$$

Принимая во внимание (8), интегрируемость $r(x)$ и неравенство Чебышева, получаем, что случайная величина A_n равномерно ограничена и, следовательно, в силу леммы 1

$$O_p \left(\frac{a_n \ln n}{\sqrt{n}} \right) \cdot A_n \xrightarrow{\mathrm{P}} 0 \quad \text{при} \quad n \longrightarrow \infty,$$

так как по условию $n^{-1} a_n^2 \longrightarrow 0$.

Далее, используя интегрируемость $r(x)$ и

$$\max_{1 \leq i \leq p} \sup_{-\infty < x < \infty} |\varepsilon_{n_i}^{(3)}(x)| = O_p \left(\frac{\sqrt{a_n} \ln n}{\sqrt{n}} \right),$$

получаем

$$\sqrt{a_n} A_n^{(2)} = O_p \left(\frac{a_n^{3/2} \ln^2 n}{n} \right).$$

Следовательно,

$$\sqrt{a_n} \left(T_n - T_n^{(1)} \right) = o_p(1). \tag{9}$$

Перейдем к изучению предельного распределения функционала $T_n^{(1)}$.

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

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

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

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

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

Однако

$$\mathrm{E}\xi_j(t_1)\xi_j(t_2) = a_n \int K(a_n(t_1 - u))K(a_n(t_2 - u)) f_0(u)\, du + O(a_n^{-1}) =$$
$$= f_0(t_1) K_0(a_n(t_2 - t_1)) + O(a_n^{-1}), \tag{10}$$

причем оценка $O(\,\cdot\,)$ равномерна по $t_1,\ t_2$ и $K_0 = K * \widetilde{K},\ \widetilde{K} = K(-x)$.

Из (10) следует, что

$$R_n^{(i)}(t_1, t_2) = f_0(t_1) K_0\left(a_n(t_2 - t_1)\right) + O(a_n^{-1}). \tag{11}$$

Семиинвариант $\chi_n^{(i)}(s)$ порядка s случайной величины $U_n^{(i)}$ задается формулой [9]

$$\chi_n^{(i)}(s) = (s-1)! \cdot 2^{s-1} \int \ldots \int R_n^{(i)}(x_1, x_2) R_n^{(i)}(x_2, x_3) \ldots R_n^{(i)}(x_s, x_1) \times$$
$$\times r(x_1) r(x_2) \ldots r(x_s)\, dx_1\, dx_2 \ldots dx_s. \tag{12}$$

Из (11) и (12) нетрудно устанавливается, что

$$\begin{aligned} \mathrm{E}U_n^{(i)} &= \chi_n^{(i)}(1) = R(K) \int f_0(x) r(x)\, dx + O(a_n^{-1}), \\ \operatorname{Var} U_n^{(i)} &= \chi_n^{(i)}(2) = 2R(K_0) a_n^{-1} \int f_0^2(x) r^2(x)\, dx + o(a_n^{-1}), \end{aligned} \tag{13}$$

и s-й семиинвариант $\chi_n^{(i)}(s)$ с точностью до членов высшего порядка малости равен [9]

$$(s-1)!\, 2^{s-1} (a_n^{-1})^{s-1} [K * \widetilde{K}]^{(s)}(0) \int f_0^s(x) r^s(x)\, dx, \tag{14}$$

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

Из соотношений (13), (14) следует [5, 9], что

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

распределена асимптотически нормально с математическим ожиданием 0 и дисперсией $2R(K_0)\displaystyle\int f_0^2(u)r^2(u)\,du$ и, следовательно, $\sqrt{a_n}\,(T_n^{(1)}-\mu)$ асимптотически нормальна с математическим ожиданием 0 и дисперсией σ^2, где

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

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

$$R(g)=\int g^2(x)\,dx.$$

Далее, принимая во внимание (9) и представление

$$\sqrt{a_n}(T_n-\mu)=\sqrt{a_n}(T_n^{(1)}-\mu)+o_p(1),$$

заключаем, что $a_n^{1/2}(T_n-\mu)$ распределена в пределе нормально с математическим ожиданием 0 и дисперсией σ^2.

Очевидно,

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

Поэтому для завершения доказательства теоремы достаточно показать, что

$$a_n^{1/2}\left(\mu_n^*-\int f_0(x)r(x)\,dx\right)=o_p(1), \tag{15}$$

$$\Delta_n^2-\int f_0^2(x)r^2(x)\,dx=o_p(1), \tag{16}$$

где μ_n^* и Δ_n^2 определены в (3). Но (16) непосредственно следует из теоремы 2.1 [10] (см. также [5, 11]).

Докажем (15). Имеем

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

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

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

Применяя формулу Тейлора, находим

$$\mathrm{E}f_n^*(x)=\frac{1}{k}\sum_{j=1}^{p}k_jf_0(x)+O(a_n^{-1})=f_0(x)+O(a_n^{-1}).$$

Отсюда в силу интегрируемости $r(x)$ получаем

$$A_{2n}\le c_4a_n^{-1/2}.$$

Далее, имеем

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

$$\le c_5 a_n^{1/2} \left\{ \frac{a_n}{n} \int f_0(u)\, du \int K^2(t) r \left(u + \frac{t}{a_n} \right) dt \right\}^{1/2} \le c_6 \frac{a_n}{\sqrt{n}}.$$

Следовательно,

$$A_{1n} + A_{2n} \le c_7 \left(a_n^{-1/2} + \frac{a_n}{\sqrt{n}} \right) \longrightarrow 0.$$

Теорема 1 доказана.

Теорема 1 позволяет построить асимптотический критерий уровня α проверки гипотезы H_0, согласно которой $f_1(x) = f_2(x) = \ldots = f_p(x)$. Критическая область для проверки этой гипотезы устанавливается неравенством

$$T_n \ge d_n(\alpha), \tag{17}$$

где

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

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

3. Теперь исследуем асимптотическое поведение критерия (17), т. е. поведение функции мощности при $n \to \infty$.

Справедлива следующая теорема.

Теорема 2. *Пусть для $K(x)$, $f_i(x)$, $i = 1, \ldots, p$, и $r(x)$ выполняются условия* (i), (ii) *и* (iii) *соответственно. Если $n^{-1} a_n^2 \to 0$, то*

$$\mathbb{P}_{H_1}(T_n \ge d_n(\alpha)) \longrightarrow 1 \quad \text{при} \quad n \to \infty,$$

т. е. критерий, определенный в (17), *состоятелен против любой альтернативы $H_1 : f_i(x) \ne f_j(x)$ при некоторых $i \ne j$, $i, j = 1, \ldots, p$, на множестве положительной меры Лебега.*

Доказательство. Представим T_n в виде суммы [3]

$$T_n = T_n^{(2)} + A_{1n} + A_{2n}, \tag{18}$$

где

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

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

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

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

Далее, в силу формулы Тейлора можно показать, что

$$\mathrm{E}\widehat{f}_i(x) = f_i(x) + O(a_n^{-1})$$

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

$$A_{2n} = \frac{n}{a_n}\, \Delta + O\left(\frac{n}{a_n^2}\right), \tag{19}$$

где

$$\Delta = \sum_{i=1}^{p} k_i \int \left[f_i(x) - \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j f_j(x) \right]^2 r(x)\, dx.$$

Далее,

$$\mathrm{E}|A_{1n}| \leq (\mathrm{E}A_{1n}^2)^{1/2} \leq$$

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

где

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

Нетрудно видеть, что

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

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

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

Поэтому

$$\mathrm{E}|A_{1n}| \leq c_8 \sqrt{n}\, \times$$

$$\times \left\{ \sum_{i=1}^{p} k_i \int f_i(u)\, du \left[\int K(a_n(x-u)) A_i(x) r(x)\, dx \right]^2 \right\}^{1/2}. \tag{20}$$

Поскольку $\sup\limits_{x} |A_i(x)| \leq c_9$ и $r(x)$ ограничена, из (20) получаем

$$\mathrm{E}|A_{1n}| \le c_{10}\sqrt{\frac{n}{a_n}}\,.$$

Значит,

$$A_{1n} = O_p\left(\sqrt{\frac{n}{a_n}}\right). \tag{21}$$

Таким образом, из (18) – (20) следует, что

$$T_n = T_n^{(2)} + \frac{n}{a_n}\,\Delta + O\left(\frac{n}{a_n^2}\right) + O_p\left(\sqrt{\frac{n}{a_n}}\right). \tag{22}$$

Далее, рассуждая так же, как и при выводе (4), (6) и (9), получаем

$$\begin{gathered} T_n^{(2)} = T^{(3)} + o_p\left(\frac{1}{\sqrt{a_n}}\right) + O_p\left(\sqrt{\frac{a_n \ln^2 n}{n}}\right), \\ T_n^{(3)} = \sum_{i=1}^{p-1}\int\left[\sum_{j=1}^{p} c_{ij}\xi_j(x)\right]^2 r(x)\,dx, \end{gathered} \tag{23}$$

$$\xi_j(x) = a_n^{1/2}\int\left[K(a_n(x-t)) - \int K(a_n(x-u))\,dF_j(u)\right]\,dW_j(F_j(t)), \tag{24}$$

где $F_j(x)$ — функция распределения с плотностью $f_i(x)$ при альтернативе H_1 и $W_j(t)$, $j = 1,\ldots,p,$ — независимые стандартные винеровские процессы на $[0,1]$.

Итак, из (18), (19), (21) – (23) следует, что

$$\mathbb{P}_{H_1}\{T_n \ge d_n(\alpha)\} = \mathbb{P}_{H_1}\left\{T_n^{(3)} \ge -\frac{n}{a_n}\,\Delta + R_n\right\}, \tag{25}$$

где

$$R_n = o_p(a_n^{-1/2}) + O(na_n^{-2}) + O_p\left(\frac{n}{\sqrt{a_n}}\right) + O_p\left(\frac{a_n\ln^2 n}{n}\right) + \mu_n + \lambda_\alpha\sigma_n a_n^{-1/2}. \tag{26}$$

Вероятность выполнения неравенства $T_n^{(3)} \ge -\dfrac{n}{a_n}\,\Delta + R_n$ не меньше, чем вероятность выполнения неравенства $V_n^{(i)} \ge -\dfrac{n}{a_n}\,\Delta + R_n$ при любом фиксированном i, $i = 1,\ldots,p-1$, например $i = 1$, где $V_n^{(i)}$ — слагаемые суммы $T_n^{(3)}$, т. е.

$$\mathbb{P}\left\{T_n^{(3)} \ge -\frac{n}{a_n}\,\Delta + R_n\right\} \ge \mathbb{P}\left\{V_n^{(1)} \ge -\frac{n}{a_n}\,\Delta + R_n\right\}. \tag{27}$$

Здесь

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

Теперь установим предельное распределение функционала $V_n^{(1)}$. Ковариационная функция $\overline{R}_n^{(1)}(t_1,t_2)$ гауссовского процесса $\sum_{j=1}^{p} c_{1j}\xi_j(t)$ (см. (10)) имеет вид

$$\overline{R}_n^{(1)}(t_1,t_2)=\varphi(t_1)K_0(a_n(t_2-t_1))+O\left(\frac{1}{a_n}\right), \tag{28}$$

где

$$\varphi(t)=\sum_{j=1}^{p} c_{1j}^2 f_j(t),$$

оценка $O(\,\cdot\,)$ равномерна по t_1, t_2 и $K_0=K*\widetilde{K}(x)$, $\widetilde{K}(x)=K(-x)$.

Семиинвариант $\overline{\chi}_n^{(1)}(s)$ порядка s случайной величины $V_n^{(1)}$ задается формулой [9]

$$\overline{\chi}_n^{(1)}(s)=(s-1)!\cdot 2^{s-1}\int\ldots\int \overline{R}_n^{(1)}(t_1,t_2)\overline{R}_n^{(1)}(t_2,t_3)\ldots\overline{R}_n^{(1)}(t_s,t_1)\times$$

$$\times r(t_1)r(t_2)\ldots r(t_s)\,dt_1\,dt_2\ldots dt_s. \tag{29}$$

Из (28) и (29) можно установить, что

$$\mathrm{E}V_n^{(1)}=a_0+O\left(a_n^{-1}\right),\quad \mathrm{Var}\,V_n^{(1)}=a_n^{-1}b_0^2+o\left(a_n^{-1}\right), \tag{30}$$

и k-й семиинвариант $\overline{\chi}_n(s)$ с точностью до членов высшего порядка малости равен [9]

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

где

$$a_0=\int\varphi(x)r(x)\,dx\cdot R(K),$$

$$b_0^2=\int K_0^2(x)\,dx\int\varphi^2(x)r^2(x)\,dx.$$

Из соотношений (29), (30) следует, что [9]

$$Z_n=a_n^{1/2}(V_n^{(1)}-a_0)b_0^{-1}$$

распределена асимптотически нормально с математическим ожиданием 0 и дисперсией 1.

Далее, принимая во внимание (25) и (27), получаем неравенство

$$\mathbb{P}_{H_1}\{T_n\geq d_n(\alpha)\}\geq\mathbb{P}_{H_1}\left\{Z_n\geq-\frac{n}{b_0\sqrt{a_n}}\,[\Delta+R_n^*]\right\},$$

где

$$R_n^*=\frac{a_n}{n}\,(R_n-a_0)=o_p\left(\frac{\sqrt{a_n}}{n}\right)+O\left(\frac{1}{a_n}\right)+O\left(\frac{\sqrt{a_n}}{\sqrt{n}}\right)+O_p\left(\frac{a_n^2\ln^2 n}{n^2}\right)+$$

$$+\frac{a_n}{n}\,(\mu_n-a_0)+\lambda_\alpha\sigma_n\,\frac{\sqrt{a_n}}{n}\,.$$

По аналогии с доказательством (15) и (16) $\dfrac{a_n}{n}(\mu_n-a_0)\to 0$ и

$$\sigma_n^2\to 2(p-1)R(K_0)(\overline{k})^{-1}\sum_{j=1}^{p}k_j\int f_j^2(x)r^2(x)\,dx,$$

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

По вероятности при гипотезе H_1 и по условию $n^{-1}a_n^2 \to 0$ находим, что $R_n^* = o_p(1)$.

Наконец, поскольку Z_n распределена при гипотезе H_1 асимптотически нормально, $na_n^{-1/2} \to \to \infty$ и $\Delta > 0$, то мощность $\mathbb{P}_{H_1}\{T_n \geq d_n(\alpha)\} \to 1$ и, следовательно, критерий (17) состоятелен.

Теорема 2 доказана.

Литература

1. *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. Multivariate Anal. – 1994. – **50**, № 1. – P. 41 – 54.
2. *Nadaraya E. A.* Limit distribution of the quadratic deviation of two nonparametric estimators of the density of a distribution // Soobshch. Akad. Nauk Gruz.SSR. – 1975. – **78**. – P. 25 – 28 (in Russian).
3. *Бабилуа П. К., Надарая Э. А., Сохадзе Г. А.* Проверка гипотез о равенстве плотностей распределения // Укр. мат. журн. – 2016. – **68**, № 5. – P. 586 – 600.
4. *Rosenblatt M.* A quadratic measure of deviation of two-dimensional density estimates and test independence // Ann. Statist. – 1975. – **3**. – P. 1 – 14.
5. *Nadaraya E. A.* Nonparametric estimation of probability densities and regression curves // Math. and Appl. (Soviet Ser.). – Dordrecht: Kluwer Acad. Publ. Group, 1989. – **20**.
6. *Komlós J., Major P., Tusnády G.* An approximation of partial sums of independent RV's and the sample DF. I // Z. Wahrscheinlichkeitstheor. und verw. Geb. –1975. – **32**. – S. 111 – 131.
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. *Prakasa Rao B. L. S.* Nonparametric functional estimation. – London: Acad. Press, 1983.
9. *Bickel P. J., Rosenblatt M.* On some global measures of the deviations of density function estimates // Ann. Statist. – 1973. – **1**. – P. 1071 – 1095.
10. *Bhattacharyya G. K., Roussas G. G.* Estimation of a certain functional of a probability density function // Skand. Aktuarietidskr. – 1969. – **1969**. – P. 201 – 206.
11. *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.

Получено 27.07.18