Н.К.Бакиров

УТОЧНЕНИЕ АСИМПТОТИКИ МОЩНОСТИ КВАНТИЛЬНОГО КРИТЕРИЯ

I. Квантильный критерий, близкий по своим свойствам к критерию согласия Пирсона, был предложен Л.Н.Большевым [I] и изучался, в частности, в работах [2] , [3] . Хорошо известно, что статистика квантильного критерия Z_n^2 при нулевой гипотезе H_0 и возрастании числа наблюдений $n \to \infty$ имеет предельное χ^2 -квадрат распределение и имеет нецентральное χ^2 -квадрат распределение при $n \to \infty$ в случае, когда справедлива альтернатива $H_{1,n}$, сближающаяся с проверяемой гипотезой H_0 при $n \to \infty$ со скоростью $1/\sqrt{n}$. В настоящей работе эти результаты уточняются с помощью асимптотических представлений.

2. Пусть проверяется гипотеза H_0 , согласно которой независимые, одинаково распределенные наблюдения $X_1, X_2, \dots, X_n$ имеют непрерывную функцию распределения $F_0(x)$. Образуем из случайных величин X_i вариационный ряд $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Зафиксируем вектор положительных вероятностей $\bar{p} = (p_1, p_2, \dots, p_k)$, $k > 2$, такой, что $\sum_{i=1}^{k} p_i = 1$. Далее, рассмотрим выборочные квантили $\xi_i = X_{(m_i)}$, $m_i = [(n+1)u_i]$ ($u_i = p_1 + p_2 + \dots + p_i$, $i = 1, 2, \dots, k-1$, $[\cdot]$ – целая часть числа, $\xi_0 = -\infty$, $\xi_k = +\infty$) и определим величины

$$\Delta_i = \sqrt{\frac{n}{p_i}}\left(F_0(\xi_i) - F_0(\xi_{i-1}) - p_i\right), \quad \delta_i = \sqrt{\frac{n}{p_i}}\left(F_n(g_i) - F_n(g_{i-1}) - p_i\right),$$

здесь $F_n(x)$ – эмпирическая функция распределения, $g_i = F_0^{-1}(u_i)$, $g_0 = -\infty$, $g_k = +\infty$. Пусть $\vec{\Delta} = (\Delta_1, \Delta_2, \dots, \Delta_k)$, $\vec{\delta} = (\delta_1, \delta_2, \dots, \delta_k)$. Статистика квантильного критерия определяется равенством

$$Z_n^2 = \|\vec{\Delta}\|^2 = \sum_{i=1}^{k} \Delta_i^2 ,$$

здесь $\|\cdot\|$ евклидова норма в R^k . В наших обозначениях $\|\vec{\delta}\|^2$ есть статистика χ^2 – критерия согласия Пирсона; для нее в [4] было выведено следующее соотношение, справедливое при всех фиксированных $x > 0$,

$$P\{\|\vec{\delta}\|^2 \leq x\} - P\{\chi^2_{K-1} \leq x\} = O(n^{-\frac{K-1}{K}}), \qquad (I)$$

здесь χ^2_{K-1} – случайная величина, имеющая $\chi И$ -квадрат распределение с $K-1$ степенью свободы. В работе [5] соотношение (I) было уточнено, там же отмечалось, что (I) выражает сильное свойство статистики $\|\vec{\delta}\|^2$, поскольку ее распределение сходится к предельному при $n \to \infty$ со скоростью больше чем $1/\sqrt{n}$. Задача уточнения асимптотики функции распределения статистики $\|\vec{\delta}\|^2$, а также мощности χ^2-критерия согласия Пирсона наталкивается на определенные трудности [5], их удается избежать при рассмотрении квантильного критерия.

ТЕОРЕМА I. Для произвольного компакта $A \subset R^1$ равномерно по $x \in A$ выполнено асимптотическое соотношение:

$$P\{Z_n^2 \leq x\} = P\{\chi^2_{K-1} \leq x\} + \frac{1}{n}\left(V_1(x,\vec{p}) - \frac{d}{dx}P\{\chi^2_{K-1} \leq x\}\right) + O\left(\frac{1}{n^2}\right), \qquad (2)$$

где в случае равных вероятностей $\vec{p} = p_0 = \left(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K}\right)$

$$V_1(x,\vec{p}_0) = \frac{1-K^2}{12}P\{\chi^2_{K-1} \leq x\} + \frac{K}{2}\int_0^x z\,dP\{\chi^2_{K-1} \leq z\} -$$

$$-\frac{(K-1)^2}{4}\int_{-\infty}^{\infty} z^4 \phi(z) P\{\chi^2_{K-1} \leq x - z^2\}dz + \frac{(K-1)(K-2)}{18}\int_{-\infty}^{\infty} z^6 \phi(z)dz. \qquad (3)$$

$$\cdot P\{\chi^2_{K-1} \leq x - z^2\}dz - \frac{(K-1)(K-2)}{6}\iint_{-\infty}^{\infty} z^2 y^4 \phi(z)\phi(y) P\{\chi^2_{K-3} \leq x - y^2 - z^2\}dy\,dz,$$

здесь $\phi(z)$ плотность стандартной гауссовской величины. Для вероятностей $\vec{p} \neq \vec{p}_0$ выражение для $V_1(x,\vec{p})$ дается ниже в (7).

ДОКАЗАТЕЛЬСТВО. Предположим сначала, что $(1+n)u_i$ суть целые числа и рассмотрим статистику $\tilde{Z}_n^2 = \frac{n+1}{n} Z_n^2$. Заметим, что $F_0(X_{(i)}) = U_{(i)}$ суть порядковые статистики для равномерного распределения на отрезке $[0,1]$. Известно, [6], что вектор $(U_{(m_1)}, U_{(m_2)}, \ldots, U_{(m_{K-1})})$ имеет плотность распределения вида

$$f(x_1, x_2, \ldots, x_{K-1}) = n! \prod_{i=1}^{K}\left((x_i - x_{i-1})^{n_i - 1}/(n_i - 1)!\right)$$

$$n_i = (1+n)p_i\ , \quad 0 = x_0 \leqslant x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{k-1} \leqslant x_k = 1.$$

Следовательно,

$$P\{\tilde{Z}_n^2 \leqslant x\} = \int\limits_{\Omega_n}\!\!\ldots\!\int f(x_1, x_2, \ldots, x_{k-1})\,dx_1\,dx_2 \ldots dx_{k-1} =$$

$$= C_n \int\limits_{\Omega_n}\!\!\ldots\!\int \prod_{i=1}^{k} \left(1 + \frac{y_i - y_{i-1}}{\sqrt{n+1}\,p_i}\right)^{n_i - 1} dy_1\,dy_2 \ldots dy_{k-1}\ , \tag{4}$$

здесь использована замена переменных:

$$y_i = (x_i - u_i)\sqrt{n+1}\ , \quad i = 0, 1, \ldots, k, \ (u_0 = 0,\ u_k = 1),$$

$$C_n = \frac{n!}{(\sqrt{n+1})^{k-1}} \prod_{i=1}^{k} \frac{p_i^{n_i - 1}}{(n_i - 1)!}\ ,$$

$$\Omega_n = \left\{ \vec{x} \,\middle|\, (n+1)\sum_{i=1}^{k} \frac{(x_i - x_{i-1} - p_i)^2}{p_i} \leqslant x,\ 0 \leqslant x_1 \leqslant \ldots \leqslant x_{k-1} \leqslant 1 \right\},$$

$$\vec{x} = (x_1, x_2, \ldots, x_{k-1}).$$

Множество Ω_n, по которому ведется интегрирование в (4), в переменных $\vec{y} = (y_1, y_2, \ldots, y_{k-1})$ запишется так

$$\Omega_n = \left\{ \vec{y} \,\middle|\, \sum_{i=1}^{k} \frac{(y_i - y_{i-1})^2}{p_i} \leqslant x,\ u_i + \frac{y_i}{\sqrt{n+1}} \geqslant u_{i-1} + \frac{y_{i-1}}{\sqrt{n+1}},\ i = 1, 2, \ldots, k \right\} =$$

$$= \left\{ \vec{y} \,\middle|\, \sum_{i=1}^{k} \frac{(y_i - y_{i-1})^2}{p_i} \leqslant x \right\} = \Omega_x\ .$$

Второе равенство здесь, очевидно, выполнено, если $n+1 \geqslant x / \min\limits_i p_i$, т.о. при достаточно больших n $\Omega_n = \Omega_x$. Далее, поскольку $(y_i - y_{i-1})/\sqrt{n+1} \xrightarrow[n\to\infty]{} 0$ на Ω_x мы можем в (4) преобразовать подинтегральное выражение, используя разложение логарифма:

$$\prod_{i=1}^{k} \left(1 + \frac{y_i - y_{i-1}}{\sqrt{n+1}\,p_i}\right)^{n_i - 1} = \exp\left\{ -\frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - y_{i-1})^2}{p_i} + \sum_{s=1}^{\infty} G_s(\vec{y}) \right\} =$$

$$= \exp\left\{ -\frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - y_{i-1})^2}{p_i} \right\} \cdot \left(1 + \sum_{s=1}^{\infty} \mu_s(\vec{y})\right), \tag{5}$$

где

$$G_1(\vec{y})=\sum_{i=1}^{K}\frac{(y_i-y_{i-1})(n_i-1)}{\sqrt{n+1}\,p_i}=-\frac{1}{\sqrt{n+1}}\sum_{i=1}^{K}\frac{y_i-y_{i-1}}{p_i},$$

$$G_2(\vec{y})=\frac{1}{n+1}\sum_{i=1}^{K}\frac{(y_i-y_{i-1})^2}{2p_i^2},\quad G_s(\vec{y})=\frac{(-1)^{s+1}}{s}\sum_{i=1}^{K}\frac{(y_i-y_{i-1})^s(n_i-1)}{(\sqrt{n+1}\,p_i)^s},$$

$s=3,4,\dots$.

Величины $\mu_s(\vec{y})$ суть однородные многочлены переменных $\vec{y}$ степени s :

$$\mu_s(\vec{y})=\sum_{(\vec{i},\vec{\alpha})=s}\frac{1}{i_1!\,i_2!\dots}\left(G_{\alpha_1}\right)^{i_1}\left(G_{\alpha_2}\right)^{i_2}\dots,$$

здесь $\vec{i}=(i_1,i_2,\dots)$, $\vec{\alpha}=(\alpha_1,\alpha_2,\dots)$ – мультиндексы, $(\vec{i},\vec{\alpha})=\sum_{k=1}^{\infty}i_k\alpha_k$.

Заметим теперь, что множество Ω_x центрально симметрично относительно нуля в R^{K-1} , а многочлены μ_s при нечетном s суть нечетные функции переменной $\vec{y}$, следовательно, величины $\mu_s(\vec{y})$, $s=1,3,5,\dots$, при интегрировании дадут нулевой вклад в вероятность (4). Используя далее формулу Старлинга, легко получим разложение

$$C_n=\frac{1}{(\sqrt{2\pi})^{K-1}\sqrt{p_1p_2\cdots p_K}}\exp\left\{\sum_{j=1}^{\infty}\frac{B_{2j}}{2j(2j-1)(n+1)^{2j-1}}\left(1-\sum_{i=1}^{K}\frac{1}{p_i^{2j-1}}\right)\right\},\qquad(6)$$

где B_{2j} – числа Бернулли порядка $2j$. Формулы (4) – (6) позволяют теперь выписать асимптотическое разложение для вероятности $P\{\tilde{Z}_n^2\leq x\}$ по степеням $1/(n+1)$:

$$P\{\tilde{Z}_n^2\leq x\}=P\{\chi_{K-1}^2\leq x\}+\sum_{s=1}^{\infty}\frac{V_s(x,\vec{p})}{(n+1)^s}.$$

Первый член асимптотики дается формулой

$$V_1(x,\vec{p})=\int\limits_{\Omega_x}\!\!\dots\!\int p_0(\vec{y})\left\{\frac{1}{12}\left(1-\sum_{i=1}^{K}\frac{1}{p_i}\right)+\frac{1}{2}\sum_{i=1}^{K}z_i^2+\right.$$
$$\left.+\frac{1}{2}\left(\sum_{i=1}^{K}z_i-\frac{1}{3}\sum_{i=1}^{K}p_iz_i^3\right)^2-\frac{1}{4}\sum_{i=1}^{K}p_iz_i^4\right\}d\vec{y},\qquad(7)$$

здесь $z_i = (y_i - y_{i-1})/p_i$

$$p_0(\vec{y}) = \frac{1}{(\sqrt{2\pi})^{k-1}\sqrt{p_1 p_2 \cdots p_k}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\frac{(y_i - y_{i-1})^2}{p_i}\right\}.$$

Нетрудно подсчитать, что $V_1(\infty, \vec{p}) = 0$. Выражение для $V_1(x, \vec{p})$ в случае равных вероятностей $\vec{p} = \vec{p}_0$ можно преобразовать к виду (3), используя равнокоррелированность случайных величин $Z_i = (Y_i - Y_{i-1})/p_i$, где случайный вектор $(Y_1, Y_2, \ldots, Y_{k-1})$ распределен с плотностью $p_0(\vec{y})$, и находя совместную плотность распределения сл.в. Z_1, Z_2 и $\sum_{i=1}^{k} Z_i^2$.

Анализируя приведенные выше рассуждения, можно прийти к выводу, что для статистики квантильного критерия Z_n^2 имеет место асимптотическая формула (2),,заметим здесь, что случай, когда величины $(1+n)u_i$ не обязательно целые сводится к уже рассмотренному с $u_i' = \frac{[(1+n)u_i]}{1+n} = u_i + O\left(\frac{1}{n}\right)$. Теорема доказана.

3. Рассмотрим теперь последовательность альтернатив $H_{1,n}$ согласно которым наблюдения $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ имеют непрерывные функции распределения $F_{1,n}(x) = F_0(x) + F_1(x)/\sqrt{n+1}$. Предположим, что для некоторых ограниченных $A_k(x)$, $k=0,1$, $\forall i = 1, 2, \ldots, i-1$

$$F_1\left(F_0^{-1}(u_i + y)\right) = A_0(u_i) + y A_1(u_i) + O(y^2).$$

Из свойств функций распределения выводим $F_1(F_0^{-1}(0)) = F_1(F_0^{-1}(1)) = 0$, положим формально $A_k(0) = A_k(1) = 0$, $k = 0, 1$. Найдем теперь асимптотическое представление для мощности β_n критерия проверки H_0 против альтернативы $H_{1,n}$, отвергающего H_0 при достаточно больших Z_n^2, т.е. с критическим множеством вида $\{Z_n^2 \geqslant z_0\}$.

ТЕОРЕМА 2. Имеет место асимптотическая формула

$$\beta_n = P\left\{\chi_{k-1}^2(\rho) \leqslant x_\alpha\right\} + \frac{R(x_\alpha)}{\sqrt{n+1}} + O\left(\frac{1}{n}\right), \qquad (8)$$

где α – уровень значимости критерия, x_α квантиль уровня $1-\alpha$ для χ^2-квадрат распределения с $k-1$ степенью свободы, $\chi_{k-1}^2(\rho)$ – случайная величина, имеющая нецентральное χ^2-квадрат распределение с $k-1$ степенью свободы и параметром нецентральности $\rho = \sum_{i=1}^{k}\left(A_0(u_i) - A_0(u_{i-1})\right)^2/p_i$

$$R(x)=\int\limits_{\Omega_x}\!\!\ldots\!\int p_1(\vec{y})\left\{-\sum_{i=1}^{K}\frac{(z_i-z_{i-1})\left(A_1(u_i)y_i-A_1(u_{i-1})y_{i-1}\right)}{p_i}+\right.$$

$$\left.+\frac{1}{3}\sum_{i=1}^{K}\frac{(z_i-z_{i-1})^3}{p_i^2}-\sum_{i=1}^{K}\frac{z_i-z_{i-1}}{p_i}\right\}d\vec{y},$$

ГДЕ $z_i=y_i+A_0(u_i)$,

$$P_1(\vec{y})=\frac{1}{(\sqrt{2\pi})^{K-1}\sqrt{p_1p_2\cdots p_K}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{K}\frac{(z_i-z_{i-1})^2}{p_i}\right\}.$$

ДОКАЗАТЕЛЬСТВО. Пусть сначала $(1+n)p_i$ суть целые числа. Имеем

$$P_n\left\{\tilde{Z}_n^2\leqslant x\right\}=\int\limits_{V_n}\!\!\ldots\!\int f(x_1,x_2,\ldots,x_{K-1})d\vec{x},$$

где P_n распределение, соответствующее альтернативе $H_{1,n}$,

$$V_n=\left\{\vec{x}\,\middle|\,(n+1)\sum_{i=1}^{K}\frac{1}{p_i}\left(F_0\left(F_{1,n}^{-1}(x_i)\right)-F_0\left(F_{1,n}^{-1}(x_{i-1})\right)\right)^2\leqslant x\right\}$$

$$0=x_0\leqslant x_1\leqslant\ldots\leqslant x_{K-1}\leqslant x_K=1$$

Введем новые переменные $y_i=\sqrt{n+1}\left(F_0\left(F_{1,n}^{-1}(x_i)\right)-u_i\right)$, $i=0,1,\ldots,K$ и заметим что как и прежде $V_n=\Omega_x$ для достаточно больших n. Плотность $f(\vec{x})$ для $\vec{x}\in V_n$ перепишется в новых переменных

$$f(\vec{x})=C_n\prod_{i=1}^{K}\left(1+\frac{y_i-y_{i-1}}{p_i\sqrt{n+1}}+\sum_{s=0}^{1}\frac{A_s(u_i)y_i^s-A_s(u_{i-1})y_{i-1}^s}{p_i(\sqrt{n+1})^{s+1}}+O\left(\frac{1}{n^{3/2}}\right)\right)^{n_i-1}$$

Используя как и ранее разложение логарифма в ряд, а также разложение (6), можно получить, что $\forall x>0$

$$P_n\left\{\tilde{Z}_n^2\leqslant x\right\}=P\left\{\chi_{K-1}^2(\rho)\leqslant x\right\}+\frac{R(x)}{\sqrt{n+1}}+O\left(\frac{1}{n}\right).$$

Отсюда нетрудно вывести требуемое соотношение (8), используя представление (2) и проводя известные рассуждения [2]. В заключение отметим, что (I) справедливо и в случае, когда величины $(1+n)u_i$ не обязательно целые. Теорема доказана.

Утверждения теорем I, 2 останутся в силе, если в определении квантильного критерия вместо m_i взять $m_i'=m_i+O(1)$

4. В настоящем пункте оценивается в определенном смысле среднеквадратическая близость критериев Пирсона и квантильного. Из результатов работы [8] , в которой изучалась задача аппроксимации выборочных квантилей произвольного непрерывного распределения линейной функцией от эмпирической функции распределения, в частности, следует, что с вероятностью единица

$$\lim_{n\to\infty} \sup \|\vec{\Delta}+\vec{\delta}\|^2 \sqrt{\frac{n}{(\ln \ln n)^3}} = c_0 ,$$

где c_0 – некоторая конечная, положительная константа. В нашем специальном случае, ситуация сводится к рассмотрению выборочных характеристик равномерного на $[0,1]$ распределения.

ТЕОРЕМА. Справедливо представление:

$$E\|\vec{\Delta}+\vec{\delta}\|^2 = \frac{c}{\sqrt{n}} + \frac{\gamma_n}{n} ,$$

$$c = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{k} \frac{1}{p_i}\left(\sqrt{u_i(1-u_i)} + \sqrt{u_{i-1}(1-u_{i-1})}\right), \qquad (9)$$

здесь γ_n – ограниченная последовательность.

ДОКАЗАТЕЛЬСТВО. Предположим сначала, что np_i , $i=1,2,\dots,k$ суть целые числа. Заметим, что $F(X_{(i)})=U_{(i)}$ суть порядковые статистики для равномерного на $[0,1]$ распределения. Для удобства перепишем величины Δ_i , δ_i в следующем виде:

$$\Delta_i - E\Delta_i = \sqrt{\frac{n}{p_i}}\left(U_{(nu_i)} - \frac{nu_i}{1+n} - \left[U_{(nu_{i-1})} - \frac{nu_{i-1}}{1+n}\right]\right),$$

$$\delta_i = \sqrt{\frac{n}{p_i}}\ \frac{1}{n}\sum_{s=1}^{n}\left(\varepsilon_i(U_s) - p_i\right) = \frac{1}{\sqrt{np_i}}\sum_{s=1}^{n}\left(\varepsilon_i(U_{(s)}) - p_i\right),$$

здесь U_s , $s=1,2,\dots,n$ выборка из равномерного распределения на отрезке $[0,1]$, $\varepsilon_i(x)$ – характеристическая функция интервала (u_{i-1}, u_i) . Используя известные соотношения [6] , можно вывести, что $|E\Delta_i| \le \frac{1}{\sqrt{n+1}}$, $i=1,2,\dots,k$, т.е. нам достаточно показать, что

$$\sup_n \sqrt{n}\left|\sqrt{n}E\|\vec{\Delta}-E\vec{\Delta}+\vec{\delta}\|^2 - c\right| < +\infty .$$

Заметим сначала, что

$$E\|\vec{\delta}\|^2 = n\sum_{i=1}^{k}\frac{1}{p_i}E\left(F_n(g_i)-F_n(g_{i-1})-p_i\right)^2 = k-1 .$$

Используя известное соотношение для ковариации выборочных квантилей [6], также найдем

$$E\|\vec{\Delta}-E\vec{\Delta}\|^2 = n\sum_{i=1}^{K}\frac{1}{p_i}E\left(U_{(m_i)}-U_{(m_{i-1})}-\frac{np_i}{1+n}\right)^2 =$$

$$=\frac{n}{n+2}\sum_{i=1}^{K}\frac{1}{p_i}\frac{np_i}{1+n}\left(1-\frac{np_i}{1+n}\right)=K-1+O\left(\frac{1}{n}\right).$$

Найдем теперь ковариацию величин $\vec{\Delta}$ и $\vec{\delta}$, запишем сначала

$$E\left(\vec{\Delta}-E\vec{\Delta},\vec{\delta}\right)=\sum_{i=1}^{K}\frac{1}{p_i}\left(-\frac{n^2p_i^2}{1+n}+\right.$$

$$\left.+\sum_{s=1}^{n}E\,\varepsilon_i\left(U_{(s)}\right)\cdot\left(U_{(nu_i)}-U_{(nu_{i-1})}\right)\right).$$

Обозначим $nu_i=r$; нам нужно, в частности, найти величину

$$\sum_{s=1}^{n}E\,U_{(r)}\,\varepsilon_i\left(U_{(s)}\right)=\sum_{s=1}^{r-1}+E\,U_{(r)}\,\varepsilon_i\left(U_{(r)}\right)+\sum_{s=r+1}^{n}=J_1+J_2+J_3.$$

Используя выражение для совместной плотности распределения двух членов вариационного ряда для равномерного на [0,1] распределения,[6] и после несложных упрощений, найдем

$$J_1=\sum_{s=1}^{r-1}\frac{n!}{(s-1)!(r-s-1)!(n-r)!}\iint\limits_{x>y}x\,\varepsilon_i(y)\,y^{s-1}(x-y)^{r-s-1}(1-x)^{n-r}dx\,dy=$$

$$=\frac{n!}{(n-r)!(r-2)!}\int_0^1 x^{r-1}(1-x)^{n-r}\int_0^x\varepsilon_i(y)\,dy\,dx,$$

$$J_3=\sum_{s=r+1}^{n}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\iint\limits_{x<y}x\,\varepsilon_i(y)\,x^{r-1}(y-x)^{s-r-1}(1-y)^{n-s}dx\,dy=$$

$$=\frac{n!}{(r-1)!(n-r-1)!}\int_0^1 x^{r}(1-x)^{n-r-1}\int_x^1\varepsilon_i(y)\,dy\,dx,$$

$$J_2=\frac{n!}{(r-1)!(n-r)!}\int_0^1 x^{r}(1-x)^{n-r}\,\varepsilon_i(x)\,dx.$$

Ниже исследуется асимптотика интегралов $\mathcal{J}_2$ и $\mathcal{J}_1+\mathcal{J}_3$. Сначала рассмотрим величину $\mathcal{J}_2$. Используя формулу Стирлинга и замену переменных $x=u_i-z/\sqrt{n}$, получим:

$$\mathcal{J}_2=\frac{u_i}{\sqrt{2\pi u_i(1-u_i)}}\int\limits_0^{p_i\sqrt{n}}\left(1-\frac{z}{u_i\sqrt{n}}\right)^{nu_i}\left(1+\frac{z}{\sqrt{n}(1-u_i)}\right)^{(1-u_i)n}dz\,\cdot$$

$$\cdot\left(1+O\left(\frac{1}{n}\right)\right)=\int\limits_0^{R_n}+\int\limits_{R_n}^{p_i\sqrt{n}}.$$

Выберем $R_n=n^{0.01}$, тогда в силу монотонности по z подинтегральной функции

$$\int\limits_{R_n}^{p_i\sqrt{n}}\leqslant \mathrm{const}\cdot\sqrt{n}\left(1-\frac{R_n}{u_i\sqrt{n}}\right)^{nu_i}\left(1+\frac{R_n}{(1-u_i)\sqrt{n}}\right)^{(1-u_i)n}=O\left(\frac{1}{n}\right).$$

Используя разложение логарифма в окрестности единицы, также получим

$$\int\limits_0^{R_n}=\frac{u_i}{\sqrt{2\pi u_i(1-u_i)}}\int\limits_0^{R_n}\exp\left\{-\frac{z^2}{2u_i(1-u_i)}-\frac{z^3}{3\sqrt{n}}\left(\frac{1}{u_i^2}-\frac{1}{(1-u_i)^2}\right)-\right.$$

$$\left.-\frac{z^4}{4n}\left(\frac{1}{u_i^3}+\frac{1}{(1-u_i)^3}\right)\right\}dz+O\left(\frac{1}{n}\right)=O\left(\frac{1}{n}\right)+$$

$$+\frac{u_i}{2}-\frac{1}{\sqrt{n}}\,\frac{u_i}{3}\sqrt{\frac{2}{\pi}}\,\frac{1-2u_i}{\sqrt{u_i(1-u_i)}}\,.$$

Перейдем к изучению суммы $\mathcal{J}_1+\mathcal{J}_3$, после несложных преобразований

$$\mathcal{J}_1+\mathcal{J}_3=r\int\limits_0^1\varepsilon_i(y)dy+\frac{n!}{(r-1)!\,(n-r)!}\int\limits_0^1\varepsilon_i(x)\,\cdot$$

$$\cdot\int\limits_x^1 z^{r-1}(1-z)^{n-r-1}\big(r-1-(n-1)z\big)dz\,dx=nu_ip_i+$$

$$+\frac{n!}{(r-1)!\,(n-r)!}\int\limits_{u_{i-1}}^{u_i}\int\limits_x^1 z^{r-1}(1-z)^{n-r-1}\big(r-1-z(n-1)\big)dz\,dx=$$

$$=nu_ip_i-p_i-\frac{n!}{(r-1)!\,(n-r)!}\int\limits_{u_{i-1}}^{u_i}\int\limits_0^x z^{r-1}(1-z)^{n-r-1}\big(r-1-z(n-1)\big)dz\,dx.$$

Используя теперь формулу Стирлинга и замены переменных $x = u_i - y/\sqrt{n}$, $z = u_i - t/\sqrt{n}$, получим для $u_i \neq 0 \vee 1$

$$J_1 + J_3 = nu_i p_i - p_i - \frac{1}{(1-u_i)\sqrt{2\pi u_i(1-u_i)}} \int_0^{p_i\sqrt{n}} dy \int_y^{u_i\sqrt{n}} \left(1 - \frac{t}{u_i\sqrt{n}}\right)^{nu_i - 1}$$

$$\cdot \left(1 + \frac{t}{(1-u_i)\sqrt{n}}\right)^{(1-u_i)n-1} \left(\frac{n-1}{n} t - \frac{1-u_i}{\sqrt{n}}\right) dt \cdot \left(1 + O\left(\frac{1}{n}\right)\right).$$

Используя далее разложение логарифма и действуя как и ранее, получим

$$J_1 + J_3 = nu_i p_i - p_i - \frac{1}{(1-u_i)\sqrt{2\pi u_i(1-u_i)}} \int_0^{\infty} dy \int_y^{\infty} e^{-\frac{t^2}{2u_i(1-u_i)}} \cdot$$

$$\cdot \left(t - \frac{1-u_i}{\sqrt{n}} + \frac{t^2(1-2u_i)}{\sqrt{n}\, u_i(1-u_i)} - \frac{t^4(1-2u_i)}{3\sqrt{n}\, u_i^2(1-u_i)^2}\right) dt + O\left(\frac{1}{n}\right) =$$

$$= nu_i p_i - p_i - \frac{u_i}{2} + \frac{\sqrt{u_i(1-u_i)}}{\sqrt{2\pi n}} \left(1 + \frac{2}{3}\, \frac{1-2u_i}{1-u_i}\right) + O\left(\frac{1}{n}\right).$$

Собирая последние равенства воедино, в итоге получаем, что

$$J_1 + J_2 + J_3 = nu_i p_i - p_i + \sqrt{\frac{u_i(1-u_i)}{2\pi n}} + O\left(\frac{1}{n}\right).$$

Аналогично можно вывести, что

$$\sum_{s=1}^{n} E U_{(nu_{i-1})} \varepsilon_i\left(U_{(s)}\right) = nu_{i-1} p_i - \sqrt{\frac{u_{i-1}(1-u_{i-1})}{2\pi n}} + O\left(\frac{1}{n}\right),$$

тем самым доказана справедливость представления (9). Нетрудно показать, что формула (I) останется справедливой и в случае, когда $u_i = 1$, либо $u_{i-1} = 0$. Общий случай, когда величины np_i не обязательно целые, сводится к рассмотренному с $u_i' = [nu_i]/n$, здесь следует отметить, что все асимптотические равенства остаются в силе для u_i', поскольку $u_i' = u_i + O\left(\frac{1}{n}\right)$. Теорема доказана.

Идея преобразований плотности $f(\bar{x})$ принадлежит Н.В. Смирнову (1937), Ф.Мостеллеру (1946) и развита, в частности, в работе [7] .

Автор благодарит Д.М.Чибисова за полезные замечания.

Литература

1. Bolsev L.N. Cluster analysis. — Bull. Intern. Statist. Inst., 1969, v. 43, p. 411-425.
2. Чибисов Д.М. Некоторые критерии типа Хи-квадрат для непрерывных распределений. - Теория вероятн. и ее примен., 1971, т.16, в.1, с.3-20.
3. Никулин М.С. О квантильном критерии. - Теория вероятн. и ее примен., 1974, т.19, в.2, с.431-434.
4. Esseen C.G. Fourier analysis of distribution function, A mathematical study of the Laplace-Gaussian law. — Acta Mathematica, 1945, v. 77, p. 1-125.
5. Калинин В.М. Специальные функции и предельные свойства вероятностных распределений.П. - В кн.: Исследования по классическим проблемам теории вероятностей и математической статистики.П. Зап.научн.семин.ЛОМИ, 1972, т.26, с.5-88.
6. Дэйвид Г. Порядковые статистики. М.: Наука, 1979. 336 с.
7. Reiss R.D. The asymptotic normality and asymptotic expantions for the joint distribution of several order statistics. — Colloquia Mathematica Societatis János Bolyai, 1977, v. 11, p. 297-340.
8. Kiefer J. On Bahadur's represantation of sample quantiles. — Ann. Mathem. Stat., 1967, v. 38, No. 5, p. 1323-1343.