Записки научных
семинаров ПОМИ
Том 408, 2012 г.

К. Ю. Волкова, Я. Ю. Никитин

КРИТЕРИИ СОГЛАСИЯ СО СТЕПЕННЫМ ЗАКОНОМ, ОСНОВАННЫЕ НА ХАРАКТЕРИЗАЦИИ ПУРИ–РУБИНА, И ИХ АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ

§1. Введение

Проверка гипотез согласия для параметрических семейств остается одной из важных и интересных статистических задач. Пусть $\mathcal{P}$ является семейством распределений *степенного закона* с функциями распределения (ф.р.)

$$F(x) = x^{\lambda}, \quad x \in (0,1),\ \lambda > 0, \tag{1}$$

зависящими от положительного параметра λ.

Это распределение является частью семейства бета-распределений, а также "инверсией" распределения Парето. Степенное распределение довольно часто встречается в прикладных областях, таких как теория массового обслуживания [4], экономика ценообразования [22] и надежность электрических систем [12].

Нас будут интересовать критерии для проверки согласия с указанным семейством, не зависящие от неизвестного параметра λ. Насколько нам известно, попытки построить подобные критерии были сделаны ранее лишь Мартыновым [10,11], который предложил использовать известный метод Дурбина [3], основанный на эмпирических процессах с оцененными параметрами.

В данной работе мы предлагаем совершенно иной путь для построения и анализа новых критериев согласия со степенным семейством, основанный на использовании характеризации степенного закона. Рассмотрим следующую характеризацию Пури и Рубина [20]:

Пусть X и Y – независимые неотрицательные случайные величины с ф.р. F. Тогда равенство по распределению X и $\min\left(\frac{X}{Y}, \frac{Y}{X}\right)$

Ключевые слова: степенной закон, U-статистики, характеризация, бахадуровская эффективность, проверка гипотез, локальная оптимальность.

Работа поддержана грантами РФФИ 10-01-00154, НШ-1216.2012.1 и ФЦП 2010-1.1-111-128-033.

имеет место в том и только том случае, когда X имеет степенное распределение с ф.р. из семейства $\mathcal{P}$.

Заметим, что указанный результат получен путем монотонного преобразования из характеризации экспоненциального закона, найденного в [20]. Это общий и традиционный метод для переформулировки характеризационных теорем. Однако, как было замечено в [5, с. 169], "несмотря на то, что свойство интересно для некоторого распределения, оно может потерять свою привлекательность после применения к нему преобразования". Тем не менее, мы считаем указанную выше характеризацию степенного закона весьма удобной для задачи проверки гипотез.

Критерии, основанные на характеризациях, и их эффективности в настоящее время мало изучены, и в этой области существует лишь небольшое количество публикаций, см., например, работы [2,6,13,19]. Мы считаем, что данная область является перспективным направлением исследований, в результате которых появляются простые критерии с довольно высокой эффективностью.

Пусть $X_1, X_2, \ldots$ – независимые одинаково распределенные наблюдения (н.о.р.с.в.), имеющие непрерывную ф.р. F. Рассмотрим задачу проверки сложной гипотезы $H_0 : F \in \mathcal{P}$ против альтернативы $H_1 : F \notin \mathcal{P}$, предполагая, что альтернативная ф.р. также сосредоточена на $(0,1)$.

Построим эмпирическую ф.р. $F_n(t) = n^{-1} \sum\limits_{i=1}^{n} \mathbf{1}\{X_i < t\}, t \in R^1$, основанную на выборке $X_1, \ldots, X_n$. В соответствии с характеризацией Пури–Рубина построим так называемую U-эмпирическую ф.р., см. [8, 9], с помощью формулы

$$H_n(t) = \binom{n}{2}^{-1} \sum_{1 \leqslant i < j \leqslant n} \mathbf{1}\left\{\min\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) < t\right\}, \quad t \in (0,1).$$

Рассмотрим две новые статистики для проверки гипотезы H_0 против H_1:

$$I_n^{PR} = \int_0^1 (H_n(t) - F_n(t))\, dF_n(t), \tag{2}$$

$$D_n^{PR} = \sup_{t \in [0,1]} |H_n(t) - F_n(t)|. \tag{3}$$

Первая из статистик – интегрального типа и схожа с классической статистикой ω_n^1, а вторая является статистикой типа Колмогорова. Мы опишем их предельные распределения при H_0 и вычислим локальную бахадуровскую эффективность новых критериев для ряда параметрических альтернатив. С этой целью нами будут изучены вероятности больших уклонений построенных статистик при основной гипотезе. Мы обсудим также условия их локальной асимптотической оптимальности в бахадуровском смысле.

За основными сведениями из теории Бахадура мы отсылаем к монографиям [1] и [14]. Указанный тип эффективности наиболее предпочтителен для решения нашей задачи, поскольку статистика типа Колмогорова не является асимптотически нормальной, а потому подход Питмена к ней не применим.

В теории Бахадура за меру качества последовательности статистик $\{T_n\}$ принимается точный наклон $c_T(\theta)$, описывающий степень экспоненциального убывания их P-значений при альтернативе. Известно, что всегда выполнено следующее неравенство Бахадура–Рагавачари

$$c_T(\theta) \leqslant 2K(\theta),$$

где $K(\theta)$ является "расстоянием" Кульбака–Лейблера между нулевой и альтернативной гипотезами, зависящим от действительного параметра θ. Поэтому естественно определить локальную бахадуровскую эффективность последовательности T_n формулой

$$\mathrm{eff}(T) := \lim_{\theta\to 0} c_T(\theta)/2K(\theta).$$

§2. Статистика I_n^{PR}

Статистика I_n^{PR} асимптотически эквивалентна U-статистике степени 3 с центрированным ядром

$$\Psi_{PR}(X,Y,Z) = \frac{1}{3}\left(\mathbf{1}\left\{\min\left(\frac{X}{Y},\frac{Y}{X}\right) < Z\right\}\right.$$
$$\left. + \mathbf{1}\left\{\min\left(\frac{X}{Z},\frac{Z}{X}\right) < Y\right\} + \mathbf{1}\left\{\min\left(\frac{Y}{Z},\frac{Z}{Y}\right) < X\right\}\right) - \frac{1}{2}.$$

Заметим, что статистики I_n^{PR} и D_n^{PR} при H_0 инвариантны относительно замены $X \to X^{1/\lambda}$. Поэтому можно положить $\lambda = 1$, т.е. считать исходную выборку равномерной на $(0,1)$.

Как известно, см. [7,9], невырожденные U- и V-статистики асимптотически нормальны. Чтобы показать, что ядро $\Psi_{PR}(X,Y,Z)$ невырождено, вычислим его проекцию ψ_{PR}. Для фиксированного $X=s$ получаем

$$\begin{aligned}\psi_{PR}(s) &:= \mathbf{E}(\Psi_{PR}(X,Y,Z) \mid X=s)\\ &= \frac{2}{3}\,\mathbf{P}\{\min\left(\frac{s}{Y},\frac{Y}{s}\right) < Z\} + \frac{1}{3}\,\mathbf{P}\left\{\min\left(\frac{Y}{Z},\frac{Z}{Y}\right) < s\right\} - \frac{1}{2}.\end{aligned}$$

Первая вероятность легко вычисляется:

$$\mathbf{P}\left\{\min\left(\frac{s}{Y},\frac{Y}{s}\right) < Z\right\} = 1 - \int\limits_s^1 \frac{s}{y}\,dy - \int\limits_0^s \frac{y}{s}\,dy = 1 + s\,\ln s - \frac{1}{2}\,s,$$

а из используемой характеризации следует, что

$$\mathbf{P}\{\min\left(\frac{Y}{Z},\frac{Z}{Y}\right) < s\} = \mathbf{P}\{Y<s\} = s, \quad 0 \leqslant s \leqslant 1.$$

Поэтому окончательно получаем выражение для проекции ядра:

$$\psi_{PR}(s) = \frac{1}{6} + \frac{2}{3}\,s\ln s, \quad 0 \leqslant s \leqslant 1. \tag{4}$$

Следовательно, дисперсия проекции равна

$$\Delta_{PR}^2 = \int\limits_0^1 \psi_{PR}^2(s)\,ds = \frac{5}{972},$$

и положительна. Таким образом, ядро $\Psi_{PR}(X,Y,Z)$ невырождено. Отсюда по теореме Хёффдинга [7,9],

$$\sqrt{n}\,I_n^{PR} \xrightarrow{d} \mathcal{N}\left(0,\frac{5}{108}\right).$$

Ядро Ψ_{PR} не только центрировано и невырождено, но и ограничено. Применяя теорему о больших уклонениях невырожденных U- и V-статистик из [17], см. также [15], получаем:

Теорема 1. *При* $a>0$

$$\lim_{n\to\infty} n^{-1}\ln\mathbf{P}\left\{I_n^{PR} > a\right\} = -f(a),$$

где функция f непрерывна при достаточно малых $a > 0$, причем

$$f(a) \sim \frac{a^2}{18\Delta_{PR}^2} = \frac{54}{5}\,a^2 \quad \textit{при } a \to 0.$$

Когда исходное распределение является равномерным, или, в более общем случае, степенным, не существует общепринятых стандартных альтернатив. Поэтому мы рассмотрим три альтернативы: альтернативу загрязнения и две альтернативы, не имеющие устоявшегося названия и сосредоточенные на $(0,1)$. Выражения для ф.р. этих альтернатив записываются формулами:

$$G_1(x,\theta) = (1-\theta)x + \theta x^r, \quad 0 \leqslant \theta \leqslant 1,\ r > 1,\ x \in (0,1).$$

$$G_2(x,\theta) = x - \theta\sin(\pi x), \quad 0 \leqslant \theta < 1/\pi,\ x \in (0,1).$$

$$G_3(x,\theta) = x + \theta\int\limits_0^x \left(\frac{1}{6} + \frac{2}{3}y\ln y\right)dy, \quad 0 \leqslant \theta \leqslant 1,\ x \in (0,1).$$

Обозначим их плотности через $g_j(x,\theta)$, $j = 1,2,3$, соответственно.

В дальнейших рассуждениях нам потребуются выражения при $\theta \to 0$ "расстояния" Кульбака–Лейблера между нулевой гипотезой и указанными альтернативами. Выведем выражение для этого расстояния в общем виде при $\theta \to 0$, с учетом того, что основная гипотеза является сложной. Положим

$$K(\theta) = \inf_{\lambda>0}\int\limits_0^1 \ln\frac{g(x,\theta)}{\lambda x^{\lambda-1}}\,g(x,\theta)\,dx. \tag{5}$$

Лемма 1. *Пусть $g(x,\theta)$ – альтернативная плотность на $(0,1)$, регулярная настолько, чтобы любая производная по параметру под знаком интеграла в доказательстве имела смысл и информация Кульбака–Лейблера (5) была строго определена. Тогда*

$$2K(\theta) \sim \theta^2\left[\int\limits_0^1 (g_\theta'(x,0))^2\,dx - \left(\int\limits_0^1 g_\theta'(x,0)\ln x\,dx\right)^2\right], \quad \theta \to 0. \tag{6}$$

Доказательство. Инфимум в (5) достигается при

$$\lambda^* = -\Big(\int\limits_0^1 g(x,\theta)\ln x\,dx\Big)^{-1}$$

и равен

$$K(\theta) = \int_0^1 g(x,\theta)\ln g(x,\theta)\,dx - (\lambda^* - 1)\int_0^1 g(x,\theta)\ln x\,dx - \ln\lambda^*. \qquad (7)$$

При $\theta \to 0$ функция $K(\theta)$ ведет себя как

$$K(\theta) \ \sim K(0) + K'(0)\cdot\theta + \frac{1}{2}K''(0)\cdot\theta^2.$$

Нетрудно видеть, что $K(0) = 0$ и что $K'(0) = 0$. Дважды продифференцировав правую часть (7) по θ, находим

$$K''(\theta) = \int_0^1 g''_{\theta^2}(x,\theta)(1 + \ln g(x,\theta) + \ln x)\,dx + \int_0^1 \left(g'_\theta(x,\theta)\right)^2 / g(x,\theta)\,dx$$

$$+ \int_0^1 g''_{\theta^2}(x,\theta)\ln x\,dx \left(\int_0^1 g(x,\theta)\ln x\,dx\right)^{-1} - \left(\frac{\int_0^1 g'_\theta(x,\theta)\ln x\,dx}{\int_0^1 g(x,\theta)\ln x\,dx}\right)^2.$$

Подставив в последнее выражение $\theta = 0$, получаем требуемое. □

Вычислим теперь локальный бахадуровский наклон и локальную эффективность последовательности статистик I_n^{PR} для альтернативы с ф.р. $G(x,\theta)$ и плотностью $g(x,\theta)$, предполагая их регулярность и возможность дифференцировать под знаком интеграла. Эти требования выполнены для всех трех рассматриваемых альтернатив. Обозначим также $h(x) = g'_\theta(x,0)$. Заметим, что $\int_0^1 h(x)\,dx = 0$.

Согласно закону больших чисел для U-статистик [9], предел по вероятности статистик I_n^{PR} при любой альтернативе при $\theta \to 0$ равен

$$b_I(\theta; PR) = \mathbf{P}_\theta\left\{\min\left(\frac{X}{Y}, \frac{Y}{X}\right) < Z\right\} - \frac{1}{2}$$

$$= 2\int_0^1 g(z,\theta)\,dz \int_0^1 g(y,\theta)G(yz,\theta)\,dy - \frac{1}{2} \sim J(0) + J'(0)\cdot\theta. \qquad (8)$$

Легко видеть, что

$$J(0) = 0,\ J'(0) = 2\int\limits_0^1 h(z)z\,dz + 2\int\limits_0^1 dz\int\limits_0^1 dy\int\limits_0^{yz} h(x)\,dx.$$

Дважды изменив порядок интегрирования, получим

$$\int\limits_0^1 dz\int\limits_0^1 dy\int\limits_0^{yz} h(x)\,dx = \int\limits_0^1 dz\int\limits_0^z h(x)\left(1-\frac{x}{z}\right)\,dx$$

$$= \int\limits_0^1 h(x)\,(x\ln x - x)\,dx.$$

Поэтому заключаем, что

$$b_I(\theta; PR) \sim 3\theta\int\limits_0^1 \psi_{PR}(x)\,h(x)\,ds. \tag{9}$$

Альтернатива загрязнения. После ряда нетрудных вычислений из (9) при $\theta \to 0$ получаем, что $b_I(\theta) \sim (r-1)^2/2(r+1)^2\cdot\theta$. Следовательно, локальный точный наклон последовательности статистик I_n^{PR} при $\theta \to 0$ допускает представление

$$c_I(\theta; PR) \sim \frac{108}{5}\,b_I^2(\theta; PR) = \frac{27(r-1)^4}{5(r+1)^4}\,\theta^2.$$

Легко получить, используя (6), что для альтернативной ф.р. G_1

$$2K(\theta) \sim \frac{(r-1)^4}{r^2(2r-1)}\theta^2, \quad \theta \to 0. \tag{10}$$

Следовательно, локальная бахадуровская эффективность нашего критерия равна

$$\mathrm{eff}(r; I^{PR}) = \lim_{\theta\to 0}\frac{c_I(\theta; PR)}{2K(\theta)} = \frac{27(2r-1)r^2}{5(r+1)^4}, quad r > 1.$$

Найденная эффективность довольно высока при умеренных значениях r и достигает своего максимума при $r = 2+\sqrt{3}$, когда она равна 0.970.

Вторая альтернатива. Вычисление локальной бахадуровской эффективности для случая альтернативы G_2 проходит аналогичным образом. Используя результат (9), имеем $b_I(\theta; PR) \sim 0.224 \cdot \theta^2$, таким образом, локальный точный наклон последовательности статистик I_n^{PR} при $\theta \to 0$ допускает представление $c_I(\theta; PR) \sim 1.083 \cdot \theta^2$.

Для второй альтернативы информация Кульбака¬Лейблера легко вычисляется с помощью (6), имеем

$$2K(\theta) \sim 1.505 \cdot \theta^2, \quad \theta \to 0. \tag{11}$$

Следовательно, локальная бахадуровская эффективность нашего критерия равна $\mathrm{eff}(I^{PR}) = 0.719$.

Третья альтернатива. В случае третьей альтернативы вычисления проходят аналогично, после некоторых выкладок мы получаем, что $c_I(\theta; PR) \sim \frac{5\theta^2}{972}$ при $\theta \to 0$. Информация Кульбака¬Лейблера в этом случае также удовлетворяет соотношению

$$2K(\theta) \sim \frac{5\theta^2}{972}, \quad \theta \to 0. \tag{12}$$

Следовательно, локальная бахадуровская эффективность нашего критерия равна 1, и интегральный тест является локально оптимальным в бахадуровском смысле, см. [14, гл. 6]. Мы вернемся к объяснению этого явления в последнем параграфе работы.

Таблица 1. Локальная бахадуровская эффективность для статистики I_n^{PR}.

Альтернатива	Эффективность
G_1	0.970, $r \approx 3.7$
G_2	0.719
G_3	1. 000

§3. Статистика типа Колмогорова D_n^{PR}

Теперь рассмотрим статистику колмогоровского типа (3). В этом случае при фиксированном t разность $H_n(t) - F_n(t)$ является семейством U-статистик с ядрами

$$\Xi_{PR}(X, Y; t) = \mathbf{1}\{\min(X/Y, Y/X) < t\} - \frac{1}{2}\mathbf{1}\{X < t\} - \frac{1}{2}\mathbf{1}\{Y < t\},$$

зависящим от $t \in (0,1)$. Проекция этого ядра при фиксированном $t \in (0,1)$ имеет вид

$$\xi_{PR}(s;t) := \mathbf{E}\left(\Xi_{PR}(X,Y;t) \mid X = s\right)$$
$$= \mathbf{P}\{\min(s/Y, Y/s) < t\} - \frac{1}{2}\mathbf{1}\{s < t\} - \frac{1}{2}\mathbf{P}\{Y < t\}.$$

После нетрудных вычислений получаем

$$\xi_{PR}(s;t) = \mathbf{1}\{s < t\}\Big(\frac{1}{2} - \frac{s}{t}\Big) + st - \frac{1}{2}\,t. \tag{13}$$

Теперь вычислим дисперсию $\delta^2_{PR}(t)$ этой проекции. После ряда простых вычислений получаем

$$\delta^2_{PR}(t) := \mathbf{E}\,\xi^2_{PR}(X_1;t) = \frac{1}{12}t(1+t-2t^2), \quad 0 < t < 1. \tag{14}$$

Нетрудно видеть, что супремум функции $\delta^2_{PR}(t)$ достигается в точке $t^* = \frac{1+\sqrt{7}}{6}$ и равен $\delta^2_{PR} \approx 0.044$. Поэтому наше семейство ядер $\Xi_{PR}(X;t)$ невырождено, см. [15].

Предельное распределение статистики D_n^{PR} неизвестно. Опираясь на методы работы [21], можно показать, что U-эмпирический процесс

$$\eta_n(t) = \sqrt{n}\,(H_n(t) - F_n(t))\,, \quad t \in (0,1),$$

слабо сходится при $n \to \infty$ к некоторому центрированному гауссовскому процессу $\eta(t)$ со сложной ковариацией. Тогда последовательность статистик $\sqrt{n}D_n^{PR}$ слабо сходится к величине $\sup\limits_t |\eta(t)|$, найти распределение которой не представляется возможным. Поэтому критические значения для статистик D_n^{PR} целесообразно искать с помощью моделирования их выборочного распределения.

Семейство ядер $\{\Xi_{PR}(X,Y;t)\}, t \in (0,1)$ не только центрировано, но и ограничено. Применяя теорему [15] о больших уклонениях семейств невырожденных U-статистик, получаем:

Теорема 2. *Для достаточно малых $a > 0$ выполнено*

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}(D_n^{PR} > a) = -k(a),$$

где функция k непрерывна, причем

$$k(a) = \frac{a^2}{8\delta^2_{PR}}(1 + o(1)) \sim 2.84a^2 \quad \textit{при} \;\; a \to 0.$$

Альтернатива загрязнения. Вычислим локальный бахадуровский наклон и локальную эффективность статистики (3) для альтернативы с ф.р. $G_1(x,\theta)$. По теореме Гливенко–Кантелли для U-статистических ф.р. [8] предел D_n^{PR} по вероятности для произвольной альтернативы равен при $\theta \to 0$

$$b_D(t,\theta;PR) = \sup_{0\leqslant t\leqslant 1}\left|2\int\limits_0^1 g(y,\theta)\,G(ty,\theta)\,dy - G(t,\theta)\right|. \tag{15}$$

Предполагая регулярность альтернативной ф.р., мы заключаем

$$b_D(t,\theta;PR) \sim 2\sup_{0\leqslant t\leqslant 1}\left|\int\limits_0^1 \xi_{PR}(s;t)\,h(s)\,ds\right|\cdot\theta, \tag{16}$$

где $\xi_{PR}(s;t)$ – проекция из (13). Применяя полученную формулу, мы имеем для нашей альтернативы

$$b_D(t,\theta;PR) \sim \frac{(r-1)^2}{r+1}r^{-\frac{r}{r-1}}\theta, \quad \theta\to 0.$$

Таким образом, локальный точный наклон последовательности статистик D_n^{PR} при $\theta \to 0$ допускает представление

$$c_D(\theta;PR) \sim \frac{5.68(r-1)^4 r^{-\frac{2r}{r-1}}}{(r+1)^2}\,\theta^2.$$

Информация Кульбака–Лейблера удовлетворяет (10). Следовательно, локальная бахадуровская эффективность нашего критерия равна

$$\mathrm{eff}(r;D^{PR}) = \frac{5.68(2r-1)r^{-\frac{2}{r-1}}}{(r+1)^2}.$$

Можно показать, что максимальное значение локальной эффективности последовательности $\{D_n^{PR}\}$ достигается для $r = 4.64$ и равно 0.636, причем его значение при $3 \leqslant r \leqslant 12$ больше, чем 0.5.

Вторая альтернатива. Вычисление локальной бахадуровской эффективности для случая альтернативы G_2 происходит аналогичным образом. С помощью (16) при $\theta \to 0$ получаем $b_D(\theta;PR) \sim 0.367\cdot\theta$. Таким образом, локальный точный наклон последовательности статистик D_n^{PR} асимптотически равен $c_D(\theta;PR) \sim 0.765\cdot\theta^2$. Информация

Кульбака–Лейблера в этом случае вычисляется с помощью (11). Следовательно, локальная бахадуровская эффективность нашего критерия равна $\mathrm{eff}(D^{PR}) = 0.508$.

Третья альтернатива. В случае альтернативной ф.р. $G_3(x,\theta)$ при $\theta \to 0$ мы получаем, что $b_D(\theta; PR) \sim 0.0249{\cdot}\theta$. Поэтому локальный точный наклон последовательности статистик D_n^{PR} при $\theta \to 0$ допускает представление $c_D(\theta; PR) \sim 0.00352 \cdot \theta^2$. Мы знаем, что информация Кульбака–Лейблера в этом случае удовлетворяет (12). Следовательно, локальная бахадуровская эффективность нашего критерия равна 0.685.

Таблица 2. Локальная бахадуровская эффективность для статистики D_n^{PR}.

Альтернатива	Эффективность
G_1	0.636, $r \approx 4.64$
G_2	0.508
G_3	0.685

Заметим, что эффективность статистик типа Колмогорова меньше, чем у интегральных статистик I_n^{PR}, как это обыкновенно бывает в задачах проверки гипотез, см. также [14].

§4. Условия локальной асимптотической оптимальности

Найдем условия локальной асимптотической оптимальности (ЛАО) в смысле Бахадура для обеих статистик I_n^{PR} и D_n^{PR}, то есть опишем семейства альтернатив, для которых изучаемая статистика имеет наибольшую локальную эффективность, такую, что выполнено соотношение, см. [14], [18],

$$c_T(\theta) \sim 2K(\theta), \quad \theta \to 0,$$

Такой класс альтернатив образует область ЛАО для рассматриваемой последовательности статистик.

Рассмотрим функции

$$H(x) = G_\theta^{'}(x,\theta)\,|_{\theta=0}, \quad h(x) = g_\theta^{'}(x,\theta)\,|_{\theta=0}\,.$$

Будем предполагать, что выполнены следующие условия регулярности, см. [18]:

$$\int_0^1 h^2(x)\,dx < \infty, \quad \text{где} \quad h(x) = H'(x), \tag{17}$$

$$\frac{\partial}{\partial\theta}\int_0^1 g(x,\theta)\,q(x)\,dx\,|_{\theta=0} = \int_0^1 h(x)\,q(x)\,dx, \quad \forall q \in L_1(0,1). \tag{18}$$

Обозначим через $\mathcal{G}$ класс плотностей $g(x,\theta)$ с ф.р. $G(x,\theta)$, удовлетворяющими условиям регулярности (17) - (18). Выведем условия оптимальности в терминах функции $h(x)$.

Для альтернативных плотностей из класса $\mathcal{G}$ верны все рассуждения леммы 1, поэтому выполнена асимптотика

$$2K(\theta) \sim \left\{\int_0^1 h^2(x)\,dx - \left(\int_0^1 h(x)\ln x\,dx\right)^2\right\}\theta^2, \quad \theta \to 0.$$

Сначала рассмотрим интегральную статистику I_n^{PR} с проекциями $\psi_{PR}(x) = \frac{1}{6} + \frac{2}{3}x\ln x$. Введем вспомогательную функцию

$$h_0(x) = h(x) - (\ln x + 1)\int_0^\infty \ln u\,h(u)\,du.$$

Простыми вычислениями получаем

$$\int_0^1 h^2(x)\,dx - \left(\int_0^1 h(x)\ln x\,dx\right)^2 = \int_0^1 h_0^2(x)\,dx,$$

$$\int_0^1 \psi_{PR}(x)\,h(x)\,dx = \int_0^1 \psi_{PR}(x)\,h_0(x)\,dx,$$

$$\int_0^1 \xi_{PR}(x;t)\,h(x)\,dx = \int_0^1 \xi_{PR}(x;t)\,h_0(x)\,dx \quad \text{для любого } t \in (0,1).$$

Поэтому локальная асимптотическая эффективность, с учетом (9), принимает вид

$$\begin{aligned}\operatorname{eff}(I_n^{PR}) &= \lim_{\theta\to 0} b_I^2(\theta; PR) \big/ \left(9\Delta_{PR}^2 \cdot 2K(\theta)\right) \\ &= \left(\int_0^1 \psi_{PR}(x)\, h_0(x)\, dx\right)^2 \Bigg/ \left(\int_0^1 \psi_{PR}^2(x)\, dx \cdot \int_0^1 h_0^2(x)\, dx\right).\end{aligned}$$

С помощью неравенства Коши–Буняковского–Шварца находим, что выражение в правой части равно 1 тогда и только тогда, когда $h_0(x) = C_1\psi_{PR}(x)$ для некоторой константы $C_1 > 0$, так что $h(x) = C_1\psi_{PR}(x) + C_2(\ln x + 1)$ для некоторых констант $C_1 > 0$ и C_2. Распределения, для которых функция $h(x)$ указанного вида, образуют область ЛАО в классе $\mathcal{G}$. Примером такой альтернативы может служить та, у которой плотность $g(x,\theta)$ при всех $\theta > 0$ выражается формулой:

$$g(x,\theta) = 1 + \theta\left(\frac{1}{6} + \frac{2}{3}\, x \ln x\right), \quad 0 \leqslant x \leqslant 1. \tag{19}$$

Этим объясняется, почему третья альтернатива является локально оптимальной для теста, основанного на статистике I_n^{PR}. Это соответствует результатам работы [16], где аналогичная задача была решена для случая, когда нулевая гипотеза является *простой*.

Теперь рассмотрим статистику колмогоровского типа D_n^{PR} с семействами ядер $\Xi_{PR}(X, Y; t)$ и проекциями

$$\xi_{PR}(x; t) = \mathbf{1}\{x < t\}\Big(\frac{1}{2} - \frac{x}{t}\Big) + xt - \frac{t}{2}.$$

В этом случае нетрудно видеть, что верна асимптотика:

$$b_D(\theta; PR) \sim 2\theta \sup_{t\in(0,1]} \left|\int_0^1 \xi_{PR}(x; t) h_0(x)\, dx\right|. \tag{20}$$

Поэтому локальная асимптотическая эффективность имеет вид

$$\operatorname{eff}(D^{PR}) = \lim_{\theta\to 0}\left[b_D^2(\theta;PR)/\sup_{t\in(0,1)}\left(4\delta_{PR}^2(t)\right)\cdot 2K(\theta)\right]$$
$$= \frac{\sup_{t\in(0,1)}\left(\int\limits_0^1 \xi_{PR}(x;t)h_0(x)\,dx\right)^2}{\sup_{t\in(0,1)}\left(\int\limits_0^1 \xi_{PR}^2(x,t)\,dx\cdot\int\limits_0^1 h_0^2(x)\,dx\right)} \leqslant 1.$$

Снова применяем неравенство Коши–Буняковского–Шварца к интегралу в (20). Тогда статистика D_n^{PR} локально оптимальна тогда и только тогда, когда $h(x) = C_3\xi_{PR}(x,t_0) + C_4(\ln x + 1)$ для

$$t_0 = \arg\sup_{t\in(0,1)} \delta_{PR}^2(t) = \frac{1+\sqrt{7}}{6} \tag{21}$$

и для некоторых констант $C_3 > 0$ и C_4. Распределения, для которых функция $h(x)$ указанного вида, образуют область ЛАО в соответствующем классе $\mathcal{G}$. Простейший пример такой альтернативы с плотностью $g(x,\theta)$ при всех $\theta > 0$ выражается формулой:

$$g(x,\theta) = 1 + \theta\left(\mathbf{1}\{x < t_0\}\left(\frac{1}{2} - \frac{x}{t_0}\right) + xt_0 - \frac{t_0}{2}\right), \quad 0 \leqslant x \leqslant 1, \tag{22}$$

где t_0 из (21).

Таким образом, мы видим, что существуют альтернативные плотности специального вида (19) и (22) относительно простой формы, для которых наши последовательности статистик являются локально оптимальными.

§5. Заключение

Нами были предложены два новых критерия для проверки согласия со степенным законом, основанные на классической характеризации Пури и Рубина. Мы описали их предельное распределение и большие уклонения, а также вычислили локальную бахадуровскую эффективность для ряда альтернатив, которая оказалась довольно высокой. Нами были также найдены специальные альтернативы, для которых новые критерии являются локально оптимальными. Полученные результаты отражают достоинства новых критериев и потенциальную возможность применения их на практике.

Литература

1. R. R. Bahadur, *Some Limit Theorems in Statistics.* SIAM, Philadelphia, 1971.
2. L. Baringhaus, N. Henze, *Tests of fit for exponentiality based on a characterization via the mean residual life function.* — Statist. Papers **41** (2000), 225–236.
3. J. Durbin, *Weak convergence of the sample distribution function when parameters are estimated.* — Ann. Statist. **1** (1973), 279–290.
4. M. A. Ferreira, M. Andrade, *The $M/G/\infty$ queue busy period distribution exponentiality.* — J. Appl. Math. **4** (2011), 249–260.
5. J. Galambos, *Characterizations.* In: Handbook of the Logistic Distribution, ed. by N. Balakrishnan, Marcel Dekker, New York (1992), 170–188.
6. N. Henze, S. Meintanis, *Goodness-of-fit tests based on a new characterization of the exponential distribution.* — Commun. Statist. Theor. Meth. **31** (2002), 1479–1497.
7. W. Hoeffding, *A class of statistics with asymptotically normal distribution.* — Ann. Math. Statist. **19** (1948), 293–325.
8. P. L. Janssen, *Generalized Empirical Distribution Functions With Statistical Applications.* Limburgs Universitair Centrum, Diepenbeek, 1988.
9. V. S. Korolyuk, Yu. V. Borovskikh, *Theory of U-Statistics.* Kluwer, Dordrecht, 1994.
10. G. V. Martynov, *Cramér-von Mises test for the Weibull and Pareto distributions.* — Proc. Dobrushin Intern. Conf., Moscow (2009), 117–122.
11. G. V. Martynov, *Cramér-von Mises test for the Weibull and Pareto distributions.* Abstr. intern. conf."Probability and Statistics with applications" dedicated to the 100th anniv. of the birthday of Béla Gyires. Debrecen (2009), 42–44.
12. M. Meniconi, D. M. Barry, *The power function distribution. A useful and simple distribution to assess electrical component reliability.* — Microelectr. Reliab. **36** (1996), 1207–1212.
13. K. Morris, D. Szynal, *Goodness-of-fit tests using characterizations of continuous distributions.* — Appl. Math. (Warsaw) **28** (2001), 151–168.
14. Y. Nikitin, *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, New York, 1995.
15. Ya. Yu. Nikitin, *Large deviations of U-empirical Kolmogorov–Smirnov tests, and their efficiency.* — J. Nonpar. Stat. **22** (2010), 649–668.
16. Ya. Yu. Nikitin, I. Peaucelle, *Efficiency and local optimality of distribution-free tests based on U- and V-statistics.* — Metron **LXII** (2004), 185–200.
17. Ya. Yu. Nikitin, E. V. Ponikarov, *Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics.* — Proc. St.Petersburg Math. Soc. **7** (1999), 124–167. Engl. transl. in AMS Transl., ser. 2 **203** (2001), 107–146.
18. Ya. Yu. Nikitin, A. V. Tchirina, *Bahadur efficiency and local optimality of a test for the exponential distribution based on the Gini statistic.* — Statist. Meth. Appl. **5** (1996), 163–175.
19. Ya. Yu. Nikitin, K. Yu. Volkova, *Asymptotic efficiency of exponentiality tests based on order statistics characterization.* — Georgian Math. J. **17** (2010), 749–763.
20. P. S. Puri, H. Rubin, *A characterization based on the absolute difference of two iid random variables.* — Ann. Math. Statist. **41** (1970), 2113–2122.

21. B. W. Silverman, *Convergence of a class of empirical distribution functions of dependent random variables.* — Ann. Probab. **11** (1983), 745–751.
22. Z. Wu, B. Kazaz, S. Webster, K. K. Yang, *Ordering, pricing, and lead-time quotation under lead-time and demand uncertainty.* — Production Operat. Manag., **21** (2012), 576–589.

Volkova K. Yu., Nikitin Ya. Yu. Goodness-of-fit tests for the power function distribution based on Puri–Rubin characterization and their efficiencies.

We construct integral and supremum type goodness-of-fit tests for the family of power distribution functions. Test statistics are functionals of $U-$empirical processes and are based on the classical characterization of power function distribution family belonging to Puri and Rubin. We describe the logarithmic large deviation asymptotics of test statistics under null-hypothesis, and calculate their local Bahadur efficiency under common parametric alternatives. Conditions of local optimality of new statistics are given.

Санкт-Петербургский государственный университет, математико-механический факультет, Университетский пр. 28, Петродворец 198504, Россия

E-mail: efrksenia@gmail.com
yanikit47@gmail.com

Поступило 2 октября 2012 г.