Записки научных
семинаров ПОМИ
Том 454, 2016 г.

Г. Т. Букия, Я. Ю. Никитин

АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ НОВЫХ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ СИММЕТРИИ ДЛЯ ОБОБЩЕННЫХ СКОШЕННЫХ АЛЬТЕРНАТИВ

§1. Введение

Симметрия распределения случайной величины с непрерывной функцией распределения (ф.р.) F означает выполнение условия

$$F(x) = 1 - F(-x) \qquad \text{для всех } x \in \mathbb{R}^1. \tag{1}$$

Предположим, что у нас имеется выборка $X_1, \dots, X_n$ с непрерывной ф.р. F и перед нами стоит задача проверки гипотезы $\mathcal{H}$ о том, что F симметрична, т.е. выполняется соотношение (1), против альтернативы о том, что это соотношение нарушается хотя бы в одной точке. Проверка гипотезы симметрии – одна из классических задач непараметрической статистики. Существует целый ряд критериев симметрии, например, критерий знаков и знаково-ранговый критерий Вилкоксона, а также более сложные критерии (см. [18,20], [28, гл. 4]), идеи которых весьма разнятся. Среди них лишь немногочисленные критерии основаны на характеризациях симметрии и тем самым используют внутренние, определяющие особенности свойства симметрии. Среди таких исключений работы [10, 12, 23].

В отличие от критериев симметрии, построение и исследование *критериев согласия*, основанных на характеризационных свойствах распределений, в последние годы быстро развивается. Примерами могут служить работы [19, 26, 27, 32–34] и ряд других.

В статье Ахсалуннаха [1] доказано, что симметрия характеризуется тождественностью распределений экстремальных порядковых статистик. Справедлива

Ключевые слова: критерий симметрии, U-эмпирическое распределение, бахадуровская эффективность, скошенная альтернатива.

Работа второго автора поддержана грантом РФФИ 16-01-00258.

Теорема 1. *Пусть* $X_1, \ldots, X_n$ – *независимые наблюдения, имеющие общую непрерывную ф.р.* F. *Рассмотрим экстремальные порядковые статистики*

$$X_{1,n} := \min(X_1, \ldots, X_n), \quad X_{n,n} := \max(X_1, \ldots, X_n).$$

Распределение X_i *симметрично, т.е. ф.р.* F *удовлетворяет условию* (1) *тогда и только тогда, когда* $|X_{1,n}| \stackrel{d}{=} |X_{n,n}|$.

Это свойство легло в основу свободного от распределения критерия симметрии, разработанного Никитиным и Ахсалуннахом в [30].

Зафиксируем произвольное целое $k > 1$ и рассмотрим V-эмпирические функции распределения величин $|X_{1,n}|$ и $|X_{n,n}|$:

$$G_n^{(k)}(t) = n^{-k} \sum_{1 \leqslant i_1, \ldots, i_k \leqslant n} \mathbf{1}\{|\min(X_{i_1}, \ldots, X_{i_k})| < t\}, \quad t \in \mathbb{R},$$

$$H_n^{(k)}(t) = n^{-k} \sum_{1 \leqslant i_1, \ldots, i_k \leqslant n} \mathbf{1}\{|\max(X_{i_1}, \ldots, X_{i_k})| < t\}, \quad t \in \mathbb{R}.$$

По теореме Гливенко–Кантелли для V-эмпирических функций распределения [17] при справедливости гипотезы $\mathcal{H}$ функции $G_n^{(k)}(t)$ и $H_n^{(k)}(t)$ должны с ростом n неограниченно сближаться, что позволяет построить на их разности статистики для критериев симметрии.

Мы будем рассматривать два типа статистик, предложенных в [1]: интегральную

$$J_n^{(k)} = \int_{\mathbb{R}} [G_n^{(k)}(t) - H_n^{(k)}(t)] \, dF_n(t), \tag{2}$$

где F_n – эмпирическая функция распределения выборки $|X_i|$, и типа Колмогорова

$$D_n^{(k)} = \sup_t |G_n^{(k)}(t) - H_n^{(k)}(t)|. \tag{3}$$

Таким образом, для всех значений параметра $k = 2, \ldots, n-1$ рассматриваются по два критерия, соответствующие интегральной и колмогоровской статистике. Номер в скобке вверху соответствует степени соответствующего ядра U или V-статистики.

Никитин и Ахсануллах вычислили в [30] локальную бахадуровскую асимптотическую эффективность этих статистик для альтернатив сдвига. Бахадуровская эффективность выбирается среди других видов эффективности, поскольку может применяться для статистик, распределение которых при нулевой гипотезе отлично от нормального. Это особенно ценно для статистик колмогоровского типа, имеющих ненормальные предельные распределения.

Для большинства рассматриваемых распределений указанная эффективность оказалась необычно высокой. Например, для последовательности интегральных статистик $\{J_n^{(3)}\}$ ее значение оказалось равно 0,977 в случае нормального распределения и 0,938 в случае логистического распределения. В то же время колмогоровская статистика $D_n^{(3)}$, как обычно, несколько менее эффективна – для указанных распределений значения эффективности 0,764 и 0,750.

Интересно было бы проверить, окажутся ли новые критерии столь же эффективными для более сложного и практически важного класса альтернатив. Поэтому в данной статье мы продолжим исследование локальной бахадуровской эффективности рассматриваемых статистик для обширного класса альтернатив, которые можно назвать *обобщенными скошенно-симметричными альтернативами.* При таких альтернативах плотность наблюдений асимметрична и имеет вид:

$$h(x,\theta) := c(\theta)\, G(w(x,\theta)) f(x), \tag{4}$$

где четная функция f – плотность распределения при выполнении гипотезы $\mathcal{H}$, G – некоторая гладкая неотрицательная функция, $c(\theta)$ – нормирующий множитель, а функция $w(\theta, x)$, называемая функцией скоса, будет определена в дальнейшем. Предполагается, что при нулевом значении параметра θ плотность обобщенного скоса совпадает с $f(x)$, поэтому наибольший интерес представляет случай близких к нулю значений параметра. Более подробно эти альтернативы будут описаны ниже.

Заметим, что литература, касающаяся асимптотического поведения критериев симметрии при скошенных альтернативах, довольно скудна, и мы можем назвать лишь статью [14], посвященную в основном ранговым критериям, имеются интересные работы [11] и [22], отдаленно связанные с темой нашего исследования, однако в них рассматриваются другие критерии, другие альтернативы и другой тип эффективности.

Мы вычислим значения локальных бахадуровских эффективностей статистик (2) и (3) для альтернатив (4) и в заключение построим альтернативы из рассматриваемого класса, наиболее эффективные по

Бахадуру для данных статистик, используя технику предыдущих работ, описанную в [14–16]. В процессе вычислений мы для краткости будем опускать малосущественные условия регулярности, которые легко восстановить.

§2. Структура обобщенных скошенно-симметричных альтернатив

Скошенно-симметричные распределения как гибкие и реалистические модели альтернатив известны в математической статистике уже более века. Первые исследования механизмов скоса симметричных распределений принадлежат де Эльгеро [13], см. [21] и особенно [7]. Популярность скошенно-симметричным распределениям принесла известная работа Аззалини [3] о скошенном нормальном распределении, в которой была рассмотрена плотность

$$\varphi(x,\theta) = 2\,\varphi(x)\,\Phi(x\theta), \qquad -\infty < x < \infty, \tag{5}$$

где Φ и φ – функция распределения и плотность стандартного нормального закона.

Эта статья дала начало многочисленным исследованиям и обобщениям скошенных распределений. Среди множества работ на эту тему отметим [2,4,6,24,35], а также обобщающую монографию Аззалини и Капитанио [8].

Широкое обобщение скошенно-симметричного распределения на многомерный случай было предложено в статье [35]. Для одномерного случая рассмотренные там плотности имеют вид

$$2\,G(x-\theta)\,f(x-\theta), \tag{6}$$

где функция скоса G удовлетворяет условию $0 \leqslant G \leqslant 1$, условию симметрии $G(x) = 1 - G(-x)$, $x \in \mathbb{R}$, f – четная плотность, а θ – произвольный вещественный параметр. Авторы [35] доказали, что всякую непрерывную плотность можно представить в виде (6), иными словами, вообще любое распределение является в определенном смысле скошенным для некоторого симметричного распределения. Тот же класс плотностей, но в другой записи был рассмотрен в [5], см. [35, Prop. 2].

Мы используем несколько иной, но близкий способ построения обобщенных скошенно-симметричных альтернатив. Возвратимся к плотности (4) и опишем более детально ее компоненты. Нормирующий множитель c зависит лишь от параметра θ. Введем для удобства и обратную величину:

$$\widetilde{c}(\theta) = 1/c(\theta) = \int_{-\infty}^{\infty} G(w(x,\theta))\, f(x)\, dx.$$

Структурную функцию $G : \mathbb{R} \longrightarrow \mathbb{R}^+$ и функцию скоса $w(\theta, x) : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ подчиним следующим условиям:

(1) G дважды дифференцируема в окрестности нуля и $G(0) > 0$;
(2) $\int_{-\infty}^{\infty} G(w(x,\theta))\, f(x)\, dx < \infty$ для всех θ;
(3) $w(x,\theta) = -w(-x,\theta)$ для любого x;
(4) w дважды дифференцируема по θ в окрестности нуля;
(5) $w(x,0) = w(0,\theta) = 0$ для всех x и θ;
(6) $\int_{-\infty}^{\infty} (w'_\theta(x,0))^2 f(x)\, dx < \infty$.

Будем обозначать $H(x,\theta)$ ф.р., отвечающую плотности $h(x,\theta)$ и пусть $g(x) = G'(x)$.

Приведем ряд конкретных примеров обобщенных скошенно-симметричных альтернатив.

1) Классическая скошенно-нормальная плотность (5) в то же время является обобщенной скошенно-симметричной плотностью в нашем понимании. В самом деле, обозначив в (4) $w(x,\theta) := x\theta$, $c(\theta) := 2$, $G(x) = \Phi(x)$, мы обеспечим выполнение свойств 1–6.

2) Вообще говоря, альтернатива сдвига не является обобщенной скошенной альтернативой. Однако нормальная плотность со сдвигом $\varphi(x-\theta)$ удовлетворяет условиям обобщенного (но не классического) скоса. Действительно,

$$\varphi(x-\theta) = \frac{e^{-(x-\theta)^2/2}}{\sqrt{2\pi}} = e^{-\theta^2/2} \cdot e^{\theta x} \cdot \varphi(x),$$

где функции $c(\theta) = e^{-\theta^2/2}$, $G(x) = e^x$ и $w(x,\theta) = x\theta$ удовлетворяют условиям 1–6.

3) Частный случай обобщенного скоса нормальной плотности. Рассмотрим альтернативную плотность

$$h(x,\theta) = 2\,\varphi(x)\,\Phi\left(\frac{x\theta}{\sqrt{1+x^2\theta^2}}\right), \quad x \in \mathbb{R}.$$

Здесь

$$w(x,\theta) = \frac{x\theta}{\sqrt{1+x^2\theta^2}},$$

и все необходимые свойства выполнены. Такая модель была представлена в [2] и впоследствии неоднократно исследовалась, см. об этом [8, с. 48].

4) Гибкий класс плотностей с кубической функцией скоса.
Плотность

$$h(x,\theta) = 2\,\varphi(x)\,\Phi((x+x^3)\,\theta), \quad x \in \mathbb{R},$$

впервые рассматривалась в [24], а затем во многих публикациях, см. подробнее [8, с. 49].

Функцию распределения альтернативы можно разложить в ряд при малых значениях параметра θ, когда скошенная альтернатива близка симметричному распределению. Из определения функций w следует, что

$$w'_\theta(x,\theta) = -w'_\theta(-x,\theta), \qquad w''_\theta(x,\theta) = -w''_\theta(-x,\theta).$$

Дважды дифференцируемый множитель $c(\theta)$ имеет соответствующие производные по θ:

$$\begin{aligned} c'(\theta) &= -\widetilde{c}'(\theta)/\widetilde{c}^2(\theta), \\ c''(\theta) &= 2\,(\widetilde{c}'(\theta))^2/(\widetilde{c}(\theta))^3 - (\widetilde{c}(\theta))''/(\widetilde{c}(\theta))^2, \end{aligned}$$

где

$$\widetilde{c}(\theta) = \int_{-\infty}^{\infty} G(w(x,\theta))\,f(x)\,dx,$$

$$(\widetilde{c}(\theta))' = \int_{-\infty}^{\infty} w'_\theta(x,\theta) g(w(x,\theta))\,f(x)\,dx,$$

$$(\widetilde{c}(\theta))'' = \int_{-\infty}^{\infty} w''_\theta(x,\theta) g(w(x,\theta))\,f(x)\,dx + \int_{-\infty}^{\infty} (w'_\theta)^2(x,\theta) g'(w(x,\theta))\,f(x)\,dx,$$

что дает нам

$$c(0) = 1/G(0); \quad c'(0) = 0; \quad c''(0) = -g'(0)/G^2(0) \int_{-\infty}^{\infty} (w'_\theta(0,x))^2\,f(x)\,dx.$$

Ясно, что при $\theta = 0$ получается $H(x,0) = F(x)$. Найдем значение $H'_\theta(x,0)$. Поскольку

$$H'_\theta(x,\theta) = \int\limits_{-\infty}^{x} \bigl(c'(\theta)\, G(w(u,\theta) + c(\theta)\, w'_\theta(u,\theta)\, g(w(u,\theta))\bigr)\, f(u)\, du,$$

получаем

$$H'_\theta(x,0) = \frac{g(0)}{G(0)} \int\limits_{-\infty}^{x} w'_\theta(u,0)\, f(u)\, du =: \frac{g(0)}{G(0)}\, F_w(x),$$

где для краткости обозначено

$$F_w(x) = \int\limits_{-\infty}^{x} w'_\theta(u,0)\, f(u)\, du, \quad x \in \mathbb{R}.$$

Нечетность функции $w(x,\theta)f(x)$ влечет за собой следующее свойство для всех $x \geqslant 0$:

$$\int\limits_{-\infty}^{x} w'_\theta(u,0)\, f(u)\, du = -\int\limits_{x}^{\infty} w'_\theta(u,0)\, f(u)\, du.$$

В частности,

$$F_w(-\infty) = F_w(\infty) = 0, \qquad F_w(-x) = F_w(x).$$

Теперь мы можем найти главные члены разложения в ряд Тейлора функции распределения при альтернативной гипотезе. Итак, при $\theta \to 0$ и любом $k > 1$ имеем:

$$\begin{aligned} H^k(x,\theta) &\sim F(x)^k + k\,\frac{g(0)}{G(0)}\, F_w(x)\, F^{k-1}(x)\cdot\theta, \\ H^k(-x,\theta) &\sim F(-x)^k + k\,\frac{g(0)}{G(0)}\, F_w(x)\, F^{k-1}(-x)\cdot\theta, \\ (1-H(x,\theta))^k &\sim F(-x)^k - k\,\frac{g(0)}{G(0)}\, F_w(x)\, F^{k-1}(-x)\cdot\theta, \\ (1-H(-x,\theta))^k &\sim F(x)^k - k\,\frac{g(0)}{G(0)}\, F_w(x)\, F^{k-1}(x)\cdot\theta. \end{aligned} \tag{7}$$

§3. Информация Кульбака–Лейблера

Обозначим через F ф. р. наблюдений при нулевой гипотезе, через H – ф. р. наблюдений при альтернативе, а соответствующие плотности пусть будут f и h. В качестве "информационного расстояния" между функциями распределения рассмотрим информацию Кульбака–Лейблера, которая играет важную роль в теории асимптотической проверки гипотез, см., например, [9]:

$$K(h,f) = \int\limits_{-\infty}^{\infty} h(x)\, \ln \frac{h(x)}{f(x)}\; dx.$$

(предполагается, что ф.р. H абсолютно непрерывна относительно F). В нашем случае для плотности симметричного распределения f и плотности скоса при альтернативе h, которая задается в (4), при проверке сложной основной гипотезы H можно использовать верхнюю границу для точных наклонов тестовых статистик (о которых говорится ниже), см. [28, гл. 4]:

$$\begin{aligned} K(\theta) = \int\limits_{-\infty}^{\infty} \ln \frac{2\,h(x,\theta)}{h(x,\theta)+h(-x,\theta)}\, h(x,\theta)\, dx \\ = \int\limits_{-\infty}^{\infty} c(\theta)\, G(w(\theta,x))\, \ln\left\{c(\theta)\, G(w(\theta,x))\right\}\; f(x)\, dx. \quad (8) \end{aligned}$$

Теперь получим значение главного члена этого выражения при $\theta \to 0$ при заданных условиях регулярности для G и w. Для этого выражение $c(\theta)\, G(w(\theta,x))$ разложим в ряд Тейлора в окрестности $\theta = 0$ до θ^2 и, учитывая свойства $c(\theta), w(x,\theta), F_w(x)$ и $G(x)$, сформулированные выше, получим асимптотику для информации Кульбака–Лейблера:

$$K(\theta) \sim \frac{g^2(0)}{2\,G^2(0)} \int\limits_{-\infty}^{\infty} (w'_\theta(x,0))^2\, f(x)\, dx \cdot \theta^2, \quad \theta \to 0. \qquad (9)$$

В частном случае, для классической модели скоса Аззалини (5) при $w(x,\theta) = x\theta$ это соотношение принимает вид

$$K(\theta) \sim 2\, g^2(0) \cdot \mathbf{E}_F\, X_1^2 \cdot \theta^2,$$

что совпадает с вычислениями в [15].

§4. Вычисление бахадуровской эффективности

Продолжим вычислять локальную бахадуровскую эффективность для статистик $J_n^{(k+1)}$ и $D_n^{(k)}$, см. (2) и (3). Более детальный разбор этой методики представлен в [9] и [28]. Отметим также, что ключевым понятием теории Бахадура является понятие точного наклона для последовательности тестовых статистик $\{T_n\}$. Эта величина показывает скорость экспоненциального убывания достигаемого уровня $\{T_n\}$ при альтернативе. Для последовательности $\{T_n\}$ при $n \to \infty$ должны быть выполнены два следующих свойства:

1. $T_n \to b_T(\theta)$, $n \to \infty$, по вероятности при альтернативе;
2. $n^{-1} \ln P(T_n \geqslant u) \to -\zeta(u)$, $n \to \infty$, при $\mathcal{H}$,

где $\zeta(u)$ – непрерывная функция на множестве значений $b_T(\theta)$. Тогда точный наклон Бахадура [9] определяется как $c_T(\theta) = 2\,\zeta(b_T(\theta))$, а абсолютная локальная эффективность Бахадура выражается через

$$e_T = \lim_{\theta \to 0+} \frac{c_T(\theta)}{2K(\theta)}, \tag{10}$$

где $K(\theta)$ было определено в (8).

При проверке условий (1) и (2) следует учитывать, что первое из них, как правило, является простым следствием закона больших чисел, тогда как второе условие (вид логарифмической асимптотики больших уклонений) нетривиально. В нашем случае эта асимптотика уже вычислена в [30]. Следующие разделы посвящены нахождению локальной эффективности интегральной и колмогоровской статистик.

§5. Эффективность интегральной статистики

Рассуждения, необходимые в начале этого раздела, большей частью повторяют изложенное в [30], поэтому мы ограничимся лишь кратким описанием.

Статистика (2) не зависит от распределения. Она представима в виде:

$$J_n^{(k+1)} = n^{-k-1} \sum_{1 \leqslant i_1, \ldots, i_{k+1} \leqslant n} \bigl(\mathbf{1}\{|\min(X_{i_1}, \ldots, X_{i_k})| < |X_{i_{k+1}}|\} \\ - \mathbf{1}\{|\max(X_{i_1}, \ldots, X_{i_k})| < |X_{i_{k+1}}|\}\bigr),$$

которая после симметризации сводится к V-статистике. Для нее асимптотика больших уклонений выражается, согласно [31], следующим образом:

$$\zeta_{k+1}(t) \sim \frac{t^2}{2(k+1)^2\sigma_{k+1}^2}, \qquad t \to 0,$$

где

$$\sigma_{k+1}^2 = \frac{1}{2^{2k-2}(k+1)^2}\int\limits_0^1 \left((1+s)^k + (1-s)^k - 2\right)^2 \, ds > 0.$$

Схожим образом в виде V-статистики представляется выражение:

$$b_J^{(k+1)}(\theta) = \mathbf{P}_\theta\{|\min(Y_1,\ldots,Y_k)| < |X|\} - \mathbf{P}_\theta\{|\max(Y_1,\ldots,Y_k)| < |X|\}. \quad (11)$$

Теперь нам нужно выделить главную часть (11) для $\theta \to 0$ при альтернативе (4). Заметим, что

$$\mathbf{P}_\theta\{|\max(Y_1,\ldots,Y_k)| < |Z|\} = \int\limits_0^\infty \left(H^k(x,\theta) - H^k(-x,\theta)\right) \, d(H(x,\theta) - H(-x,\theta)).$$

Это легко получить, проинтегрировав условное распределение максимумов при фиксированном $|X|$. Аналогичными рассуждениями получаем также

$$\mathbf{P}_\theta\{|\min(Y_1,\ldots,Y_k)| < |Z|\} = \int\limits_0^\infty \left((1 - H(-x,\theta))^k - (1 - H(x,\theta))^k\right) \, d(H(x,\theta) - H(-x,\theta)).$$

Используя полученные представления, имеем, обозначая для краткости $\overline{H}(x,\theta) := 1 - H(x,\theta)$, $H_0(x,\theta) := H(x,\theta) - H(-x,\theta)$,

$$b_J^{(k+1)}(\theta) = \int\limits_0^\infty \left(\overline{H}^k(-x,\theta) - \overline{H}^k(x,\theta) - H^k(x,\theta) + H^k(-x,\theta)\right) dH_0(x,\theta).$$

Из (7) следует, что

$$H_0(x,\theta) \sim F(x) - F(-x), \quad \theta \to 0,$$

а следовательно,

$$dH_0(x,\theta) \sim 2\, f(x)\, dx.$$

С помощью (7) получаем теперь асимптотику $b_J^{(k+1)}(\theta)$:

$$b_J^{(k+1)}(\theta) \sim 4\, k\, \theta\, \frac{g(0)}{G(0)} \int\limits_0^\infty F_w(x)\, (F^{k-1}(-x) - F^{k-1}(x))\, f(x)\, dx.$$

Учитывая, что

$$d/dx\ \left(F^k(x) + F^k(-x)\right) = k\, f(x) \left(F^{k-1}(x) - F^{k-1}(-x)\right),$$

мы можем интегрировать по частям. Следовательно, используя (2), можно получить

$$\begin{aligned} b_J^{(k+1)}(\theta) &\sim 4\, k\, \theta \frac{g(0)}{G(0)} \int\limits_0^\infty F_w(x)(F^{k-1}(-x) - F^{k-1}(x))\, f(x)\, dx \\ &\sim -4\, \theta\, \frac{g(0)}{G(0)} \int\limits_0^\infty F_w(x)\, d(F^k(-x) + F^k(x)) \\ &= 4\, \theta\, \frac{g(0)}{G(0)} \left(\frac{1}{2^{k-1}} F_w(0) + \int\limits_0^\infty (F^k(-x) + F^k(x))\, w'_\theta(x,0)\, f(x)\, dx \right) \\ &= 4\, \theta\, \frac{g(0)}{G(0)} \left(\int\limits_0^\infty \left(F^k(-x) + F^k(x) - \frac{1}{2^{k-1}} \right) w'_\theta(x,0)\, f(x)\, dx \right). \end{aligned}$$

Теперь, учитывая асимптотику (11), находим при $\theta \to 0$

$$c_J^{(k+1)}(\theta) \sim \frac{16\, g^2(0) \Big(\int\limits_0^\infty \left(F^k(-x) + F^k(x) - 1/2^{k-1} \right) w'_\theta(0,x)\, f(x)\, dx \Big)^2}{G^2(0)(k+1)^2 \sigma_{k+1}^2}\, \theta^2.$$

Наконец, локальная точная абсолютная эффективность Бахадура (10) последовательности интегральных статистик $\{J_n^{(k+1)}\}$ с учетом асимптотики (9) информации Кульбака–Лейблера выглядит следующим образом:

$$e_{J^{(k+1)}} = \frac{2^{2k+1}\left(\int\limits_0^\infty \left(F^k(-x) + F^k(x) - 1/2^{k-1}\right) w'_\theta(0,x)\, f(x)\, dx\right)^2}{\int\limits_0^\infty (w'_\theta(0,x))^2\, f(x)\, dx \int\limits_0^1 ((1+s)^k + (1-s)^k - 2)^2\, ds}. \tag{12}$$

§6. Эффективность колмогоровской статистики

Теперь займемся вычислением эффективности последовательности колмогоровских статистик $D_n^{(k)}$, заданной в (3). Как и в первом случае, они свободны от распределения при $\mathcal{H}$, а асимптотики больших уклонений получены в [30] на основе результата [29]. Там было доказано, что

$$\zeta_{k+1}(t) \sim \frac{t^2}{2k^2 s_k^2}, \qquad t \to 0,$$

где

$$s_k^2 = \sup_{0<s<1} (1-s)\left((1+s)^{k-1}/2^{k-1} - (1-s)^{k-1}/2^{k-1}\right)^2.$$

Вычисляя предел при альтернативе, авторы [30] получили, что

$$\begin{aligned} b_D^{(k)}(\theta) &= \lim_{n\to\infty} D_n^{(k+1)} \\ &= \sup_{x\geqslant 0} \left|\mathbf{P}_\theta\{|\min(Y_1,\ldots,Y_k)| < x\} - \mathbf{P}_\theta\{|\max(Y_1,\ldots,Y_k)| < x\}\right|. \end{aligned}$$

Аналогично рассуждениям в предыдущем разделе, можно получить для случая скошенной альтернативы, используя (7):

$$\begin{aligned} b_D^{(k)}(\theta) &= \sup_{x\geqslant 0}\left|H^k(x,\theta) + H_0^k(x,\theta) - H^k(-x,\theta) - H_0^k(-x,\theta)\right| \\ &\sim \theta\cdot 2k\frac{g(0)}{G(0)} \sup_{x\geqslant 0} |(F^{k-1}(x) - F^{k-1}(-x))\, F_w(x))|. \end{aligned}$$

Тогда точный наклон для колмогоровской статистики допускает асимптотику:

$$c_{D^{(k)}}(\theta) \sim \frac{4\, g^2(0)\, |\sup_{x\geqslant 0} |(F^{k-1}(x) - F^{k-1}(-x))\, F_w(x))|^2}{G^2(0)\, s_k^2}.$$

Наконец, ее локальная абсолютная бахадуровская эффективность равна

$$e_{D^{(k)}} = \frac{4\sup_{x\geqslant 0}|(F^{k-1}(-x) - F^{k-1}(x)) \int\limits_{-\infty}^{x} w'_\theta(0,u)\ f(u)du|^2}{s_k^2 \int\limits_0^\infty (w'_\theta(0,x))^2\ dF}. \qquad (13)$$

§7. Обсуждение найденных эффективностей

Хотя обобщенные скошенно-симметричные распределения представляют собой довольно широкий класс альтернатив, значение асимптотической эффективности и для интегрального, и для колмогоровского критерия зависит только от производной в нуле функции скоса $w'_\theta(0,x)$ и совершенно не зависит от "структурной" функции G.

Чтобы получить еще один полезный результат, связанный с эффективностью критериев, введем вспомогательную функцию $R(x) = 2F(x) - 1$. Тогда

$$F(x) = \frac{1+R(x)}{2}, \quad F(-x) = \frac{1-R(x)}{2},$$
$$dF(x) = \frac{1}{2}dR(x), \quad R(0) = 0, \quad R(\infty) = 1.$$

Теперь эффективности (12) и (13) представляются в виде:

$$e_{J^{(k+1)}} = \frac{\left(\int\limits_0^\infty \big((1+R(x))^k + (1-R(x))^k - 2\big)\, w'_\theta(0,x)\ dR(x)\right)^2}{\int\limits_0^\infty (w'_\theta(0,x))^2\ dR(x) \int\limits_0^\infty ((1+s)^k + (1-s)^k - 2)^2\ ds},$$

$$e_{D^{(k)}} = \frac{\sup\limits_{x>0} |((1+R(x))^{k-1} - (1-R(x))^{k-1}) \int\limits_{-\infty}^{x} w'_\theta(0,u)\ dR(u)|^2}{s_k^2 \int\limits_0^\infty (w'_\theta(0,x))^2\ dR(x)}.$$

Наконец, после некоторых упрощений, остается

$$e_{J^{(k+1)}} = \frac{\left(\int\limits_0^\infty \left(2\sum\limits_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l}R(x)^{2l}\right)w'_\theta(x,0)\,dR(x)\right)^2}{\int\limits_0^\infty (w'_\theta(x,0))^2\,dR(x)\int\limits_0^\infty\left(2\sum\limits_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l}R^{2l}(x)\right)^2\,dR(x)}, \tag{14}$$

$$e_{D^{(k)}} = \frac{\sup\limits_{x>0}\left|\left(2\sum\limits_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l-1}R(x)^{2l-1}\right)\int\limits_{-\infty}^{x} w'_\theta(u,0)\,dR(u)\right|^2}{\int\limits_0^\infty (w'_\theta(x,0))^2\,dR(x)\sup\limits_{x>0}(1-x)\left(2\sum\limits_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l-1}x^{2l-1}\right)^2}. \tag{15}$$

Из (14)–(15) видно, что при $k=2$ и $k=3$ значения эффективностей совпадают и для интегрального, и для колмогоровского критерия. Это вполне согласуется с результатами для этих критериев, полученными в [23] и в [30] для альтернатив сдвига.

Теперь приступим к вычислению локальной бахадуровской эффективности для нескольких стандартных распределений. Рассмотрим следующие симметричные плотности исходной выборки:

(1) Нормальную: $f_1(x) = (2\pi)^{-1/2}e^{-x^2/2}$;
(2) Логистическую: $f_2(x) = e^x/(1+e^x)^2$;
(3) Арксинус на $[-1,1]$: $f_3(x) = (\pi\sqrt{1-x^2})^{-1}\mathbf{1}\{-1<x<1\}$;
(4) Равномерную на $[-1,1]$: $f_4(x) = \frac{1}{2}\mathbf{1}\{-1<x<1\}$;
(5) Стьюдент-3: $f_5(x) = 2/(\pi(1+x^2)^2)$.

Эти распределения рассматривались в [14, 15]. Для каждой из выбранных плотностей f рассмотрим соответствующее скошенно-симметричное распределение (4) с произвольной функцией G и такой функцией уклонения w, что $w'(x,0)=x$, потому что этому условию соответствует наиболее распространенный случай $w(x,\theta)=x\theta$. Значения эффективностей получены по соответствующим формулам (12) и (13) с помощью системы Matlab.

Представим таблицу для $k = 2, 3, 4$. В скобках указаны (если известны) эффективности соответствующих тестов, полученные Никитиным и Ахсануллахом [30] для альтернатив сдвига.

Заметим прежде всего, что тесты показали довольно высокую эффективность даже для сравнительно редких распределений, поэтому

Таблица 1. Локальные эффективности интегрального и колмогоровского тестов

Тест	Интегральный		Колмогоровский	
Распределение	$k = 2,3$	$k = 4$	$k = 2,3$	$k = 4$
Нормальное	0,977 (0,977)	0,975 (0,975)	0,764 (0,764)	0,733 (0,733)
Логистическое	0,962 (0,938)	0,964 (0,925)	0,747 (0,750)	0,725 (0,696)
Арксинус	0,868	0,848	0,698	0,635
Равномерное	0,938	0,923	0,750	0,697
Стьюдент-3	0,766	0,777	0,585	0,306

мы можем их уверенно рекомендовать для проверки гипотезы симметрии при обобщенных скошенно-симметричных альтернативах. Более эффективный интегральный критерий требует бóльших вычислительных затрат. Он уже нашел свое применение в прикладных задачах. Стоит отметить, что иногда интегральный тест для $k = 4$ показывает лучшую эффективность, чем для $k = 2$ и 3.

Анализ данной таблицы и таблицы 2 из статьи [30] показывает, что значения эффективностей для разных критериев при альтернативе сдвига и скошенной альтернативе почти совпадают. Это удобно для прикладных задач: на практике структура распределения альтернативы редко известна, и данное свойство позволяет использовать лучшие критерии независимо от альтернативы. Еще одно интересное наблюдение состоит в том, что для нормального распределения при сдвиговой и скошенной альтернативах значения бахадуровской эффективности совпадают. Это свойство нормального распределения было отмечено в статье [15].

Мы подсчитали локальные эффективности и для скошенно-симметричных альтернатив с кубической функцией скоса (см. пример 4 в разделе 2) для нескольких распределений. Как видно, результаты оказались вполне сопоставимы.

Таблица 2. Эффективности для скошенных альтернатив с кубической функцией скоса.

Тест	Интегральный		Колмогоровский	
Распределение	$k=2,3$	$k=4$	$k=2,3$	$k=4$
Нормальное	0.670	0.689	0.548	0.575
Арксинус	0.943	0.928	0.778	0.721
Равномерное	0.991	0.985	0.803	0.767

§8. Наиболее благоприятная альтернатива

Следующий вопрос был поставлен Бахадуром и развит в [28, Гл. 6]. Какое распределение при альтернативе будет приводить к локальной асимптотической оптимальности (ЛАО) новых статистик по Бахадуру? Иными словами, какая альтернатива является наиболее благоприятной и обеспечивает единичную локальную эффективность, т.е. $e_{J^{(k+1)}} = 1$ или $e_{D^{(k)}} = 1$ для некоторого $k > 1$.

Возвращаясь к представлению эффективностей в форме (14)–(15), можно увидеть, что числитель и знаменатель связаны неравенством Коши. Действительно:

$$\left(\int_0^\infty \left(2\sum_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l}R^{2l}(x)\right)w'_\theta(x,0)\,dR(x)\right)^2 \leqslant \int_0^\infty (w'_\theta(x,0))^2\,dR(x)\int_0^\infty\left(2\sum_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l}R^{2l}(x)\right)^2 dR(x);$$

$$\sup_{x>0}\left|\left(2\sum_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l-1}R^{2l-1}(x)\right)\int_{-\infty}^{-x}w'_\theta(u,0)\,dR(u)\right|^2$$
$$= \sup_{x>0}\left(2\sum_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l-1}R^{2l-1}(x)\right)^2\left(\int_0^\infty \mathbf{1}\{u>x\}w'_\theta(u,0)\,dR(u)\right)^2$$
$$\leqslant \sup_{x>0}\left(2\sum_{l=1}^{\lfloor k/2\rfloor}\binom{k}{2l-1}R^{2l-1}(x)\right)^2\int_x^\infty dR(x)\int_0^\infty (w'_\theta(u,0))^2\,dR(u)$$

$$= \int_0^\infty (w'_\theta(x,0))^2 \, dR(x) \sup_{x>0}(1-x)\left(2\sum_{l=1}^{\lfloor k/2 \rfloor}\binom{k}{2l-1}x^{2l-1}\right)^2.$$

Таким образом, распределение асимптотически оптимально по Бахадуру, если в соответствующем неравенстве Коши достигается знак равенства, а это возможно при пропорциональности подынтегральных выражений. Но для колмогоровской статистики оптимальность означает выполнение

$$\mathbf{1}\{u > x\} = C(x) \cdot w'_\theta(u,0),$$

что невозможно из-за свойств функции w. Подобный результат был получен в случае классического колмогоровского критерия для скошенных альтернатив в [15]. Таким образом, область ЛАО пуста.

Для интегральных критериев оптимальность достигается, если и только если

$$\sum_{l=1}^{\lfloor k/2 \rfloor}\binom{k}{2l}R^{2l}(u) = C \cdot w'_\theta(u,0), \tag{16}$$

при какой-либо положительной константе C. Приведем пример оптимальных (наиболее благоприятных) альтернатив для $k = 2$ и $k = 3$. Плотность таких распределений имеет вид:

$$f_{2,3}(x) = C \cdot |w'_\theta(x,0)|^{-1/2}, \quad -b \leqslant x \leqslant b, \quad b > 0.$$

Если $w'_\theta(x,0) = x$, что отвечает классическому случаю $w(x,\theta) = x\theta$, то в качестве наиболее благоприятной альтернативы получается простое симметричное распределение с плотностью

$$f_{2,3}(x) = 1/4\sqrt{b|x|}, \quad -b \leqslant x \leqslant b,$$

которое никогда прежде не появлялось в подобных задачах. Из равенства (16) можно получить и более сложные примеры.

Заключение

Мы нашли теоретические значения локальной эффективности Бахадура для последовательности статистик (2) и (3), предложенных в [30], для проверки симметрии распределения против обобщенных скошенно-симметричных альтернатив и вычислили ее значения для некоторых наиболее распространенных случаев. Значения эффективности оказались весьма высоки, причем одни и те же критерии показали лучшие результаты как для альтернатив сдвига, так и для скошенных альтернатив. Всё это говорит о том, что новые критерии довольно удачны и могут применяться на практике. Кроме того, в статье обсуждается вопрос о наиболее благоприятных альтернативах из данного класса.

Литература

1. M. Ahsanullah, *On some characteristic property of symmetric distributions.* — Pakist. J. Statist. **8** (1992), 171–179.
2. R. B. Arellano-Valle, H. W. Gomez, E. A. Quintana, *A new class of skew-normal distributions.* — Commun. Statist. Theor. Meth. **33** (2004), 1465–1480.
3. A. Azzalini, *A class of distributions which includes the normal ones.* — Scand. J. Statist. **12** (1985), 171–178.
4. A. Azzalini, *The skew-normal distribution and related multivariate families.* — Scand. J. Statist. **32** (2005), 159–188.
5. A. Azzalini, A. Capitanio, *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.* — J. Roy. Statist. Soc. **B65** (2003), 367–389.
6. A. Azzalini, A. Dalla Valle, *The multivariate skew-normal distribution.* — Biometrika **83** (1996), 715–726.
7. A. Azzalini, G. Regoli, *The work of Fernando de Helguero on non-normality arising from selection.* — Chilean J. Statist. **3** (2012), No. 2, 113–128.
8. A. Azzalini with the collaboration of A. Capitanio, *The Skew-Normal and Related Families.* Cambridge University Press, Cambridge, 2013.
9. R. R. Bahadur, *Some Limit Theorems in Statistics.*, SIAM, Philadelphia, 1971.
10. L. Baringhaus, N. Henze, *A characterization of and new consistent tests for symmetry.* — Commun. Statist. Theor. Meth. **21** (1992), 1555–1566.
11. D. Cassart, M. Hallin, D. Paindaveine, *Optimal detection of Fechner–asymmetry.* — J. Statist. Plann. Infer. **138** (2008), No. 8, 2499–2525.
12. S. K. Chatterjee, P. K. Sen, *On Kolmogorov–Smirnov-type tests for symmetry.* — Ann. Inst. Statist. Math. **25** (1973), 287–299.
13. F. De Helguero, *Sulla rappresentazione analitica delle curve abnormali.* In G. Castelnuovo (Ed.), Atti del IV Congresso Internazionale dei Matematici (Roma, 611 Aprile 1908) **III** (1909), sezione III-B, 288–299. Roma: R. Accademia dei Lincei.
14. A. Durio, Ya. Yu. Nikitin, *On asymptotic efficiency of certain distribution-free symmetry tests under skew alternatives.* Studi in Onore di Angelo Zanella, a cura di B. V. Frosini, U. Magagnoli, G. Boari, Vita e Pensiero, Milano, 2002, 223–239.
15. A. Durio, Ya. Yu. Nikitin, *Local asymptotic efficiency of some goodness-of-fit tests under skew alternatives.* — J. Statist. Plann. Infer. **115** (2003), 171–179.
16. A. Durio, Ya. Yu. Nikitin, *Local efficiency of integrated goodness-of-fit tests under skew alternatives.* — Statist. Probab. Lett. **117** (2016), 136–143.
17. R. Helmers, P. Janssen, R. Serfling, *Glivenko–Cantelli properties of some generalized empirical DF's and strong convergence of generalized L – statistics.* — Probab. Theor. Relat. Fields **79** (1988), 75–93.

18. T. Hettsmansperger, *Statistical Inference based on Ranks*, Wiley, New York, 1984.
19. M. Jovanovic, B. Miloševic, Ya. Yu. Nikitin, M. Obradovic, K. Yu. Volkova, *Tests of exponentiality based on Arnold-ЦVillaseñor characterization and their efficiencies.* — Comp. Statist. Data Anal. **90** (2015), 100–113.
20. E. L. Lehmann, H. J. D' Abrera, *Nonparametrics: Statistical Methods based on Ranks.* Springer, New York, 2006.
21. C. Ley, *Flexible modelling in statistics: past, present and future.* — J. Soc. Franç. Statist. **156** (2015), 76–97.
22. C. Ley, D. Paindaveine, *Le Cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions.* — J. Nonpar. Statist. **21** (2009), No. 8, 943–967.
23. V. V. Litvinova, *New nonparametric test for symmetry and its asymptotic efficiency.* — Vestnik St. Petersb. Univ. Matem. **34** (2001), 12–14.
24. Y. Ma, M. G. Genton, *Flexible class of skew–symmetric distributions.* — Scand. J. Statist. **31** (2004), No. 3, 459–468.
25. B. Miloševic, M. Obradovic, *Characterization based symmetry tests and their asymptotic efficiencies.* — Statist. Probab. Lett. **119** (2016), 155–162.
26. K. Morris, D. Szynal, *Goodness–of–fit tests using characterizations of continuous distributions.* — Appl. Math. (Warsaw) **28** (2001), 151–168.
27. P. Muliere, Ya. Yu. Nikitin, *Scale–invariant test of normality based on Polya's characterization.* — Metron **60** (2002), 21–33.
28. Ya. Nikitin, *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, New York, 1995.
29. Ya. Yu. Nikitin, *Large deviations of U–empirical Kolmogorov–Smirnov test, and their efficiency.* — J. Nonpar. Statist. **22** (2010), 649–668.
30. Ya. Yu. Nikitin, M. Ahsanullah, *New U-empirical tests of symmetry based on extremal order statistics, and their efficiencies.* In: Mathematical Statistics and Limit Theorems: Festschrift in Honor of Paul Deheuvels, Springer, 2015, 231–248.
31. Ya. Yu. Nikitin, E. V. Ponikarov, *Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics.* — Proc. of Saint-Petersburg Math. Soc. **7** (1999), 124–167; English translation in AMS Translations, ser. 2, **203** (2001), 107–146.
32. Ya. Yu. Nikitin, K. Yu. Volkova, *Asymptotic efficiency of exponentiality tests based on order statistics characterization.* — Georgian Math. J. **17** (2010), No. 4, 749–763.
33. M. Obradovic, M. Jovanovic, B. Miloševic, *Goodness–of–fit tests for Pareto distribution based on a characterization and their asymptotics.* — Statistics **49** (2015), No. 5, 1026–1041.
34. K. Yu. Volkova, Y. Y. Nikitin, *On the asymptotic efficiency of normality tests based on the Shepp property.* — Vestnik St. Petersb. Univ. Matem. **42** (1999), No. 4, 256–261.
35. J. Wang, J. Boyer, M. G. Genton, *A skew–symmetric representation of multivariate distribution.* — Statist. Sinica **14** (2004), 1259–1270.

Bookiya G. T., Nikitin Ya. Yu. Asymptotic efficiency of new distribution-free tests of symmetry for generalized skew alternatives.

We calculate Bahadur efficiency of new nonparametric tests for symmetry recently proposed by Nikitin and Ahsanullah [30]. In contrast with that paper where only location alternatives were discussed, we are interested in generalized skew alternatives. It is shown that the new tests are highly efficient for a vast class of skew alternatives. The problem of most favorable alternatives is also explored.

Поступило 28 октября 2016 г.

Филологический факультет,
С.-Петербургский государственный университет,
Университетская наб. 7/9,
199034 С.-Петербург, Россия
E-mail: gregorybookia@yandex.ru

Математико-механический факультет,
С.-Петербургский государственный университет,
Университетская наб. 7/9,
199034 С.-Петербург, Россия
E-mail: y.nikitin@spbu.ru