Записки научных
семинаров ПОМИ
Том 351, 2007 г.

В. В. Литвинова, Я. Ю. Никитин

ОБ АСИМПТОТИЧЕСКОЙ ЭФФЕКТИВНОСТИ И ЛОКАЛЬНОЙ ОПТИМАЛЬНОСТИ КРИТЕРИЕВ, ОСНОВАННЫХ НА ДВУХВЫБОРОЧНЫХ U– И V–СТАТИСТИКАХ

1. Введение

Многие непараметрические критерии согласия и симметрии основаны на U- и V-статистиках. Примерами могут служить критерий Вилкоксона, критерий Маесоно [9], критерии ω^2 Крамера-фон Мизеса–Смирнова и U^2 Ватсона и многие другие. Теория U- и V-статистик хорошо разработана и излагается, например, в [8].

Для сравнения статистических тестов в больших выборках обычно применяется тот или иной вид асимптотической эффективности. Результаты, полученные в этой области, изложены в монографии [11], где бахадуровская асимптотическая эффективность используется чаще других. Как известно, ключевым для вычисления этой эффективности является знание асимптотики больших уклонений черновского типа тестовых статистик при основной гипотезе.

Большие уклонения черновского типа для U- и V-статистик стали изучаться сравнительно недавно. Эйхельсбахер и Лёве [6] доказали для них принцип больших уклонений, но этот результат недостаточен для вычисления асимптотической эффективности.

Явная форма для асимптотики вероятностей больших уклонений одновыборочных U-статистик была найдена в работах [12, 13]. Это позволило вычислить локальную бахадуровскую эффективность для ряда конкретных критериев согласия и симметрии, основанных на U- и V-статистиках. Позже в работе [15] была описана так называемая область локальной асимптотической оптимальности (ЛАО) критериев, основанных на U- и V-статистиках. Это означает, что находятся те распределе-

Работа поддержана грантами РФФИ 06-01-00179 и 07-01-00159.

ния наблюдений, для которых данная последовательность тестовых статистик является локально асимптотически оптимальной в бахадуровском смысле. Важность этой “обратной” задачи была отмечена Бахадуром [1] и Сэвиджем [16], но конкретные результаты были впервые установлены в [10, 11]. В [15] подобные результаты были получены не для конкретных ядер, а для целого их класса.

Критерии однородности, основанные на двухвыборочных U- и V-статистиках, изучены значительно меньше. В [14] были найдены явные выражения для асимптотики вероятностей больших уклонений невырожденных двухвыборочных U- и V-статистик с ограниченными ядрами. Это позволяет вычислить в настоящей работе асимптотическую эффективность соответствующих критериев однородности и описать их области ЛАО. Заметим, что питменовская эффективность рассматриваемых критериев совпадает с их локальной бахадуровской эффективностью в силу асимптотической нормальности невырожденных U- и V-статистик [1, 11].

2. Большие уклонения двухвыборочных U- и V-статистик

Для удобства читателя приведем здесь основной результат из [14]. Пусть $X_1, X_2, \dots, X_{n_1}$ и $Y_1, Y_2, \dots, Y_{n_2}$ две независимые выборки с одной и той же непрерывной функцией распределения (ф.р.) F на $\mathbb{R}^1$, обозначим $n = n_1 + n_2$ и пусть объемы выборок n_1 и n_2 стремятся к бесконечности так, что

$$\lim_{n_1, n_2 \to \infty} \frac{n_1}{n} = r_1 > 0, \qquad \lim_{n_1, n_2 \to \infty} \frac{n_2}{n} = r_2 > 0.$$

Пусть $\Phi : \mathbb{R}^{m_1 + m_2} \to \mathbb{R}^1$ – вещественное ядро, симметричное относительно перестановок первых m_1 и последних m_2 аргументов. Для этого ядра двухвыборочная U-статистика степени (m_1, m_2) определяется формулой

$$U_{n_1, n_2} = (C_{n_1}^{m_1})^{-1} \ (C_{n_2}^{m_2})^{-1} \sum \Phi(X_{i_1}, \dots, X_{i_{m_1}}; Y_{j_1}, \dots, Y_{j_{m_2}}),$$

где суммирование берется по множеству индексов

$$\{1 \leqslant i_1 < \dots < i_{m_1} \leqslant n_1, 1 \leqslant j_1 < \dots < j_{m_2} \leqslant n_2\},$$

а функционал Мизеса, или V – статистика степени (m_1, m_2), задается соотношением

$$V_{n_1,n_2} = n_1^{-m_1} n_2^{-m_2} \sum_{i_1=1}^{n_1} \cdots \sum_{i_{m_1}=1}^{n_1} \sum_{j_1=1}^{n_2} \cdots \sum_{j_{m_2}=1}^{n_2} \Phi(X_{i_1}, \ldots, X_{i_{m_1}}; Y_{j_1}, \ldots, Y_{j_{m_2}}).$$

Предположим, что $|\Phi(s;t)| \le M < \infty$ для всех значений векторов s и t. Это условие не слишком ограничительно, поскольку у критериев, основанных на U- и V-статистиках, ядра обычно являются комбинациями индикаторных функций, и потому ограничены.

При изучении больших уклонений можно считать, что исходные выборки имеют равномерное распределение U на $[0,1]$. В случае необходимости мы всегда можем перейти от исходных случайных величин $\Phi(X_1, \ldots, X_{m_1}; Y_1, \ldots, Y_{m_2})$ к функциям от равномерных случайных величин U_j, $j = 1, \ldots, m_1 + m_2$, а именно

$$\tilde{\Phi}(U_1, \ldots, U_{m_1+m_2}) = \Phi(F^{-1}(U_1), \ldots, F^{-1}(U_{m_1+m_2})). \tag{1}$$

Таким образом U- и V-статистики с ядром Φ, зависящим от X_i и Y_j, могут быть преобразованы в U- или V-статистики с новым ограниченным ядром $\tilde{\Phi}$, зависящим от равномерно распределенных величин.

Переход от исходного ядра Φ к “приведенному” ядру $\tilde{\Phi}$ по формуле (1) весьма естественен для непараметрических тестовых статистик и аналогичен переходу от совместной ф.р. случайного вектора к соответствующей копуле.

Предположим, кроме того, что ядро Φ невырождено по отношению к обеим группам переменных. Это означает [8], что

$$\sigma_1^2 = E\psi_1^2(X_1) > 0, \quad \sigma_2^2 = E\psi_2^2(Y_1) > 0, \tag{2}$$

где

$$\psi_1(x) = E\{\Phi(X_1, \ldots, X_{m_1}; Y_1, \ldots, Y_{m_2}) \mid X_1 = x\}$$

и

$$\psi_2(y) = E\{\Phi(X_1, \ldots, X_{m_1}; Y_1, \ldots, Y_{m_2}) \mid Y_1 = y\}$$

являются одномерными проекциями ядра Φ. Например, для ядра статистики Вилкоксона $\Phi(s,t) = \mathbf{1}\{s < t\} - 1/2$ имеем

$$\psi_1(x) = 1/2 - F(x),\ \psi_2(y) = F(y) - 1/2,\ \tilde{\psi}_1(x) = 1/2 - x,\ \tilde{\psi}_2(y) = y - 1/2.$$

Основным результатом [15] является следующая теорема.

Теорема 1. *Пусть $|\Phi| \le M < \infty$, $E\Phi = 0$ и Φ невырождено, т.е. выполнено (2). Тогда справедливо соотношение*

$$\lim_{n\to\infty} n^{-1} \log \Pr\{V_{n_1,n_2} \geqslant a\} = \sum_{j=2}^{\infty} b_j a^j,$$

где ряд с числовыми коэффициентами b_j в правой части сходится при достаточно малых $a > 0$, причем

$$b_2 = -\frac{1}{2}\left(\frac{m_1^2\sigma_1^2}{r_1} + \frac{m_2^2\sigma_2^2}{r_2}\right)^{-1}.$$

Результат справедлив и для U-статистик с тем же ядром Φ.

Этот результат позволяет вычислить локальную бахадуровскую эффективность критериев, основанных на U- и V- статистиках.

3. Бахадуровские точные локальные наклоны

Рассмотрим проверку гипотезы однородности двух выборок против альтернативы, при которой первая выборка $X_1, \dots, X_{n_1}$ имеет плотность $f(x, 0)$, а вторая выборка $Y_1, \dots, Y_{n_2}$ имеет плотность $f(x, \theta)$, индексированную вещественным параметром $\theta \in \Theta = [0, \theta^*]$, $\theta^* > 0$. Пусть $F(x,\theta)$ – ф.р., отвечающая этой плотности, и $F(x, 0) = F(x)$.

Пусть тестовая статистика - это U- или V-статистика U_n с ядром Φ и критическими большими значениями. Обозначим

$$b_\Phi(\theta) := \mathbf{E}_{H_1}\Phi(X_1, \dots, X_{m_1}; Y_1, \dots, Y_{m_2})$$
$$= \int\limits_{R^{m_1}} \int\limits_{R^{m_2}} \Phi(x_1, \dots, x_{m_1}; y_1, \dots, y_{m_2}) \prod_{i=1}^{m_1} f(x_i; 0) dx_i \prod_{j=1}^{m_2} f(y_j; \theta) dy_j.$$

В дальнейшем мы предполагаем, что семейство $f(x,\theta)$ согласовано с ядром $\Phi(X_1, \dots, X_{m_1}; Y_1, \dots, Y_{m_2})$ так, что

$$b_\Phi(\theta) > 0, \tag{3}$$

для любого $\theta \in \Theta$. Это обеспечивает состоятельность теста, основанного на U_n. Мы предполагаем также справедливость естественного условия $b_\Phi(\theta) \to 0$ при $\theta \to 0$.

По закону больших чисел для U- и V-статистик, см. [8], при H_1 справедлива сходимость по вероятности

$$U_n \to b_\Phi(\theta), \quad \theta > 0. \tag{4}$$

Мерой бахадуровской эффективности последовательности $\{U_n\}$ является бахадуровский наклон $c_U(\theta)$. Он может быть найден с помощью следующего предложения ([2], §7; [11], гл. 1).

Теорема 2. *Пусть выполнены два следующих условия:*

$$\text{a)} \quad U_n \xrightarrow{P_\theta} b(\theta), \quad \theta > 0,$$

где $-\infty < b(\theta) < \infty$, *и* $\xrightarrow{P_\theta}$ *означает сходимость по вероятности при выполнении* H_1*;*

$$\text{b)} \quad \lim_{n\to\infty} n^{-1} \ \ln \ P_{H_0}(U_n \geqslant t\ ) \ = \ -h(t)$$

для любого t *из открытого интервала* I, *на котором* h *непрерывна и* $\{b(\theta), : \theta > 0\} \subset I$. *Тогда*

$$c_U(\theta) \ = \ 2\ h(b(\theta)).$$

Из (3), (4), а также из теоремы 1 вытекает, что условия теоремы 2 выполнены для статистики U_n. Поэтому мы можем написать для главной части точного наклона этой статистики при условии (3) следующее соотношение: при $\theta \to 0$

$$c_U(\theta) \sim (\frac{m_1^2\sigma_1^2}{r_1} + \frac{m_2^2\sigma_2^2}{r_2})^{-1}\, b_\Phi^2(\theta). \tag{5}$$

Для дальнейшего понадобится обозначение

$$\Phi^*(y_1,\ldots,y_{m_2}) := \int\limits_{R^{m_1}} \Phi(x_1,\ldots,x_{m_1};y_1,\ldots,y_{m_2}) \prod_{j=1}^{m_1} f(x_j,0)dx_j.$$

Мы упростим выражение для $b_\Phi(\theta)$ при $\theta \to 0$ и при некоторых условиях регулярности. Опишем поведение функции правдоподобия

$$L(z;\theta) = L(z_1,\ldots,z_m;\theta) = \Pi_{j=1}^m f(z_j,\theta),$$

как функции от θ. Для простоты производная f'_θ обозначается f_θ. С помощью разложения Тейлора с остаточным членом в интегральной форме (см. [5], теор. 8.14.3), получаем следующую формулу:

$$L(z;\theta) = L(z;0) + \theta L_\theta(z;0) + \frac{1}{2}\theta^2 \int_0^1 t^2(1-t)^2 L_{\theta\theta}(z;t\theta)dt. \tag{6}$$

Нам понадобятся точные выражения для производных функции $L(z;\theta)$ по θ. Ясно, что

$$L(z;0) = \Pi_{j=1}^m f(z_j,0), \quad L_\theta(z;0) = \sum_{i=1}^m f_\theta(z_i,0)\Pi_{j\neq i}^m f(z_j,0).$$

Точная формула для $L_{\theta\theta}(z;\tau)$ для любого $\tau \in [0,\theta^*)$ имеет вид

$$L_{\theta\theta}(z;\tau) = \sum_{i=1}^m f_{\theta\theta}(z_i,\tau)\Pi_{j\neq i} f(z_j,\tau) + \sum_{i\neq j}^m f_\theta(z_i,\tau) f_\theta(z_j,\tau)\Pi_{k\neq i,j} f(x_k,\tau).$$

Введем следующее условие регулярности. *Условие ND: существует $\delta > 0$ такое, что для любого $0 \le \theta \le \delta$*

$$\int_0^1 t^2(1-t)^2 \int_{R^{m_2}} \Phi^*(y_1,\ldots,y_{m_2}) f_{\theta\theta}(y_1,t\theta)\Pi_{j\neq 1} f(y_j,t\theta)dy_1\ldots dy_{m_2}dt < \infty,$$

$$\int_0^1 t^2(1-t)^2 \int_{R^{m_2}} \Phi^*(y_1,\ldots,y_{m_2}) f_\theta(y_1,t\theta) f_\theta(y_2,t\theta) \times \Pi_{k\neq 1,2} f(y_k,t\theta)dy_1\ldots dy_{m_2}dt < \infty.$$

Поскольку Φ ограничено, то условие ND выполнено если, например, для $0 \le \theta \le \delta$

$$\int_{R^1} |f_\theta(x,\theta)|dx < \infty, \quad \int_{R^1} |f_{\theta\theta}(x,\theta)|dx < \infty. \tag{7}$$

Условия (7) выполняются для многих семейств распределений.

С помощью тейлоровского разложения (6) и условия ND легко получаем

$$b_\Phi(\theta) \sim m_2 \int\limits_{R^1} \psi_2(y) f_\theta(y,0)dy \cdot \theta,\ \theta \to 0. \tag{8}$$

Используя асимптотическое представление (8) вместе с (5), находим следующие формулы для локальных точных наклонов $c_U(\theta)$.

Теорема 3. *Предположим, что $\{U_n\}$ является последовательностью невырожденных U- или V-статистик с ограниченными ядрами и что* (3) *и условие ND справедливы. Тогда, при $\theta \to 0$*

$$c_U(\theta) \sim \frac{m_2^2}{\frac{m_1^2\sigma_1^2}{r_1} + \frac{m_2^2\sigma_2^2}{r_2}} \Big(\int\limits_{R^1} \psi_2(x) f_\theta(x,0)dx \Big)^2 \cdot \theta^2. \tag{9}$$

Несимметричность этой формулы обусловлена тем, что ядро может быть несимметричным по обеим группам переменных, а альтернатива действует на вторую группу.

4. Вычисления локальной эффективности

Напомним теоретическое неравенство для точных наклонов двухвыборочных критериев однородности ([11],§3.3):

$$c_U(\theta) \le 2K(\theta), \text{ при } \theta > 0, \tag{10}$$

где $K(\theta)$ – взвешенная информация Кульбака–Лейблера

$$K(\theta) = r_1 \int\limits_{R^1} \ln \frac{f(x,0)}{h(x,\theta)} f(x,0)dx + r_2 \int\limits_{R^1} \ln \frac{f(x,\theta)}{h(x,\theta)} f(x,\theta)dx,$$

где для краткости обозначено $h(x,\theta) = r_1 f(x,0) + r_2 f(x,\theta)$.

Вполне естественно определить нелокальную бахадуровскую эффективность (БЭ) отношением $c_U(\theta)/2K(\theta)$. К сожалению, для ненулевых θ вычисления трудны или невозможны. Поэтому мы вычисляем ниже локальную бахадуровскую эффективность, которая является пределом нелокальной БЭ при $\theta \to 0$.

В типичных случаях при регулярных семействах плотностей $f(x,\theta)$ справедлива асимптотика (см. [11], §3.3):

$$2K(\theta) \sim r_1 r_2 I(f) \cdot \theta^2, \text{ при } \quad \theta \to 0, \tag{11}$$

где $I(f) = \int\limits_{R^1} [f_\theta^2(x,0)/f(x,0)]dx$ есть фишеровская информация. Тогда локальная БЭ $e_U^B(f)$ может быть записана так: при условиях теоремы 3 и (11)

$$e_U^B(f) = \Big(m_2 \int\limits_{R^1} \psi_2(x) f_\theta(x,0)dx\Big)^2 / (m_1^2\sigma_1^2 r_2 + m_2^2\sigma_2^2 r_1) I(f) \qquad (12)$$

Формула (12) справедлива для широкого класса ядер, основанных на U- и V-статистиках, и обобщает формулы для локальных эффективностей таких тестов в случае конкретных ядер из [11].

Рассмотрим несколько примеров использования полученной теоремы. Начнем со статистики Вилкоксона W, у которой

$$m_1 = m_2 = 1,\ \sigma_1^2 = \sigma_2^2 = 1/12,\ -\psi_1(x) = \psi_2(x) = F(x) - 1/2.$$

Большие уклонения этой статистики описываются теоремой 1 с $b_2 = -6$. Тогда локальный точный наклон имеет вид

$$c_W(\theta) \sim 12 r_1 r_2 \Big(\int\limits_{R^1} (F(x) - 1/2)) f_\theta(x) dx\Big)^2\ \theta^2,\ \theta \to 0.$$

В случае, когда плотность задается на всей оси, а альтернатива сдвиговая, это эквивалентно хорошо известной классической формуле

$$c_W(\theta) \sim 12 r_1 r_2 \Big(\int\limits_{R^1} f^2(y)dy\Big)^2\ \theta^2,\ \theta \to 0.$$

Рассмотрим статистику Дешпанде и Кочара, предложенную ими в [4] для проверки различия двух выборок в сдвиге. В простейшем случае $m = 2$ это двухвыборочная U-статистика с ядром

$$\Phi(s_1, s_2; t_1, t_2) = \mathbf{1}\{ttss \,\text{или}\, stts\} - \mathbf{1}\{sstt \,\text{или}\, tsst\},$$

где обозначение $\{ttss\}$ означает

$$\{t_1 \le t_2 \le s_1 \le s_2\} \cup \{t_1 \le t_2 \le s_2 \le s_1\}$$
$$\cup \{t_2 \le t_1 \le s_1 \le s_2\} \cup \{t_2 \le t_1 \le s_2 \le s_1\},$$

а другие перестановки определяются аналогично.

Вычисления, проведенные в [14], показывают, что

$$-\tilde{\psi}_1(t) = \tilde{\psi}_2(t) = -\frac{4}{3}t^3 + 2t^2 - 2t + \frac{2}{3}, \sigma_1^2 = \sigma_2^2 = \frac{38}{315}.$$

Рассмотрим в качестве альтернативы к гипотезе однородности логистическое семейство со сдвигом, т. е.

$$F(x,\theta) = \exp(x-\theta)/(1+\exp(x-\theta)).$$

Тогда $I(f) = 1/3, f_\theta(x,0) = \exp(x)(1-\exp(x))/(1+\exp(x))^3$. Условия (7) выполнены. Поэтому

$$e^B_{DK}(f) = \frac{945}{38}\Big(\int\limits_{R^1}\big(-\frac{4}{3}F^3(x) + 2F^2(x) - 2F(x) + \frac{2}{3}\big)f_\theta(x,0)dx\Big)^2$$

$$= \frac{189}{190} \approx 0.9947,$$

что является очень высоким значением.

Еще один пример – это статистика Кочара–Гупта степени $(2,2)$

$$\Phi(s_1,s_2;t_1,t_2) = \mathbf{1}\{stts\} - \mathbf{1}\{tsst\},$$

предложенная в [7] для проверки различия в масштабе. Рассмотрим в качестве альтернативы экспоненциальное распределение с параметром масштаба, т. е. с ф.р.

$$G(x,\theta) = 1 - \exp(-x\exp(\theta)), x \geq 0, \theta \geq 0.$$

В этом случае $g_\theta(x,0) = \exp(-x)(1-x)$, $I(g) = 1$. Вычисления, проведенные в [14], дают

$$-\tilde{\psi}_1(t) = \tilde{\psi}_2(t) = -2t^2 + 2t - \frac{1}{3}, \sigma_1^2 = \sigma_2^2 = \frac{1}{45}.$$

Поэтому

$$e^B_{KG}(f) = 45\Big(\int\limits_{R^+}\big(-2G^2(x,0) + 2G(x,0) - \frac{1}{3}\big)\exp(-x)(1-x)dx\Big)^2$$

$$= \frac{5}{36} \approx 0.1389.$$

Видно, что в этом примере эффективность невысока.

5. Условия локальной асимптотической оптимальности

Мы рассмотрим здесь для простоты только самый интересный и важный случай таких ядер Φ, у которых

$$m_1 = m_2 = m, \quad r_1 = r_2 = r, \quad -\psi_1 = \psi_2 = \psi, \quad \sigma_1^2 = \sigma_2^2 = \sigma^2. \quad (13)$$

Почти все примеры двухвыборочных U- и V-статистик, рассмотренные в [14], удовлетворяют этому условию.

Для заданного ядра Φ обозначим через $\mathcal{F}\ (\Phi)$ множество плотностей $f(x,\theta)$ с ф.р. $F(x,\theta)$, для которых выполняются условия регулярности из теоремы 3 и (11). Предположим также, что $\lim_{x\to-\infty} F_\theta(x,0) = 0$ и что при почти всех x

$$F_{x\theta}(x,0) = F_{\theta x}(x,0). \quad (14)$$

Вернемся к неравенству (10). Деля обе его части на θ^2 и переходя к пределу при $\theta \to 0$, получим с учетом (9), (11) и (13), что для всех семейств плотностей из $\mathcal{F}\ (\Phi)$

$$\left(m\int\limits_{R^1} \psi(x) f_\theta(x,0)dx\right)^2 \le m^2\sigma^2 I(f), \quad (15)$$

Как известно, [1, 11], локальная асимптотическая оптимальность в бахадуровском смысле последовательности статистик $\{U_n\}$ означает, что в этом неравенстве достигается равенство.

Те распределения, для которых это выполнено, образуют область ЛАО в классе $\mathcal{F}(\Phi)$. Для описания этой области полезно понятие ведущей функции последовательности статистик $\{T_n\}$, введенное и изученное в [11]. Ведущей функцией называется вещественная функция $v_T(x)$, определенная на $[0,1]$, удовлетворяющая граничным условиям $v_T(0) = v_T(1) = 0$, и такая, что условие $f(x,\theta) \in \mathcal{F}(\Phi)$ эквивалентно условию $F_\theta(F^{-1}(x),0) = Cv_T(x)$ для некоторой постоянной C и почти всех x.

Условия достижения равенства в неравенстве (15) находятся так. По неравенству Коши–Буняковского–Шварца

$$\big(m\int\limits_{R^1} \psi(x) f_\theta(x,0)dx\big)^2 \le m^2 \int\limits_{R^1} \psi^2(x) f(x,0)dx \cdot I(f) = m^2\sigma^2 I(f).$$

Следовательно, для выполнения ЛАО должно соблюдаться равенство в неравенстве Коши–Буняковского–Шварца, что имеет место в том и только том случае, когда для некоторой константы C_1 и почти всех x

$$f_\theta(x,0) = C_1\psi(x)f(x,0), \tag{16}$$

или эквивалентным образом

$$\frac{d\ln f(x,\theta)}{d\theta}|_{\theta=0} = C_1\psi(x) = C_1\tilde{\psi}(F(x)). \tag{17}$$

Уравнение (16) можно в силу (14) также переписать в виде

$$F_\theta(x,0) = C_1\int\limits_{-\infty}^{x}\psi(x)dF(x),$$

что эквивалентно

$$v_U(x) := F_\theta(F^{-1}(x),0) = C_1\int\limits_0^x \psi(F^{-1}(u))du$$

$$= C_1\int\limits_0^x \tilde{\psi}(u)du = C_1\tilde{\Psi}(x).$$

Таким образом ведущая функция последовательности статистик $\{U_n\}$ равна $\tilde{\Psi}(x) = \int_0^x \tilde{\psi}(u)du$. Этот результат можно сформулировать так.

Теорема 4. *Область ЛАО в $\mathcal{F}(\Phi)$ для невырожденных U- и V-статистик, удовлетворяющих условию теоремы 1, определяется равенством* (16) *или* (17)*, и ведущая функция есть* $\tilde{\Psi}(x)$.

В качестве примера рассмотрим двухвыборочную статистику Вилкоксона, у которой ведущая функция есть $x(1-x)$. Если для простоты считать, что альтернатива сдвиговая, то условие оптимальности приводит к дифференциальному уравнению в ф.р.

$$F' = C_2F(1-F),$$

решением которого является, как известно, логистическое семейство. Этот факт хорошо известен, см. [11], §6.4.

Применим теперь этот результат для отыскания условий оптимальности статистики Кочара-Гупты [7, 14]. В этом случае в силу [14]

$$\tilde{\psi}(s) = 2s^2 - 2s + \frac{1}{3},$$

так что ведущая функция с учетом граничных условий имеет вид $s(\frac{1}{2}-s)(1-s)$. Предположим теперь,что альтернатива - масштабная на полуоси, т.е. $F(x,\theta) = F(x\exp(\theta)), x, \theta \geq 0$. Тогда условие оптимальности сводится к дифференциальному уравнению в ф.р.

$$xF'(x) = C_3 F(x)\Big(\frac{1}{2} - F(x)\Big)(1-F(x)) \quad x \geq 0, \tag{18}$$

которое не может иметь решения в ф.р., поскольку правая часть не сохраняет знак на $(0,\infty)$. Поэтому область ЛАО в этом случае пуста.

В качестве третьего примера рассмотрим статистику Дешпанде–Кочара из [4]. В этом случае $\tilde{\psi}(x) = \frac{4}{3}x^3 - 4x^2 + 2x$, а тогда ведущая функция есть $x^2(1-x)(3-x)$. Снова рассматривая альтернативу сдвига, мы видим, что уравнение для ф.р.

$$F' = C_4 F^2(1-F)(3-F)$$

имеет решение, которое, впрочем, не выражается явно. Тем не менее это решение определяет ту альтернативу, при которой данная статистика локально оптимальна.

Литература

1. R. R. Bahadur, *Rates of convergence of estimates and test statistics.* — Ann. Math. Stat., **38** (1967), 303–324.
2. R. R. Bahadur, *Some limit theorems in statistics*, SIAM, Philadelphia (1971).
3. J. V. Deshpandé and S. C. Kochar, *Some competitors of tests based on powers of ranks for the two-sample problem.* — Sankhyā, **B42** (1980), 236–241.
4. J. V. Deshpandé and S. C. Kochar, *Some competitors of Wilcoxon–Mann–Whitney test for the location alternatives.* — J. Indian Statist. Assoc., **18** (1982), 9–18.
5. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, NY and London (1960).
6. P. Eichelsbacher, and M. Löwe, *A Large Deviation Principle for m-variate von Mises-Statistics and U-Statistics.* — J. Theoret. Prob. **8** (1995), 807–823.
7. S. C. Kochar and R. P. Gupta, *Some competitors of the Mood test for the two-sample problem.* — Commun. Statist. – Theory Methods, **15** (1986), 231–239.
8. V. S. Korolyuk and Yu. V. Borovskikh, *Theory of U–statistics*, Kluwer, Dordrecht (1994).
9. Y. Maesono, *Competitors of the Wilcoxon signed rank test*, Ann. Inst. Stat. Math., **39** (1987), Pt.A, 363–375.
10. Ya. Yu. Nikitin, *Local asymptotic Bahadur optimality and characterization problems.* — Probab. Theory Appl., **29** (1984),79–92.
11. Ya. Nikitin, *Asymptotic efficiency of nonparametric tests.* Cambridge University Press (1995).

12. Ya. Yu. Nikitin and E. V. Ponikarov, *Large deviations of Chernoff type for U-statistics and von Mises functionals.* — Proc. of St. Petersburg Math. Soc., **7**, 124–167 (in Russian). English transl. in Amer. Mathem. Soc. Transl., ser.2, ed. by N.N.Uraltseva (2001), **203**, 107–146.
13. Ya. Yu. Nikitin and E. V. Ponikarov, *Large deviations of Chernoff type for U- and V-statistics*, Doklady RAN, **369** (1999), 13–16.
14. Ya. Yu. Nikitin and E. V. Ponikarov, *On large deviations of non-degenerate two-sample U- and V-statistics with applications to Bahadur efficiency.* — Mathem. Meth. of Statist., **15**, No. 1 (2006), 103–122.
15. Ya. Yu. Nikitin and I. Peaucelle, *Efficiency and local optimality of distribution-free tests based on U- and V-statistics.* — Metron – Intern. Journ. of Statistics, LXII, No. 2 (2004), 185–200.
16. I. R. Savage, *Nonparametric statistics: a personal review.* — Sankhyā, **A 31** (1969), 107–144.

Litvinova V. V., Nikitin Ya. Yu. On asymptotic efficiency and local optimality of tests based on two-sample U- and V-statistics.

Many well-known homogeneity tests are based on two-sample U- and V-statistics. Using their large deviation asymptotics we calculate for them the local Bahadur efficiency for common alternatives and find general conditions of their local asymptotic optimality.

С.-Петербургский
государственный университет

Поступило 26 ноября 2007 г.