Записки научных
семинаров ПОМИ
Том 412, 2013 г.

К. Ю. Волкова, Я. Ю. Никитин

КРИТЕРИИ ЭКСПОНЕНЦИАЛЬНОСТИ, ОСНОВАННЫЕ НА ХАРАКТЕРИЗАЦИИ АХСАНУЛЛАХА, И ИХ ЭФФЕКТИВНОСТЬ

§1. Введение

Экспоненциальное распределение играет важную роль в теории вероятностей и математической статистике, а модели с экспоненциально распределенными наблюдениями часто встречаются в прикладных областях, таких как теория надежности, анализ данных типа времени жизни, анализ живучести систем и др. Поэтому проверка гипотезы экспоненциальности принадлежит к числу важнейших задач проверки статистических гипотез.

Существует множество критериев для проверки экспоненциальности, основанных на различных идеях [7, 9, 12, 13, 21, 22] и др. Среди характеризаций, на основе которых строились критерии проверки экспоненциальности, использовались характеризации свойством отсутствия памяти [3, 6, 19] и другие характеризации [1, 10, 15, 20, 24, 29–31].

Общая постановка задачи проверки экспоненциальности выглядит так: пусть $X_1, X_2, \ldots, X_n$ – независимые одинаково распределенные случайные величины (н.о.р.с.в.), имеющие непрерывную функцию распределения (ф.р.) F. Рассмотрим проверку сложной гипотезы экспоненциальности $H_0 : F(x)$ – ф.р. экспоненциального закона с плотностью $f(x) = \lambda e^{-\lambda x}$, $x \geqslant 0$, где $\lambda > 0$ – некоторый неизвестный параметр.

Предположим, что ф.р. F принадлежит классу распределений $\mathcal{F}$, удовлетворяющему следующим условиям: F строго монотонна, а функция интенсивности отказов $f(t)/(1 - F(t))$ монотонно возрастает или убывает для всех $t \geqslant 0$.

Ключевые слова: экспоненциальный закон, проверка гипотез, характеризация, U- и V-статистики, большие уклонения, бахадуровская эффективность, локальная оптимальность.

Работа поддержана грантами РФФИ 13-01-00172а, НШ-1216.2012.1, а также грантом НИР СПбГУ 6.38.672.2013.

Ахсануллах доказал в работе [4] следующую характеризацию экспоненциального закона свойствами порядковых статистик в рамках класса $\mathcal{F}$:

Пусть $X_1, \ldots, X_n$ – неотрицательные н.о.р.с.в. с ф.р. из класса $\mathcal{F}$. Если для некоторых i и j статистики $(k-i)(X_{i+1,k} - X_{i,k})$ и $(k-j)(X_{j+1,k} - X_{j,k})$ одинаково распределены для $1 \leqslant i < j < k$, то выборка имеет экспоненциальный закон распределения.

Задачу проверки экспоненциальности в классе $\mathcal{F}$ еще называют проверкой отсутствия старения против альтернативы, что наблюдается положительное (отрицательное) старение. Такие задачи важны для теории надежности и данная область сейчас активно исследуется, см., например, [5, 21].

Мы рассмотрим частный случай этой характеризации, когда $k = 2$, $i = 0$ и $j = 1$. Характеризация примет вид:

Пусть X и Y – неотрицательные н.о.р.с.в. с ф.р. из класса $\mathcal{F}$. Если статистики $|X - Y|$ и $2\min(X, Y)$ одинаково распределены, то X имеет экспоненциальный закон распределения.

Рассмотрим критерии для проверки экспоненциальности, использующие данную характеризацию, против альтернатив из класса $\mathcal{F}$.

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 Ахсануллаха построим две так называемые V-эмпирические ф.р., см. [17, 18] с помощью формул

$$H_n(t) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{1}\{|X_i - X_j| < t\}, \quad t \geqslant 0,$$

$$G_n(t) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{1}\{2\min(X_i, X_j) < t\}, \quad t \geqslant 0.$$

Известно, что свойства V- и U-эмпирических ф.р. схожи со свойствами обычной эмпирической ф.р., см. [14, 17]. Поэтому для больших n ф.р. H_n и G_n должны быть близки при основной гипотезе H_0, и мы сможем измерить их близость путем использования некоторых тестовых статистик.

Рассмотрим две новые статистики, инвариантные к параметру масштаба λ, для проверки гипотезы экспоненциальности:

$$I_n = \int_0^\infty (H_n(t) - G_n(t))\, dF_n(t), \tag{1}$$

$$D_n = \sup_{t \geqslant 0} \mid H_n(t) - G_n(t) \mid . \tag{2}$$

Мы опишем предельные распределения и большие уклонения обеих последовательностей статистик при гипотезе H_0, вычислим их локальную бахадуровскую эффективность при ряде альтернатив и исследуем для них условия локальной асимптотической оптимальности. При этом будем пользоваться необходимыми сведениями из теории U- и V-статистик и теории асимптотической эффективности по Бахадуру, изложенными в монографиях [8, 11, 18] и [23].

В качестве метода вычисления асимптотической относительной эффективности (АОЭ) критериев нами выбран подход Бахадура, поскольку статистика колмогоровского типа D_n не является асимптотически нормальной, поэтому подход Питмена к ней не применим. В случае интегральной статистики I_n, локальная АОЭ по Бахадуру и питменовская эффективности совпадают [23, 33].

§2. Статистика I_n

Без ограничения общности будем считать, что $\lambda = 1$. Статистика I_n асимптотически эквивалентна U-статистике степени 3 с центрированным ядром

$$3\Psi(X,Y,Z) = \mathbf{1}\{|X-Y| < Z\} + \mathbf{1}\{|Y-Z| < X\} + \mathbf{1}\{|X-Z| < Y\}$$
$$-\mathbf{1}\{2\min(X,Y) < Z\} - \mathbf{1}\{2\min(Y,Z) < X\} - \mathbf{1}\{2\min(X,Z) < Y\}.$$

Пусть X, Y, Z – независимые с.в., имеющие стандартное экспоненциальное распределение. Известно, что невырожденные U- и V-статистики асимптотически нормальны, см. [16, 18]. Чтобы показать, что ядро $\Psi(X,Y,Z)$ невырождено, вычислим его проекцию ψ. Для фиксированного $X = s$ проекция ядра имеет вид:

$$\psi(s) = \mathbf{E}(\Psi(X,Y,Z) \mid X = s) = \frac{2}{3}\mathbf{P}(|Y-s| < Z) + \frac{1}{3}\mathbf{P}(|Y-Z| < s)$$
$$- \frac{2}{3}\mathbf{P}(2\min(s,Y) < Z) - \frac{1}{3}\mathbf{P}(2\min(Y,Z) < s).$$

Первая и третья вероятности легко вычисляются:

$$\mathbf{P}(|Y-s|<Z) = \int\limits_s^\infty e^{-z}dz \int\limits_0^{z+s} e^{-y}dy + \int\limits_0^s e^{-z}dz \int\limits_{s-z}^{s+z} e^{-y}dy = e^{-s}\left(\frac{1}{2}+s\right),$$

$$\mathbf{P}(2\min(s,Y)<Z) = \frac{2}{3}e^{-3s}+\frac{1}{3},$$

а из характеризации Ахсануллаха следует, что

$$\mathbf{P}(2\min(Y,Z)<s) = \mathbf{P}(|Y-Z|<s) = 1-e^{-s}.$$

Окончательно получаем выражение для проекции ядра Ψ:

$$\psi(s) = \frac{1}{3}e^{-s}(1+2s) - \frac{4}{9}e^{-3s} - \frac{2}{9}. \tag{3}$$

Поэтому дисперсия проекции Δ^2 при гипотезе H_0 равна

$$\Delta^2 = \int\limits_0^\infty \psi^2(s)ds = \frac{647}{42525} \approx 0.015.$$

Таким образом, ядро Ψ невырождено. Применяя теорему Хёффдинга, снова см. [16,18], при $n\to\infty$ получаем

$$\sqrt{n}I_n \xrightarrow{d} \mathcal{N}\left(0, \frac{647}{4725}\right).$$

Найдем теперь логарифмическую асимптотику вероятности больших уклонений последовательностей статистик (1) при нулевой гипотезе. Ядро Ψ не только центрировано и невырождено, но и ограничено. Применяя теорему о больших уклонениях невырожденных U- и V-статистик из [26], см. также [11,24], получаем:

Теорема 1. *При $a>0$*

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

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

$$f(a) \sim \frac{a^2}{18\Delta^2} = \frac{4725}{1294}a^2 \quad \textit{при} \quad a\to 0.$$

В качестве альтернатив для рассматриваемой характеризации мы используем следующие альтернативы к гипотезе экспоненциальности, сосредоточенные на полуоси $x \geqslant 0$:

(i) альтернативу Вейбулла с плотностью

$$g_1(x,\theta) = (1+\theta)x^{\theta}\exp(-x^{1+\theta}), \quad \theta \geqslant 0, \quad x \geqslant 0;$$

(ii) альтернативу Макегама с плотностью

$$g_2(x,\theta) = (1+\theta(1-e^{-x}))\exp(-x-\theta(e^{-x}-1+x)), \quad \theta \geqslant 0, \quad x \geqslant 0;$$

(iii) гамма-альтернативу с плотностью

$$g_3(x,\theta) = \frac{x^{\theta}}{\Gamma(\theta+1)}e^{-x}, \quad \theta \geqslant 0, \quad x \geqslant 0.$$

(iv) специальную альтернативу с плотностью

$$g_4(x,\theta) = e^{-x}\left(1+\frac{\theta}{3}\left(e^{-x}(1+2x)-\frac{4}{3}e^{-3x}-\frac{2}{3}\right)+\theta(x-1)\right), \quad \theta \geqslant 0, \quad x \geqslant 0.$$

Обозначим их ф.р. как $G_j(x,\theta)$, $j = 1,2,3,4$, соответственно.

Первые 3 альтернативы являются стандартными альтернативами к гипотезе экспоненциальности. Последняя альтернатива не имеет устоявшегося названия. Нетрудно проверить, что все альтернативы принадлежат классу распределений $\mathcal{F}$.

Предположим, что при альтернативе H_1 наблюдения имеют ф.р. $G(\cdot\,,\theta)$ с плотностью $g(\cdot,\theta)$, $\theta \geqslant 0$, такую что $G(\cdot\,,0)$ является экспоненциальной ф.р. В теории Бахадура за меру качества последовательности статистик $\{T_n\}$ принимается точный наклон $c_T(\theta)$, описывающий степень экспоненциального убывания их $\mathbf{P}$-значений при альтернативе. Чтобы вычислить точный наклон, существует следующая ключевая теорема Бахадура.

Теорема 2 ([8, §7]). *Пусть последовательность статистик $\{T_n\}$ удовлетворяет следующим двум условиям:*

(a) $T_n \xrightarrow{P_\theta} b(\theta), \quad \theta > 0, \quad$ *где* $\quad -\infty < b(\theta) < \infty.$

(b) $\lim\limits_{n\to\infty} n^{-1} \ \ln \ \mathbf{P}_{H_0}\,(T_n \geqslant t\ ) = -f(t)$

для любого t из открытого интервала I, где f непрерывна и $\{b(\theta),\ \theta > 0\} \subset I$. Тогда справедлива формула

$$c_T(\theta) = 2\ f(b(\theta)), \quad \theta > 0.$$

Известно, что всегда выполнено следующее неравенство Бахадура–Рагавачари:

$$c_T(\theta) \leqslant 2K(\theta), \tag{4}$$

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

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

Если же в (4) выполняется соотношение

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

то последовательность T_n называется локально асимптотически оптимальной (ЛАО) по Бахадуру [8, 23].

Вернемся к последовательности статистик I_n. Вычислим теперь локальный бахадуровский наклон и локальную эффективность последовательности статистик I_n для альтернативы с ф.р. $G(x, \theta)$ и плотностью $g(x, \theta)$, предполагая их регулярность и возможность дифференцировать соответствующие интегралы по параметру под знаком интеграла. Эти требования выполнены для всех четырех рассматриваемых альтернатив. Обозначим также

$$h(x) = g'_\theta(x, 0). \tag{5}$$

Заметим, что $\int\limits_0^1 h(x)\,dx = 0$.

Второе условие теоремы 2 нами уже установлено ранее в (1). Для проверки первого условия теоремы 2 нам потребуется применить закон больших чисел для U-статистик, см., например, [18]. Используя введенные обозначения, легко показать (см. также [25]), что

$$b_I(\theta) \sim 3\theta \int\limits_0^\infty \psi(s)h(s)ds, \tag{6}$$

где $\psi(s)$ – проекция, вычисленная в (3).

Альтернатива Вейбулла. После ряда нетрудных вычислений, из (6) при $\theta \to 0$ получаем, что

$$b_I(\theta) \sim 3\theta \int\limits_0^\infty \left(\frac{1}{3}e^{-s}(1+2s) - \frac{4}{9}e^{-3s} - \frac{2}{9}\right) e^{-s}(1+\ln s - s\ln s)\,ds$$
$$= \frac{1}{4}(1+\ln 2)\,\theta \approx 0.4233\,\theta, \quad \theta \to 0,$$

и локальный точный наклон последовательности статистик I_n допускает при $\theta \to 0$ представление

$$c_I(\theta) = b_I^2(\theta)/(9\Delta^2) \sim 1.308\,\theta^2.$$

Информация Кульбака–Лейблера $K(\theta)$ между H_0 и H_1 для альтернативы Вейбулла вычислена, например, в [1]. Оказывается, что $K(\theta) \sim \pi^2\theta^2/12$, $\theta \to 0$. Следовательно, локальная бахадуровская эффективность нашего критерия равна

$$\operatorname{eff}(I_n) = \lim_{\theta\to 0} \frac{c_I(\theta)}{2K(\theta)} \approx 0.795.$$

Альтернатива Макегама. Пользуясь полученной формулой (6), мы имеем:

$$b_I(\theta) \sim \frac{4}{45}\,\theta, \quad \theta \to 0.$$

Локальный точный наклон допускает представление $c_I(\theta) \sim 0.058\,\theta^2$, $\theta \to 0$. Информация Кульбака–Лейблера для альтернативы Макегама известна из [1], где доказано, что $K(\theta) \sim \theta^2/24$, $\theta \to 0$. Следовательно, локальная бахадуровская эффективность равна 0.692.

Гамма-альтернатива. Аналогично предыдущим случаям, с помощью соотношения (6) получаем, что $b_I(\theta) \sim 0.269\,\theta$, $c_I(\theta) \sim 0.528\,\theta^2$. Из [2] известно, что $K(\theta) \sim (\frac{\pi^2}{12} - \frac{1}{2})\,\theta^2$, $\theta \to 0$. Поэтому локальная эффективность критерия для этой альтернативы равна 0.819.

Альтернатива $g_4(x,\theta)$. В этом случае после ряда вычислений мы получаем, что

$$c_I(\theta) \sim \frac{647}{42525}\,\theta^2 \quad \text{при} \quad \theta \to 0.$$

Выведем теперь выражение при $\theta \to 0$ “расстояния” Кульбака–Лейблера между нулевой гипотезой и альтернативой $g_4(x,\theta)$ в терминах функции $h(x)$, введенной в (5), заметив, что основная гипотеза

является сложной. Нетрудно показать, см. также [2, 27], что при указанных условиях регулярности

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

Поэтому информация Кульбака–Лейблера в этом случае удовлетворяет соотношению

$$2K(\theta) \sim \frac{647}{42525}\theta^2, \quad \theta \to 0. \tag{7}$$

Следовательно, локальная бахадуровская эффективность нашего критерия равна 1, и интегральный тест является локально оптимальным в бахадуровском смысле, см. [23, гл. 6]. Позже мы вернемся к объяснению этого явления.

Следует признать, что наш интегральный критерий оказался состоятельным и довольно эффективным для всех рассмотренных альтернатив.

Таблица 1. Локальная бахадуровская эффективность для статистики I_n.

Альтернатива	Эффективность
Вейбулл	0.795
Макегам	0.692
Гамма	0.819
g_4	1. 000

§3. Статистика типа Колмогорова D_n

Теперь рассмотрим статистику колмогоровского типа (2). В этом случае при фиксированном $t \geqslant 0$ разность $H_n(t) - G_n(t)$ является семейством V-статистик с ядрами, зависящими от $t \geqslant 0$:

$$\Xi(X, Y; t) = \mathbf{1}\{|X - Y| < t\} - \mathbf{1}\{2\min(X, Y) < t\}.$$

После ряда простых вычислений, получаем, что проекции ядер $\xi(s; t) := \mathbf{E}(\Xi(X, Y; t) \mid X = s)$ для фиксированного t равны:

$$\xi(s; t) = \begin{cases} -e^{-t-s}, & 0 \leqslant s < t/2; \\ e^{-t/2} - e^{-t-s}, & t/2 \leqslant s < t; \\ e^{-s}(e^t - e^{-t}) + e^{-t/2} - 1, & t \leqslant s < +\infty. \end{cases} \tag{8}$$

Теперь вычислим дисперсии проекций $\delta^2(t)$ при гипотезе H_0. В результате вычислений получим:

$$\int_0^\infty \xi^2(s,t)e^{-s}ds := \delta^2(t) = \frac{1}{3}e^{-t} + \frac{1}{3}e^{-2t} + \frac{1}{3}e^{-3t} - e^{-5t/2}.$$

Поэтому наше семейство ядер $\Xi(X,Y;t)$ невырождено [24] и

$$\delta^2 = \sup_t \delta^2(t) = 0.115.$$

Предельное распределение статистики D_n неизвестно, но принципиально может быть найдено с помощью методов работы [32]. Доказывается, что U-эмпирический процесс

$$\eta_n(t) = \sqrt{n}\,(H_n(t) - G_n(t))\,, \quad t \geqslant 0,$$

слабо сходится в $D(0,\infty)$ при $n \to \infty$ к некоторому центрированному гауссовскому процессу $\eta(t)$ с явно вычисляемой ковариацией. Тогда последовательность статистик $\sqrt{n}D_n$ сходится по распределению к величине $\sup_{t\geqslant 0}|\eta(t)|$, найти распределение которой в явном виде не представляется возможным. Стало быть, критические значения для статистик D_n целесообразно искать с помощью моделирования их выборочного распределения.

Семейство ядер $\{\Xi(X,Y;t)\}$, $t \geqslant 0$, не только центрировано, но и ограничено. Применяя теорему о больших уклонениях супремума семейства невырожденных U- и V-статистик [24], получаем, что справедлива

Теорема 3. *При* $a > 0$

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

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

$$f(a) = (8\delta^2)^{-1}a^2(1+o(1)) \sim 1.083\,a^2 \quad \textit{при} \quad a \to 0.$$

Альтернатива Вейбулла. Теперь снова рассмотрим альтернативу Вейбулла. По теореме Гливенко–Кантелли для супремума семейства U- и V-статистических ф.р. [17] предел D_n по вероятности равен

$$b_D(\theta) := \sup_{t\geqslant 0}|b_D(t,\theta)| = \sup_{t\geqslant 0}\left|\,\mathbf{P}_\theta(|X-Y|<t) - \mathbf{P}_\theta(2\min(X,Y)<t)\,\right|.$$

Нетрудно показать, что выполнена следующая асимптотика для функции $b_D(t,\theta)$:

$$b_D(t,\theta) \sim 2\theta \int\limits_0^\infty \xi(s;t)h(s)\,ds, \tag{9}$$

где снова $h(s) = \frac{\partial}{\partial\theta} g(s,\theta)\big|_{\theta=0}$ и $\xi(s;t)$ являются проекциями, вычисленными в (8). Таким образом, для альтернативы Вейбулла мы имеем для $t \geqslant 0$:

$$\begin{aligned} b_D(t,\theta) &\sim 2\theta \int\limits_0^{t/2} (-e^{-t-s})e^{-s}(1+\ln s - s\ln s)\,ds \\ &+ 2\theta \int\limits_{t/2}^{t} (e^{-t/2} - e^{-t-s})e^{-s}(1+\ln s - s\ln s)\,ds \\ &+ 2\theta \int\limits_{t}^{\infty} (e^{-s}(e^t - e^{-t}) + e^{-t/2} - 1)e^{-s}(1+\ln s - s\ln s)\,ds \\ &= \left(\frac{1}{2}e^{-t}(\ln t + \gamma + \ln 2 + 2t\ln 2) + \frac{1}{2}e^t\,\mathrm{Ei}\,(1,2t)\right)\theta, \quad \theta \to 0, \end{aligned}$$

где $\gamma \approx 0.5772$ – константа Эйлера, а через $\mathrm{Ei}\,(k,z)$ мы обозначаем экспоненциальный интеграл $\mathrm{Ei}\,(k,z) = \int\limits_1^\infty e^{-zt}t^{-k}\,dt$, $k \geqslant 0$, $\mathrm{Re}\,z > 0$.

Используя график функции $b_D(t,\theta)$ на рис. 1, мы видим, что

$$b_D(\theta) = \sup_{t\geqslant 0} |b_D(t,\theta)| \sim b(0.692,\theta) \approx 0.585\,\theta,$$

и локальный точный наклон последовательности статистик D_n при $\theta \to 0$ допускает представление:

$$c_D(\theta) \sim b_D^2(\theta)/(4\delta^2) \sim 0.741\,\theta^2, \quad \theta \to 0.$$

Информация Кульбака–Лейблера для альтернативы Вейбулла равна $K(\theta) \sim \frac{\pi^2\theta^2}{12}$, $\theta \to 0$, а значит локальная бахадуровская эффективность равна 0.450.

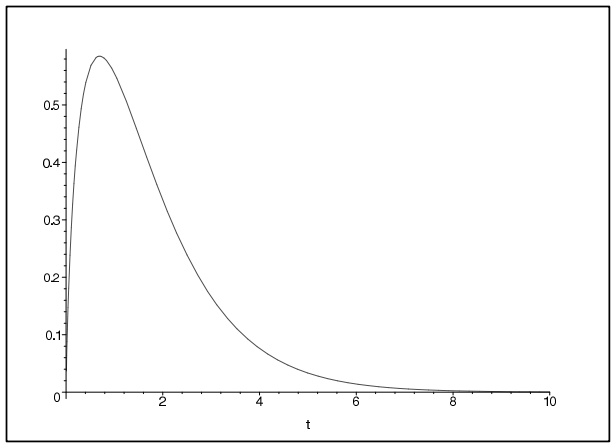

Рис. 1. График функции $b_D(t,\theta)$ при альтернативе Вейбулла.

Альтернатива Макегама. В этом случае, пользуясь полученной формулой (9), после ряда вычислений имеем:

$$b_D(t,\theta) \sim \left(\frac{4}{3}e^{-t} + \frac{2}{3}e^{-2t} - 2e^{-3t/2}\right)\theta, \quad \theta \to 0.$$

Таким образом, с помощью рис. 2 получаем, что

$$b_D(\theta) = b(0.9899,\theta) \approx 0.134\,\theta.$$

Поэтому локальный точный наклон равен $c_D(\theta) \approx 0.039\,\theta^2$.

Информация Кульбака–Лейблера для альтернативы Макегама равна $K(\theta) \sim \frac{\theta^2}{24}$, $\theta \to 0$, а эффективность нашего критерия равна 0.470.

Гамма-альтернатива. Аналогично предыдущим случаям мы получаем

$$b_D(t,\theta) \sim (e^{-t}(\ln t+\gamma-\ln 2)+2e^{-t/2}\mathrm{Ei}\,(1,t/2)-2\mathrm{Ei}\,(1,t)+e^{t}\mathrm{Ei}\,(1,2t))\theta, \quad \theta \to 0,$$

где γ – константа Эйлера, $\gamma \approx 0.5772$, а

$$\mathrm{Ei}\,(k,z) = \int\limits_1^\infty e^{-zt}t^{-k}\,dt.$$

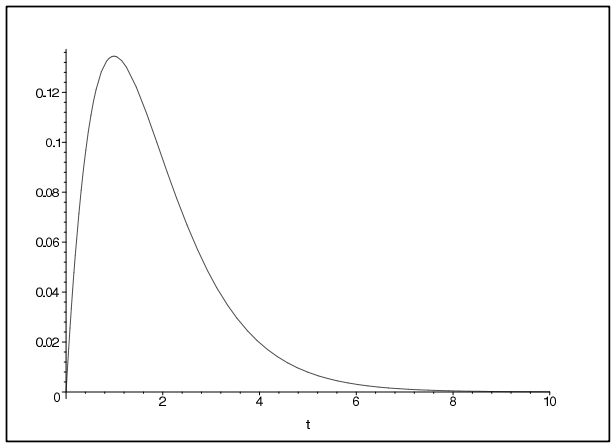

Рис. 2. График функции $b_D(t,\theta)$ при альтернативе Макегама.

Таким образом, используя рис. 3, находим, что

$$b_D(\theta) = b(0.5457, \theta) \approx 0.379\,\theta.$$

Локальный точный наклон равен $c_D(\theta) \approx 0.311\,\theta^2$. Поэтому локальная эффективность критерия для этой альтернативы равна 0.482.

Альтернатива $g_4(x,\theta)$**.** Последней мы рассмотрим альтернативу с плотностью $g_4(x,\theta)$. После некоторых вычислений с помощью рис. 4 получаем, что при $\theta \to 0$ выполнены следующие соотношения:

$$b_D(t,\theta) \sim \theta \frac{e^{-t}}{3}\left(e^{-t/2}(t+2) - \frac{2}{3}e^{-t}(3t+4) - \frac{2}{3}e^{-3t/2} + \frac{2}{15}e^{-3t} - \frac{26}{45}\right),$$
$$c_I(\theta) \sim 0.011\,\theta^2.$$

Мы знаем, что информация Кульбака–Лейблера в этом случае удовлетворяет (7). Следовательно, локальная бахадуровская эффективность нашего критерия равна 0.740.

Заметим, что эффективность статистик типа Колмогорова меньше, чем интегральных статистик I_n, как это обыкновенно бывает в задачах проверки гипотез, см. также [23].

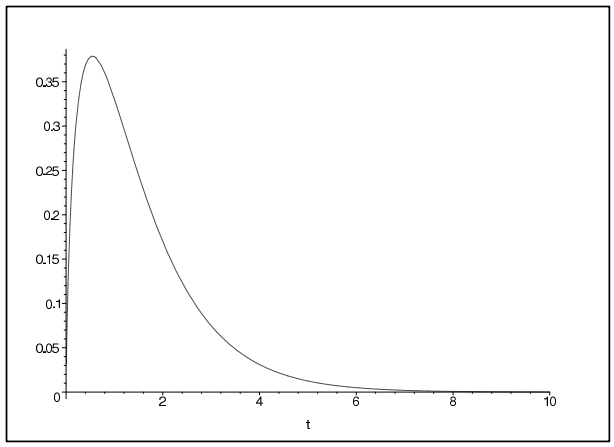

Рис. 3. График функции $b_D(t,\theta)$ при гамма-альтернативе.

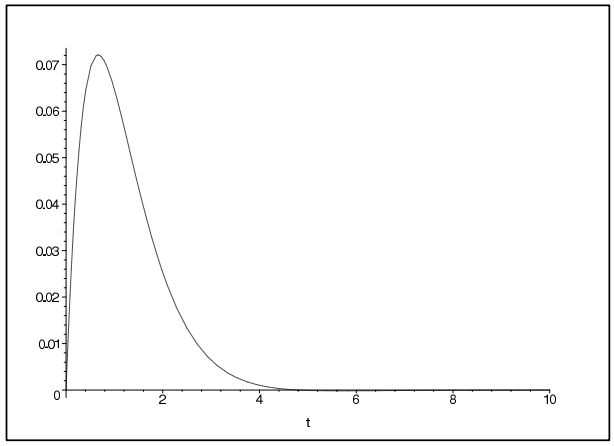

Рис. 4. График функции $b_D(t,\theta)$при альтернативе g_4.

§4. Условия локальной асимптотической оптимальности

Выясним условия локальной асимптотической оптимальности для интегральной и супремальной статистик, то есть опишем семейства

Таблица 2. Локальная бахадуровская эффективность для статистики D_n.

альтернатива	эффективность
Вейбулл	0.450
Макегам	0.470
Гамма	0.482
g_4	0.740

альтернатив, для которых изучаемая статистика является локально наилучшей в бахадуровском смысле.

Подобные задачи уже рассматривались при изучении классических непараметрических критериев в [23, гл. 6], где использовались вариационные методы.

Рассмотрим функции

$$H(x) = \frac{\partial}{\partial\theta} G(x,\theta)\big|_{\theta=0}, \quad h(x) = \frac{\partial}{\partial\theta} g(x,\theta)\big|_{\theta=0}.$$

Предположим также, что выполнены следующие условия регулярности, см. [27]:

$$h(x) = H'(x), \quad \forall x \geqslant 0, \quad \int\limits_0^\infty h^2(x) e^x \, dx < \infty, \tag{10}$$

$$\frac{\partial}{\partial\theta} \int\limits_0^\infty x g(x,\theta)\, dx\big|_{\theta=0} = \int\limits_0^\infty x h(x) dx. \tag{11}$$

Обозначим через $\mathcal{G}$ класс плотностей $g(x,\theta)$ с ф.р. $G(x,\theta)$, удовлетворяющими условиям регулярности (10)–(11). Выведем условия оптимальности в терминах функции $h(x)$.

Как уже было замечено, для альтернативных плотностей из класса $\mathcal{G}$ выполнено следующее соотношение при указанных условиях регулярности

$$2K(\theta) \sim \left\{ \int\limits_0^\infty h^2(x) e^x \, dx - \left(\int\limits_0^\infty x h(x)\, dx \right)^2 \right\} \theta^2, \quad \theta \to 0.$$

Начнем с интегральной статистики I_n с ядром $\Psi(X,Y,Z)$ и проекциями $\psi(x)$. Напомним, что выполнена следующая асимптотика для

функции $b_I(\theta)$:

$$b_I(\theta) \sim 3\theta \int\limits_0^\infty \psi(x)h(x)\,dx.$$

Будем также предполагать, что $h(x)e^x$ – элемент гильбертова пространства функций, квадраты которых интегрируемы с весом e^{-x} на $[0,\infty]$:

$$h(x)e^x \in \mathfrak{L}_2 \equiv \left\{ q(x) : \int\limits_0^\infty q^2(x)e^{-x}\,dx < \infty \right\}.$$

Введем вспомогательную функцию

$$h_0(x) = h(x) - (x-1)\exp(-x) \int\limits_0^\infty uh(u)\,du.$$

Непосредственно получаем соотношения:

$$\int\limits_0^\infty h^2(x)e^x\,dx - \left(\int\limits_0^\infty xh(x)\,dx \right)^2 = \int\limits_0^\infty h_0^2(x)e^x\,dx,$$

$$\int\limits_0^\infty \psi(x)h(x)\,dx = \int\limits_0^\infty \psi(x)h_0(x)\,dx.$$

Поэтому локальная асимптотическая эффективность принимает вид

$$\begin{aligned} \mathrm{eff}_I &= \lim_{\theta\to 0} b_I^2(\theta) / \left(9\Delta^2 2K(\theta)\right) \\ &= \left(\int\limits_0^\infty \psi(x)h_0(x)\,dx \right)^2 \Big/ \left(\int\limits_0^\infty \psi^2(x)e^{-x}\,dx \cdot \int\limits_0^\infty h_0^2(x)e^x\,dx \right). \end{aligned}$$

Локальная асимптотическая оптимальность по Бахадуру статистики I_n означает, что выражение в правой части равно 1. Поскольку непосредственно проверяется, что функция $\psi(x)$ удовлетворяет условиям регулярности (10)–(11), то она также принадлежит классу $\mathfrak{L}_2$. Из неравенства Коши–Буняковского–Шварца, см. также [25], следует, что это выполняется тогда и только тогда, когда $h_0(x)e^x = C_1\psi(x)$ для некоторой константы $C_1 \neq 0$. При этом состоятельность при малых $\theta > 0$ будет обеспечиваться при $C_1 > 0$. Таким образом, функция

$h(x) = g_\theta(x,0)$ примет вид $h(x) = e^{-x}(C_1\psi(x) + C_2(x-1))$ для некоторых констант $C_1 > 0$ и C_2. Распределения, для которых функция $h(x)$ имеет указанный вид, образуют область ЛАО в классе $\mathcal{G}$.

Примером такой альтернативы может служить та, у которой плотность $g(x,\theta)$ при всех $\theta \to +0$ выражается формулой:

$$g(x,\theta) = e^{-x}\left(1 + \frac{\theta}{3}\left(e^{-x}(1+2x) - \frac{4}{3}e^{-3x} - \frac{2}{3}\right) + \theta(x-1)\right), \quad x \geqslant 0.$$

Напомним, что рассматриваемая характеризация Ахсануллаха выполнена для функций из класса $\mathcal{F}$. Нетрудно проверить, что функция $g(x,\theta)$ находится в нужном классе. Указанному условию удовлетворяет, в частности, альтернатива g_4, что и объясняет локальную оптимальность для нее интегрального критерия.

Теперь рассмотрим статистику колмогоровского типа D_n с семействами ядер $\Xi(X,Y;t)$ и проекциями $\xi(x;t)$. В этом случае справедлива асимптотика для функции $b(\theta)$:

$$b_D(\theta) \sim 2\theta \sup_{t\geqslant 0}\left|\int_0^\infty \xi(x;t)h_0(x)\,dx\right|. \tag{12}$$

Поэтому локальная асимптотическая эффективность принимает вид

$$\begin{aligned}\mathrm{eff}_D &= \lim_{\theta\to 0} b_D^2(\theta)/\sup_{t\geqslant 0}\left(4\delta^2(t)\right)2K(\theta)\\ &= \sup_{t\geqslant 0}\left(\int_0^\infty \xi(x;t)h_0(x)\,dx\right)^2 \Big/ \sup_{t\geqslant 0}\left(\int_0^\infty \xi^2(x,t)e^{-x}\,dx \cdot \int_0^\infty h_0^2 e^x\,dx\right).\end{aligned}$$

Легко проверяется, что при фиксированном t функция $\xi(x,t)$ удовлетворяет условиям регулярности (10)–(11), поэтому $\xi(x,t) \in \mathfrak{L}_2$. Обозначим $t_0 = \arg\sup_{t\geqslant 0}\delta^2(t)$. Применяя неравенство Коши–Буняковского–Шварца к интегралу в (12) и вычисляя супремум по t, см. также [2, 28], мы приходим к условию равенства в виде:

$$\left(\int_0^\infty \xi(x;t_0)h_0(x)\,dx\right)^2 = \int_0^\infty \xi^2(x,t_0)e^{-x}\,dx \cdot \int_0^\infty h_0^2 e^x\,dx,$$

причем это равенство имеет место лишь тогда, когда

$$h_0(x) = C_1 e^{-x}\xi(x,t_0).$$

Ясно, что для функции $h_0(x)$ указанного вида локальная асимптотическая эффективность статистики D_n равна 1. Таким образом, последовательность статистик $\{D_n\}$ локально оптимальна тогда и только тогда, когда соответствующая функция $h(x)$ удовлетворяет условию $h(x) = e^{-x}(C_1\xi(x,t_0) + C_2(x-1))$ для $t_0 = \arg\sup\limits_{t\geqslant 0} \delta^2(t)$ и для некоторых констант $C_1 > 0$ и C_2. Распределения, для которых функция $h(x)$ имеет указанный вид, образуют область ЛАО в соответствующем классе $\mathcal{G}$.

Нам не удалось построить примеров альтернатив, образующих область ЛАО, поскольку в характеризации накладываются дополнительные ограничения на функции распределения наблюдений, сильно усложняющие их построение.

§5. Заключение

Нами были предложены два новых критерия для проверки экспоненциальности, основанные на известной характеризации Ахсануллаха. Мы описали их предельное распределение и большие уклонения, а также вычислили локальную бахадуровскую эффективность для ряда альтернатив. Был найден вид альтернатив, образующих область ЛАО наших критериев. Статистики оказались довольно эффективными для рассмотренных альтернатив и могут быть рекомендованы для практического применения. В целом мы считаем, что изучение U-эмпирических критериев, основанных на характеризациях экспоненциального закона, довольно перспективно.

Литература

1. В. В. Литвинова, *Асимптотические свойства критериев симметрии и согласия, основанных на характеризациях.* Кандидатская диссертация, СПбГУ, 2004.
2. А. В. Чирина, *Асимптотическая эффективность критериев экспоненциальности, свободных от параметра масштаба.* Кандидатская диссертация, СПбГУ, 2006.
3. I. Ahmad, I. Alwasel, *A goodness-of-fit test for exponentiality based on the memoryless property.* — J. Roy. Statist. Soc. **61**, Pt. 3 (1999), 681–689.
4. M. Ahsanullah, *On a characterization of the exponential distribution by spacings.* — Ann. Inst. Statist. Math. **30** (1978), 163–166.
5. M. Z. Anis, S. Dutta, *Recent tests of exponentiality against IFR alternatives: a survey.* — J. Statist. Comput. Simul. **80**, No. 12 (2010), 1373–1387.

6. J. E. Angus, *Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation.* — J. Statist. Plann. Infer. **6** (1982), 241–251.
7. S. Asher, *A survey of tests for exponentiality.* — Commun. Statist. Theory Meth. **19** (1990), 1811–1825.
8. R. R. Bahadur, *Some Limit Theorems in Statistics.* SIAM, Philadelphia, 1971.
9. N. Balakrishnan, A. Basu, *The Exponential Distribution: Theory, Methods and Applications.* Gordon and Breach, Langhorne, PA, 1995.
10. 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.
11. A. DasGupta, *Asymptotic Theory of Statistics and Probability.* Springer, New York, 2008.
12. K. A. Doksum, B. S. Yandell, *Tests of exponentiality.* — Handbook of Statist. **4** (1984), 579–612.
13. A. Grane, J. Fortiana, *A location- and scale-free goodness-of-fit statistic for the exponential distribution based on maximum correlations.* — Statist. **43** (2009), 1–12.
14. R. Helmers, P. Janssen, R. Serfling, *Glivenko–Cantelli properties of some generalized empirical DF's and strong convergence of generalized L-statistics.* — Probab. Theory Relat. Fields **79** (1988), 75–93.
15. 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.
16. W. Hoeffding, *A class of statistics with asymptotically normal distribution.* — Ann. Math. Statist. **19** (1948), 293–325.
17. P. L. Janssen, *Generalized Empirical Distribution Functions with Statistical Applications.* Limburgs Universitair Centrum, Diepenbeek, 1988.
18. V. S. Korolyuk, Yu. V. Borovskikh, *Theory of U-statistics.* Kluwer, Dordrecht, 1994.
19. H. L. Koul, *A test for new better than used.* — Commun. Statist. Theory Meth. **6** (1977), 563–574.
20. H. L. Koul, *Testing for new is better than used in expectation.* — Commun. Statist. Theory Meth. **7** (1978), 685–701.
21. C. D. Lai, M. Xie, *Stochastic Ageing and Dependence for Reliability.* Springer-Verlag, New York, 2006.
22. P. Nabendu, J. Chun, R. Crouse, *Handbook of Exponential and Related Distributions for Engineers and Scientists.* Chapman and Hall, 2002.
23. Y. Nikitin, *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, New York, 1995.
24. Ya. Yu. Nikitin, *Large deviations of U-empirical Kolmogorov–Smirnov tests, and their efficiency.* — J. Nonpar. Statist. **22** (2010), 649–668.
25. Ya. Yu. Nikitin, I. Peaucelle, *Efficiency and local optimality of distribution-free tests based on U- and V-statistics.* — Metron **LXII** (2004), 185–200.
26. Ya. Yu. Nikitin, E. V. Ponikarov, *Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics.* — In: Proc. St.Petersburg Math. Society **7** (1999), pp. 124–167.
27. 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.

28. Ya. Yu. Nikitin, A. V. Tchirina, *Lilliefors test for exponentiality: large deviations, asymptotic efficiency, and conditions of local optimality.* — Math. Meth. Statist. **16**, No. 1 (2007), 16–24.
29. Ya. Yu. Nikitin, K. Yu. Volkova, *Asymptotic efficiency of exponentiality tests based on order statistics characterization.* — Georgian Math. J. **17** (2010), 749–763.
30. R. F. Rank, *Statistische Anpassungstests und Wahrscheinlichkeiten grosser Abweichungen.* Doktor der Naturwissenschaften Dr. rer. nat. genehmigte Dissertation, Hannover, 1999.
31. J. S. Rao, E. Taufer, *The use of mean residual life to test departures from exponentiality.* — J. Nonparam. Statist. **18** (2006), 277–292.
32. B. W. Silverman, *Convergence of a class of empirical distribution functions of dependent random variables.* — Ann. Probab. **11** (1983), 745–751.
33. H. S. Wieand, *A condition under which the Pitman and Bahadur approaches to efficiency coincide.* — Ann. Statist. **4** (1976), 1003–1011.

Volkova K. Yu., Nikitin Ya. Yu. Exponentiality tests based on Ahsanullah's characterization and their efficiencies.

We construct integral and supremum type tests of exponentiality based on Ahsanullah's characterization of exponential law. We discuss limiting distributions and large deviations of new test statistics under null-hypothesis, and calculate their local Bahadur efficiency under common parametric alternatives. Conditions of local optimality of new statistics are given.

С.-Петербургский государственный университет, Университетский пр. 28, 198504, Петродворец Россия

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

Поступило 17 февраля 2013 г.