УДК 519.22 Вестник СПбГУ. Математика. Механика. Астрономия. 2019. Т. 6 (64). Вып. 2
MSC 62F03, 62G20

Критерии согласия, основанные на характеризации логистического распределения*

Я. Ю. Никитин, И. А. Рагозин

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Никитин Я. Ю., Рагозин И. А.* Критерии согласия, основанные на характеризации логистического распределения // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2019. Т. 6 (64). Вып. 2. С. 241–252.
https://doi.org/10.21638/11701/spbu01.2019.206

Логистическое семейство распределений принадлежит к числу важных семейств в теории вероятностей и математической статистике. Тем не менее, для проверки сложной гипотезы о принадлежности выборки логистическому семейству с неизвестным сдвигом против альтернатив общего вида критерии согласия почти не изучены. Разрабатываются два новых критерия согласия, интегральный критерий и критерий типа Колмогорова, основанные на недавней характеризации логистического семейства, принадлежащей Ху и Лину. Обсуждаются асимптотические свойства построенных критериев и вычисляется их бахадуровская асимптотическая эффективность для ряда естественных альтернатив.

Ключевые слова: характеризация распределений, логистическое распределение, асимптотическая эффективность, большие уклонения, информация Кульбака — Лейблера.

1. Введение. Настоящая работа посвящена критериям согласия для логистического семейства распределений. В дальнейшем под стандартной логистической функцией распределения (ф. р.) L и ее плотностью l мы будем понимать

$$L(x) = e^x/(e^x + 1) \quad \text{и} \quad l(x) = e^x/(e^x + 1)^2,\ x \in \mathbb{R}. \tag{1}$$

Несмотря на важность и популярность логистических семейств распределения (см., например, справочник [1], посвященный исключительно логистическому семейству), критерии согласия для него мало изучены. Примером очень редких публикаций на эту тему может служить статья Стивенса [2], в которой применяется известный подход Дурбина, основанный на сходимости эмпирических процессов с оцениваемыми параметрами.

В последние годы растущую популярность приобретает проверка гипотез согласия, основанная на характеризации распределений, см. обзорную статью [3]. В ней описано множество критериев экспоненциальности, нормальности и равномерности, основанных на характеризациях этих семейств распределений, причем многие из них удобны для использования на практике и имеют высокую асимптотическую эффективность. Причина этого, возможно, в скрытых свойствах распределений, выраженных именно в характеризационных терминах. Однако никаких критериев согласия

*Работа выполнена при поддержке гранта СПбГУ–ННИО 6.65.37.2017.

для проверки *сложной гипотезы* о логистичности семейства распределений, основанных на характеризациях, до настоящего времени не было известно. В недавних тезисах [4] обсуждается лишь *простая* гипотеза о стандартной логистичности.

Целью данной работы является построение и асимптотический анализ двух критериев согласия для логистического семейства со сдвигом, основанных на недавно полученной характеризации Ху и Лина [5], описываемой ниже. Первый — критерий интегрального типа, второй является вариантом критерия Колмогорова.

Известные характеризации логистического закона довольно немногочисленны. Представление о них дают статьи [6] и [7]. В недавней публикации тайваньских математиков Ху и Лина [5, теор. 4] приведена новая характеризация логистического распределения, простейшим частным случаем которой (при $n = 2$ и $k = 1$) является следующее утверждение, использующее «случайные экспоненциальные сдвиги».

Теорема 1. *Пусть X и Y — независимые одинаково распределенные случайные величины с непрерывной ф. р. F, а Z — независимая от X и Y случайная величина со стандартным экспоненциальным распределением. Тогда X и $\min(X, Y) + Z$ одинаково распределены тогда и только тогда, когда F является ф. р. логистического семейства со сдвигом с плотностью $l(x + \theta), \theta \in \mathbb{R}$, где плотность l указана в (1).*

Следует отметить, что этому результату предшествовала работа [8], в которой была получена характеризация стандартного логистического распределения, т. е. при $\theta = 0$, причем при дополнительных условиях на F.

Теорему 1 мы и положим в основу построения критериев для проверки сложной гипотезы о принадлежности семейству логистических распределений со сдвигом.

2. Построение статистик. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные наблюдения с ф. р. F. Используя теорему 1, будем проверять сложную гипотезу согласия H_0, согласно которой F есть ф. р. логистического закона (1) против альтернативы H_1, состоящей в том, что гипотеза H_0 не выполняется.

Обозначим через $F_n(t)$ обычную эмпирическую ф. р., а именно

$$F_n(t) = n^{-1} \sum_{i=1}^{n} I\{X_i < t\}, \quad t \in \mathbb{R}.$$

Теперь рассмотрим U-эмпирическую ф. р. [9]

$$\overline{U}_n(t) = \left(C_n^2\right)^{-1} \sum_{1 \le i < j \le n} I\left\{\min(X_i, X_j) < t\right\}, \quad t \in \mathbb{R},$$

а также еще одну U-эмпирическую ф. р., отвечающую сдвигу на экспоненциальную величину:

$$U_n(t) = \int_0^\infty \overline{U}_n(t-s) e^{-s} ds = \left(C_n^2\right)^{-1} \sum_{1 \le i < j \le n} \int_0^\infty I\left\{\min(X_i, X_j) < t - s\right\} e^{-s} ds =$$

$$= \left(C_n^2\right)^{-1} \sum_{1 \le i < j \le n} \left(1 - e^{(\min(X_i, X_j) - t)}\right) I\left\{\min(X_i, X_j) < t\right\}.$$

Введем две статистики: интегральную статистику

$$LU_n = \int_{-\infty}^{\infty} \left(F_n(t) - U_n(t)\right) dF_n(t) \tag{2}$$

и статистику типа Колмогорова

$$KU_n = \sup_t |F_n(t) - U_n(t)|, \tag{3}$$

которые могут служить для проверки H_0 против H_1. По теореме Гливенко — Кантелли для U-эмпирических функций [9] выражение $F_n(t) - U_n(t)$ ввиду рассматриваемой характеризации стремится п. н. к нулю равномерно по t, поэтому статистики LU_n и KU_n при основной гипотезе должны быть малы. Этот факт и используется для проверки гипотезы H_0.

3. Эффективность по Бахадуру. Одна из целей данной работы — асимптотическое сравнение построенных статистик на основе понятия бахадуровской эффективности, которая представляется наиболее удобным средством для сравнения статистик, не являющихся асимптотически нормальными. Поэтому ниже мы сжато изложим основные факты теории Бахадура [10, 11]. Мерой асимптотической эффективности последовательности статистик $\{T_n\}$ в этой теории является точный наклон $c_T(\theta)$, описывающий скорость экспоненциального убывания достигаемого уровня последовательности статистик при альтернативе. Сформулируем фундаментальную теорему Бахадура [10, 12]. Будем считать, что распределение наблюдений P_θ определяется параметром из параметрического множества Θ, причем нулевая гипотеза H_0 состоит в том, что $\theta \in \Theta_0 \subset \Theta$, а альтернатива H_1 состоит в том, что $\theta \in \Theta_1 = \Theta \setminus \Theta_0$.

Теорема 2. *Пусть последовательность статистик $\{T_n\}$ удовлетворяет следующим условиям:*

1) $T_n \to b(\theta)$ по $\mathbf{P}_\theta$-вероятности, $\theta \in \Theta_1$, где $-\infty < b(\theta) < \infty$, и

2) $\lim\limits_{n\to\infty} n^{-1} \ln \mathbf{P}_\theta(T_n \geq a) = -k(a)$ для любого $\theta \in \Theta_0$ и любых a из некоторого открытого интервала I, где функция k непрерывна на I, причем $\{b(\theta), \theta \in \Theta_1\} \subset I$.

Тогда при всех $\theta \in \Theta_1$ точный наклон $c_T(\theta)$ существует и вычисляется по формуле

$$c_T(\theta) = 2k(b(\theta)).$$

Теперь определим расстояние Кульбака — Лейблера $K(\theta)$ [10, 12] между альтернативой и нулевой гипотезой H_0. Так как в нашем случае гипотеза H_0 сложная, то для альтернативной плотности $f(x, \theta)$ величина $K(\theta)$ определяется следующим образом:

$$K(\theta) = \inf_{v\in\mathbb{R}} \int\limits_{-\infty}^{\infty} \ln \frac{f(x,\theta)}{l(x+v)} f(x,\theta)dx. \tag{4}$$

Верхняя граница для точного наклона дается, как известно [10, 12], величиной $2K(\theta)$. Поэтому естественно определить локальную бахадуровскую эффективность последовательности $\{T_n\}$, как это обычно принято, формулой

$$eff_T = \lim_{\theta\to 0} \frac{c_T(\theta)}{2K(\theta)}. \tag{5}$$

4. Вычисление информации Кульбака — Лейблера. Сначала опишем альтернативы $f_i(x, \theta), x \in \mathbb{R}, i = 1, 2, 3$, которые мы будем рассматривать в этой работе:

1) альтернатива масштаба с плотностью

$$f_1(x,\theta) = \frac{e^{\theta + xe^{\theta}}}{(1+e^{xe^{\theta}})^2}, \quad \theta \geqslant 0;$$

2) альтернатива гиперболического косинуса с плотностью

$$f_2(x,\theta) = \frac{1}{4\,\mathrm{ch}^{2+\theta}(\frac{x}{2})} = \frac{2^{\theta} e^{\frac{(2+\theta)x}{2}}}{(1+e^{x})^{2+\theta}}, \quad \theta \geqslant 0;$$

3) синус-альтернатива в духе работы [13] с функцией распределения при малых θ

$$F_3(x,\theta) = L(x) - \theta \sin(2\pi L(x))$$

и плотностью

$$f_3(x,\theta) = l(x) - 2\pi\theta \cos(2\pi L(x)) l(x).$$

Все эти альтернативы при $\theta = 0$ переходят в логистическое распределение.

Для рассматриваемых альтернатив несложно показать, что инфимум в формуле (4) достигается при $v = 0$. Тогда при естественных условиях регулярности [10, § 4], выполняющихся для рассматриваемых семейств плотностей, справедливо следующее соотношение [10]:

$$K(\theta) \sim \frac{I(0)\cdot\theta^2}{2} \quad \text{при} \quad \theta \to 0,$$

где $I(0)$ — информация Фишера в нуле для альтернативной плотности $f(x,\theta)$, которая равна

$$I(0) = \int\limits_{-\infty}^{\infty} \frac{|f'_{\theta}(x,0)|^2}{f(x,0)} dx.$$

Поэтому формула для локальной бахадуровской эффективности в (5) преобразуется следующим образом:

$$eff_T = \lim_{\theta\to 0} \frac{c_T(\theta)}{I(0)\theta^2}. \tag{6}$$

Теперь найдем поведение $K_j(\theta)$, $j = 1, 2, 3$, для наших трех альтернатив при $\theta \to 0$. Используя таблицы интегралов или численное интегрирование, получаем

$$K_1(\theta) \sim \frac{\theta^2}{2} \int\limits_{-\infty}^{\infty} \frac{e^x(e^x + x + 1 - xe^x)^2}{(1+e^x)^4} dx = 0.7150\ldots\cdot\theta^2,$$

$$K_2(\theta) \sim \frac{\theta^2}{2} \int\limits_{-\infty}^{\infty} \frac{e^x(x + 2\ln(2) - 2\ln(e^x+1))^2}{4(e^x+1)^2} dx = 0.1358\ldots\cdot\theta^2,$$

$$K_3(\theta) \sim \frac{\theta^2}{2} \int\limits_{-\infty}^{\infty} \frac{4\pi^2 e^x \cos^2(2\pi\frac{e^x}{1+e^x})}{(1+e^x)^2} dx = \pi^2\ \theta^2 = 9.8696\ldots\cdot\theta^2.$$

5. Интегральная статистика и ее свойства. Вернемся к изучению интегральной статистики, определенной в (2). Рассмотрим вспомогательную функцию

$$g(x,y,z) = \left(1 - e^{(\min(x,y)-z)}\right) I\left\{\min(x,y) < z\right\}, \quad x,y,z \in \mathbb{R}.$$

Статистика LU_n асимптотически эквивалентна U-статистике степени 3 с центрированным ядром

$$\Phi(x,y,z) = \frac{1}{2} - \frac{1}{3}\left(g(x,y;z) + g(y,z;x) + g(x,z;y)\right). \tag{7}$$

Найдем проекцию этого ядра:

$$\Psi(t) = \mathbb{E}\left(\Phi(X,Y,Z)|Z=t\right) = \mathbb{E}\left(\frac{1}{2} - \frac{1}{3}g(X,Y;t) - \frac{2}{3}g(X,t;Y)\right) =$$
$$= -\frac{2}{3}\left(Li_2(-e^t) + t\ln(e^t+1) - \frac{1}{2}\ln^2(e^t+1) + \frac{7e^t+1}{4(e^t+1)}\right), \tag{8}$$

где дилогарифм Эйлера Li_2 определяется формулой $Li_2(z) = -\int_0^z \frac{\ln(1-t)}{t}dt$, $z \in \mathbb{C}$.

Теперь вычислим дисперсию проекции. С помощью численного интегрирования получаем

$$\Delta^2 = \mathbb{E}\Psi^2(X) =$$
$$= \frac{4}{9}\int\limits_{-\infty}^{\infty}\left(Li_2(-e^x) + x\ln(e^x+1) - \frac{1}{2}\ln^2(e^x+1) + \frac{7e^x+1}{4(e^x+1)}\right)^2 \frac{e^x}{(1+e^x)^2}dx =$$
$$= \frac{4}{9}\cdot 0.00439\ldots \approx 0.00195.$$

Таким образом, ядро Φ невырождено, а тогда по теореме Хёффдинга [9] при $n \to \infty$ имеем

$$\sqrt{n}LU_n \xrightarrow{d} \mathcal{N}\left(0, 9\Delta^2\right).$$

Так как ядро Φ не только невырождено и центрировано, но и ограничено, то мы можем воспользоваться результатами о больших уклонениях U-статистик из работы [15].

Теорема 2. *При* $t > 0$

$$\lim_{n\to\infty} n^{-1}\ln\mathbb{P}(LU_n > t) = h(t),$$

где h *— некоторая непрерывная функция такая, что* $h(t) \sim -\frac{t^2}{18\Delta^2}$ *при* $t \to 0$.

Теперь мы можем вычислить локальный бахадуровский наклон нашей последовательности статистик, опираясь на теоремы 1 и 2. Ясно, что

$$c_{LU}(\theta) \sim \frac{b_{LU}^2(\theta)}{9\Delta^2} \quad \text{при} \quad \theta \to 0.$$

Для широкого класса невырожденных U-статистик асимптотика $b_{LU}(\theta)$ при $\theta \to 0$ уже вычислена в [16]. Сформулированные там условия регулярности ND выполнены для рассматриваемых альтернатив, поэтому можно написать соотношение

$$b_{LU}(\theta) \sim 3 \int\limits_{-\infty}^{\infty} \Psi(x) f'_\theta(x,0)dx \cdot \theta, \qquad \theta \to 0, \tag{9}$$

где $f(x,\theta)$ — альтернативная плотность. Пользуясь этой формулой, получим следующее выражение для локального бахадуровского наклона:

$$c_{LU}(\theta) \sim \Delta^{-2} \left(\int\limits_{-\infty}^{\infty} \Psi(x) f'_\theta(x,0)dx \right)^2 \theta^2, \qquad \theta \to 0. \tag{10}$$

Переходим к вычислению локальных эффективностей для трех указанных выше альтернатив.

5.1. Альтернатива масштаба. Используя формулы (9) и (10), получим следующее соотношение для локального бахадуровского наклона:

$$c_{LU}(\theta; f_1) \sim 1.1970\ldots \cdot \theta^2.$$

Следовательно, по формуле (6) локальная бахадуровская эффективность равна

$$eff_{LU}(f_1) = \lim_{\theta\to 0} \frac{c_{1,LU}(\theta)}{2K_1(\theta)} = \frac{1.1970\ldots}{1.4299\ldots} \approx 0.837.$$

5.2. Альтернатива гиперболического косинуса. Вычислим локальный бахадуровский наклон, используя формулы (9) и (10):

$$c_{LU}(\theta, f_2) \sim 0.1371\ldots \cdot \theta^2.$$

Следовательно, по формуле (6) локальная бахадуровская эффективность равна

$$eff_{LU}(f_2) = \lim_{\theta\to 0} \frac{c_{2,LU}(\theta)}{2K_2(\theta)} \approx \frac{0.1371\ldots}{0.2716\ldots} \approx 0.505.$$

5.3. Синус-альтернатива. По формулам (9) и (10) получим следующее соотношение для локального бахадуровского наклона:

$$c_{LU}(\theta, f_3) \sim 14.978\ldots \cdot \theta^2.$$

Следовательно, по формуле (6) локальная бахадуровская эффективность равна

$$eff_{LU}(f_3) = \lim_{\theta\to 0} \frac{c_{3,LU}(\theta)}{2K_3(\theta)} = \frac{14.978\ldots}{19.739\ldots} \approx 0.759.$$

Таблица локальных бахадуровских эффективностей для интегрального критерия будет приведена ниже вместе с эффективностями для критерия Колмогорова.

6. Статистика типа Колмогорова. Рассмотрим теперь критерий типа Колмогорова, основанный на статистике KU_n из формулы (3). Предельное распределение статистики (3) неизвестно, и для нахождения критических значений следует рассчитывать на результаты моделирования.

Эту статистику можно рассматривать как супремум по t семейства модулей U-статистик с ядрами

$$\Phi_1(X,Y;t) = \big(1 - e^{(\min(X,Y)-t)}\big) I\{\min(X,Y) < t\} - \\ - \frac{1}{2}\left(I\{X < t\} + I\{Y < t\}\right), \; t \in \mathbb{R}. \quad (11)$$

Эти ядра ограничены и в силу рассматриваемой характеризации центрированы. Чтобы применить результат о больших уклонениях для таких U-статистик, вычислим проекцию ядра. После обширных вычислений получаем

$$\Psi_1(s,t) = \mathbb{E}\left(\Phi_1(X,Y;t)|Y = s\right) =$$

$$= \mathbb{E}\left\{(1 - e^{\min(X,s)-t}) I\{\min(X,s) < t\}\right\} - \frac{1}{2}\left(\mathbb{P}(X < t) + I\{s < t\}\right) =$$

$$= \frac{e^{\min(s,t)}(1 + e^{-t})}{1 + e^{\min(s,t)}} - e^{-t}\ln(e^{\min(s,t)} + 1) + \frac{I\{s < t\}(1 - e^{s}(1 + 2e^{-t}))}{2(1 + e^{s})} - \frac{e^{t}}{2(1 + e^{t})}.$$

Теперь найдем функцию дисперсии $\Delta_1^2(t) := \mathbb{E}_X \Psi_1^2(X,t)$ рассматриваемого семейства ядер как функцию от t. После громоздких вычислений, которые мы опускаем, получаем, что

$$\Delta_1^2(t) = \frac{e^{3t} + 8e^{2t} + 8e^{t} - 4(e^{t} + 1)(e^{t} + 2)\ln(e^{t} + 1)}{4e^{2t}(e^{t} + 1)^2}, \quad t \in \mathbb{R}.$$

Теперь остается вычислить ее супремум. Для удобства обозначим e^t за новую переменную y, и вычислим супремум следующей функции при $y > 0$:

$$m(y) = \frac{y^3 + 8y^2 + 8y - 4(y + 1)(y + 2)\ln(y + 1)}{4y^2(y + 1)^2}.$$

С помощью Wolfram Mathematica находим, что супремум этой функции достигается при $y = 1.3846\ldots$, и равен $0.02322\ldots$ Для исходной функции супремум останется тем же, и достигается он в точке $t = \ln y = 0.3255\ldots$

Итак,

$$\Delta_1^2 = \sup_{t\in\mathbb{R}} \Delta_1^2(t) = 0.02322\ldots$$

Поскольку семейство ядер центрировано и ограничено, мы можем применить при справедливости H_0 теорему о больших уклонениях для U-эмпирических статистик Колмогорова [17], что приводит к следующему соотношению.

Теорема 3. *При* $z > 0$

$$\lim_{n\to\infty} n^{-1} \ln \mathbb{P}\left\{KU_n > z\right\} = w(z),$$

где w *— некоторая непрерывная функция на* $\mathbb{R}^+$, *для которой* $w(z) \sim -\frac{z^2}{8\Delta_1^2}$, $z \to 0$.

Опираясь на теорему 3, аналогично формуле (10), можно получить следующее выражение для локального бахадуровского наклона статистики типа Колмогорова [14]:

$$c_{KU}(\theta) = \Delta_1^{-2} \sup_{t\in\mathbb{R}} \left(\int\limits_{-\infty}^{\infty} \Psi_1(x;t) f'_\theta(x,0)dx \right)^2 \cdot \theta^2. \quad (12)$$

Вычислим теперь для этой статистики эффективности для рассматриваемых альтернатив.

6.1. Альтернатива масштаба. Используя формулу (12), получим следующее соотношение для локального бахадуровского наклона:

$$c_{KU,1}(\theta) = \Delta_1^{-2} \sup_{t \in \mathbb{R}} \left(\frac{1}{1+e^t} - \frac{\ln(e^t+1)}{e^t} \right)^2 \theta^2.$$

При помощи Wolfram Mathematica находим, что супремум достигается в точке $t = 0.7713\ldots$ и равен $0.01170\ldots$ Таким образом, бахадуровский наклон удовлетворяет соотношению

$$c_{KU,1}(\theta) \sim 0.5033\ldots \cdot \theta^2, \quad \theta \to 0,$$

и, подставляя в формулу (6), получаем, что локальная бахадуровская эффективность равна

$$eff_{KU,1} = \lim_{\theta \to 0} \frac{c_{KU,1}(\theta)}{2K_1(\theta) \cdot \theta^2} = \frac{0.5033\ldots}{1.4299\ldots} \approx 0.352.$$

Получившаяся эффективность существенно меньше эффективности интегральной статистики для той же альтернативы, что типично для статистик Колмогорова [12].

6.2. Альтернатива гиперболического косинуса. При вычислении локального бахадуровского наклона используем формулу (12):

$$c_{KU,f_2} =$$
$$= \sup_{t>0} \left(\frac{(2Li_2(-t) - \ln^2(t+1))(1+t) + 2t(2 - \ln t - t\ln t) + \ln(t+1)(t+t^2+2t\ln t - 2)}{4\Delta_1 t(t+1)} \right)^2 \theta^2.$$

При помощи Wolfram Mathematica получаем, что супремум достигается в точке $t \approx 2.2933\ldots$ и равен $0.00118\ldots$ Следовательно, бахадуровский наклон равен

$$c_{KU,f_2} \sim 0.0509\ldots \cdot \theta^2,$$

и, подставляя в формулу (6), получаем, что локальная бахадуровская эффективность равна

$$eff_{KU,f_2} = \lim_{\theta \to 0} \frac{c_{KU,f_2}(\theta)}{2K_2(\theta) \cdot \theta^2} = \frac{0.0509\ldots}{0.2716\ldots} \approx 0.1874.$$

6.3. Синус-альтернатива. По формуле (12) получим следующее соотношение для локального бахадуровского наклона:

$$c_{KU,f_3} = \Delta_1^{-2} \sup_{z \in [0,1]} z^{-1} \left((1-z)(Si(2\pi(1-z)) - Si(2\pi)) \right)^2 \theta^2,$$

где $Si(x) = \int\limits_0^x \frac{\sin(t)}{t} dt$. При помощи Wolfram Mathematica получаем, что супремум достигается в точке $z \sim 0.324\ldots$ и равен примерно $0.3669\ldots$, так что бахадуровский наклон равен

$$c_{KU,f_3}(\theta) \sim 15.801\ldots \cdot \theta^2.$$

Мы получаем, что по формуле (6) локальная бахадуровская эффективность равна

$$eff_{KU,f_3} = \lim_{\theta\to 0} \frac{c_{KU,f_3}(\theta)}{2K_3(\theta)} = \frac{15.801\ldots}{19.739\ldots} \approx 0.800.$$

В этом случае эффективность больше, чем в случае интегральной статистики, что происходит довольно редко.

7. Условия локальной асимптотической оптимальности. В этом разделе мы выясним условия локальной асимптотической оптимальности по Бахадуру для последовательностей статистик LU_n и KU_n, то есть опишем структуру альтернатив, при которых статистика имеет максимальную локальную эффективность. Это выполняется при следующем условии [12, 16]:

$$c_T(\theta) \sim 2K(\theta) \quad \text{при} \quad \theta \to 0.$$

Пусть $f(x,\theta)$ — альтернативная плотность в точке θ. Предположим, что выполнены следующие условия регулярности:

$$\frac{d}{d\theta}\int\limits_{-\infty}^{\infty} xf(x,\theta)dx = \int\limits_{-\infty}^{\infty} xf'_\theta(x,\theta)dx \text{ при всех } \theta, \tag{13}$$

$$\int\limits_{-\infty}^{\infty} (f'_\theta(x,0))^2 e^{-x}(1+e^x)^2 dx < \infty. \tag{14}$$

Выведем условия оптимальности для плотностей, удовлетворяющих условиям (13) и (14).

7.1. Интегральная статистика. Сначала рассмотрим интегральную статистику LU_n с ядром $\Phi(x,y,z)$ из (7) и проекцией $\Psi(t)$ из (8). Воспользуемся асимптотикой для $b_{LU}(\theta)$ из (9) и выпишем выражение для локальной бахадуровской эффективности:

$$eff_{LU} = \frac{\left(\int\limits_{-\infty}^{\infty} \Psi(x)f'_\theta(x,0)dx\right)^2}{\Delta^2 I(0)} = \frac{\left(\int\limits_{-\infty}^{\infty} \Psi(x)f'_\theta(x,0)dx\right)^2}{\left(\int\limits_{-\infty}^{\infty} \Psi^2(x)l(x)dx\right)\left(\int\limits_{-\infty}^{\infty} \frac{(f'_\theta(x,0))^2}{l(x)}dx\right)}.$$

Локальная асимптотическая оптимальность по Бахадуру статистики LU_n означает, что выражение в правой части равно 1. Из неравенства Коши — Буняковского — Шварца следует, что это выполняется тогда и только тогда, когда $f'_\theta(x,0) = C_1\Psi(x)l(x)$, для некоторой константы $C_1 > 0$. Распределения, для которых функция $f'_\theta(x,0)$ имеет такой вид, образуют область локальной асимптотической оптимальности (ЛАО), в классе функций удовлетворяющих условиям (13), (14).

Простейшим примером альтернативной плотности из этого класса может служить «подправленная» логистическая плотность

$$f(x,\theta) = l(x)\,(1-\theta\Psi(x)) \quad (\text{при малых } \theta).$$

7.2. Статистика типа Колмогорова. Теперь мы рассмотрим статистику типа Колмогорова с семейством ядер $\Phi_1(x,y;t)$ из (11) и проекцией $\Psi_1(X,t)$ из § 6. С помощью формулы (12) для бахадуровского наклона получаем следующее выражение для локальной бахадуровской эффективности (ср. с [14]):

$$eff_{KU} = \frac{\sup\limits_{t\in\mathbb{R}} \left(\int\limits_{-\infty}^{\infty} \Psi_1(x;t) f'_\theta(x,0)dx \right)^2}{\Delta_1^2 I(0)} = \frac{\sup\limits_{t\in\mathbb{R}} \left(\int\limits_{-\infty}^{\infty} \Psi_1(x;t) f'_\theta(x,0)dx \right)^2}{\sup\limits_{t\in\mathbb{R}} \left(\int\limits_{-\infty}^{\infty} \Psi_1^2(x;t) l(x)dx \cdot \int\limits_{-\infty}^{\infty} \frac{(f'_\theta(x,0))^2}{l(x)}dx \right)}.$$

Мы снова применим неравенство Коши — Буняковского — Шварца ко всей дроби и получим, что eff_{KU} равно 1 тогда и только тогда, когда $f'_\theta(x,0) = C_2\Psi_1(x;t_0)l(x)$ для $t_0 = \arg\sup\limits_{t\in\mathbb{R}} \Delta_1^2(t)$ и некоторой константы $C_2 > 0$. Альтернативные плотности такого вида образуют область ЛАО в рассматриваемом классе.

Простейшим примером может служить плотность

$$f(x,\theta) = l(x)\left(1 + \theta\Psi_1(x,t_0)\right) \quad (\text{при малых } \theta).$$

8. Заключение. Мы построили два новых критерия согласия для логистического семейства со сдвигом, основанные на характеризации, использующей случайные экспоненциальные сдвиги. Для соответствующих статистик найдена логарифмическая асимптотика вероятностей больших уклонений при нулевой гипотезе и вычислена локальная бахадуровскую эффективность для ряда подходящих альтернатив.

Для сравнения локальных бахадуровских эффективностей построенных критериев соберем их в таблицу.

Локальные бахадуровские эффективности для LU_n и DU_n

Альтернатива	LU_n	DU_n
f_1	0.837	0.353
f_2	0.505	0.187
f_3	0.759	0.800

В большинстве случаев интегральный критерий превосходит по локальной бахадуровской эффективности критерий типа Колмогорова, кроме случая f_3, где оба критерия показывают довольно высокие значения эффективности. Стоит отметить, что интегральный критерий дает высокую эффективность для альтернативы масштаба f_1, в то время, как колмогоровский критерий значительно уступает ему в этом случае. Выше для каждого из критериев были построены специальные альтернативы, при которых данный критерий оказывается локально оптимальным в бахадуровском смысле.

Литература

1. *Balakrishnan N.* Handbook of the logistic distribution. CRC Press, 1991.

2. *Stephens M. A.* Tests of fit for the logistic distribution based on the empirical distribution function // Biometrika. 1979. Vol. 66, N 3. P. 591–595.

3. *Nikitin Y. Y.* Tests based on characterizations, and their efficiencies: a survey // Acta et Commentationes Universitatis Tartuensis de Mathematica. 2017. Vol. 21, N 1. P. 3–24.
4. *Volkova K. Yu.* New tests for the logistic distribution based on the functionals of U-empirical process // Abstracts of 12th Intern. Vilnius Conf. on Prob. and Math. Stat., Vilnius, Vtex, 2018. P. 297.
5. *Chin-Yuan Hu, Gwo Dong Lin.* Characterizations of the logistic and related distributions // Journ. of Mathem. Anal. and Appl. 2018. Vol. 463, N 1. P. 79–92.
6. *Galambos J.* Characterizations // In: Balakrishnan N. Handbook of the logistic distribution. CRC Press, 1991. P. 169–188.
7. *Lin G. D., Hu C. Y.* On characterizations of the logistic distribution // Journ. Statist. Plann. Infer. 2008. Vol. 138, N 4. P. 1147–1156.
8. *Ahsanullah M., Yanev G. P., Onica C.* Characterizations of Logistic Distribution Through Order Statistics with Independent Exponential Shifts // Stochastics and Quality Control. 2011. Vol. 27, N 1. P. 85–96.
9. *Королюк В. С., Боровских Ю. В.* Теория U-статистик. Киев: Наукова думка, 1989.
10. *Bahadur R. R.* Some limit theorems in statistics. Philadelphia: SIAM, 1971.
11. *Bahadur R. R.* Stochastic comparison of tests // Ann. Math. Stat. 1960. Vol. 31, N 2. P. 276–295.
12. *Никитин Я. Ю.* Асимптотическая эффективность непараметрических статистических критериев. М.: Наука, 1995.
13. *Ley C., Paindaveine D.* Le Cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions // Journal of Nonparametric Statistics. 2009. Vol. 21, N 8. P. 943–967.
14. *Nikitin Ya. Yu., Volkova K. Yu.* Efficiency of Exponentiality Tests Based on a Special Property of Exponential Distribution // Mathematical Methods of Statistics. 2016. Vol. 25, N 1. P. 54–66.
15. *Nikitin Ya. Yu., Ponikarov E. V.* Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics // Proc. of St. Petersburg Math. Soc. 1999. Vol. 7. P. 124–167. Engl. transl. in AMS Transl., ser. 2, 2001. Vol. 203. P. 107–146.
16. *Nikitin Ya. Yu., Peaucelle I.* Efficiency and local optimality of distribution-free tests based on U- and V-statistics // Metron. 2004. Vol. LXII. P. 185–200.
17. *Nikitin Ya. Yu.* Large deviations of U-empirical Kolmogorov-Smirnov tests, and their efficiency // J. Nonpar. Stat. 2010. Vol. 22, N 5. P. 649–668.

Статья поступила в редакцию 16 ноября 2018 г.;
после доработки 16 ноября 2018 г.;
рекомендована в печать 20 декабря 2018 г.

Контактная информация:

Никитин Яков Юрьевич — д-р физ.-мат. наук; y.nikitin@spbu.ru
Рагозин Илья Андреевич — научный сотрудник; ragza@yandex.ru

Goodness-of-fit tests based on a characterization of logistic distribution

Ya. Yu. Nikitin, I. A. Ragozin

St. Petersburg State University, Universitetskaya nab., 7–9, St. Petersburg, 199034, Russian Federation

For citation: Nikitin Ya. Yu., Ragozin I. A. Goodness-of-fit tests based on a characterization of logistic distribution. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2019, vol. 6 (64), issue 2, pp. 241–252. https://doi.org/10.21638/11701/spbu01.2019.206 (In Russian)

The logistic family of distributions belongs to the class of important families in Probability and Statistics. However, the goodness-of-fit tests for the composite hypothesis of belonging to the logistic family with unknown location parameter against the general alternatives are almost unexplored. We propose two new goodness-of-fit tests: the integral and Kolmogorov-type, based on the recent characterization of logistic family by Hu and Lin. We discuss asymptotic properties of new tests and calculate their Bahadur efficiency for natural alternatives.

Keywords: characterization of distributions, logistic distribution, asymptotic efficiency, large deviations, Kullback — Leibler information.

References

1. Balakrishnan N., *Handbook of the logistic distribution* (CRC Press, 1991).

2. Stephens M. A., "Tests of fit for the logistic distribution based on the empirical distribution function", *Biometrika* **66**(3), 591–595 (1979).

3. Nikitin Y. Y., "Tests based on characterizations, and their efficiencies: a survey", *Acta et Commentationes Universitatis Tartuensis de Mathematica* **21**(1), 3–24 (2017).

4. Volkova K. Yu., "New tests for the logistic distribution based on the functionals of U-empirical process", *Abstracts of 12th Intern. Vilnius Conf. on Prob. and Math. Stat., Vilnius, Vtex*, 297 (2018).

5. Chin-Yuan Hu, Gwo Dong Lin, "Characterizations of the logistic and related distributions", *Journ. of Mathem. Anal. and Appl.* **463**(1), 79–92 (2018).

6. Galambos J., *Characterizations*, in: Balakrishnan N., *Handbook of the logistic distribution*, 169–188 (CRC Press, 1991).

7. Lin G. D., Hu C. Y., "On characterizations of the logistic distribution", *Journ. Statist. Plann. Infer.* **138** (4), 1147–1156 (2008).

8. Ahsanullah M., Yanev G. P., Onica C., "Characterizations of Logistic Distribution Through Order Statistics with Independent Exponential Shifts", *Stochastics and Quality Control* **27** (1), 85–96 (2011).

9. Korolyuk V. S., Borovskich Y. V., *Theory of U-statistics* (Springer Science & Business Media, 2013).

10. Bahadur R. R., *Some limit theorems in statistics* (Philadelphia, SIAM, 1971).

11. Bahadur R. R., "Stochastic comparison of tests", *Ann. Math. Stat.* **31**(2), 276–295 (1960).

12. Nikitin Y., *Asymptotic efficiency of nonparametric tests* (Cambridge University Press, NY, 1995; 2nd ed., paperback, 2009).

13. Ley C., Paindaveine D., "Le Cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions", *Journal of Nonparametric Statistics* **21** (8), 943–967 (2009).

14. Nikitin Ya. Yu., Volkova K. Yu., "Efficiency of Exponentiality Tests Based on a Special Property of Exponential Distribution", *Mathematical Methods of Statistics* **25** (1), 54–66 (2016).

15. Nikitin Ya. Yu., Ponikarov E. V., "Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics", *AMS Transl., ser.2* **203**, 107–146 (2001).

16. Nikitin Ya. Yu., Peaucelle I., "Efficiency and local optimality of distribution-free tests based on U- and V-statistics", *Metron* **LXII**, 185–200 (2004).

17. Nikitin Ya. Yu., "Large deviations of U-empirical Kolmogorov-Smirnov tests, and their efficiency", *J. Nonpar. Stat.* **22** (5), 649–668 (2010).

Received: November 16, 2018
Revised: November 16, 2018
Accepted: December 20, 2018

Author's information:

Yakov Yu. Nikitin — y.nikitin@spbu.ru,
Ilya A. Rogozin — ragza@yandex.ru