Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра теории вероятностей и математической статистики

ДИПЛОМНАЯ РАБОТА

студента группы 14.С02-мм
Курова Андрея Олеговича

Статистические критерии, основанные на ядерных оценках плотности

Научный руководитель, проф.
Никитин Я. Ю.

Рецензент, доц.
Литвинова В.В.

Санкт-Петербург
-2019 г.-

Saint-Petersburg State University
Faculty of Mathematics and Mechanics
Department of Probability Theory and Mathematical Statistics

Andrei O. Kurov

Graduation Thesis

Statistical tests based on kernel density estimators

Thesis supervisor, D.Sci., Prof.
Ya.Yu. Nikitin

Thesis Reviewer, doc.
V.V. Litvinova

Saint Petersburg
-2019-

Содержание

1 Введение

Проверка статистических гипотез - одна из основных задач математической статистики. Классическими разделами проверки гипотез принято считать проверку согласия, проверку симметрии, проверку однородности, проверку независимости и проверку случайности [14]. В настоящей работе мы уделяем главное внимание проверке симметрии, где на основании выборки проверяется симметричность генеральной плотности распределения, а также проверке однородности, где на основании двух выборок проверяется их принадлежность к одному и тому же распределению.

В центре наших интересов лежит интегральный критерий симметрии, основанный на L_1-расстоянии для ядерных оценок плотностей. Этот критерий был предложен в 2006 г. в статье французских статистиков Берраху и Луани [8]. Авторы нашли большие уклонения критериальной статистики при гипотезе симметрии, вычислили ее бахадуровскую эффективность для альтернативы сдвига в случае простейших распределений и исследовали вопрос о ее локальной асимптотической оптимальности.

Однако далее эта статистика никем не исследовалась. Нас интересует вопрос о том, какова ее локальная бахадуровская эффективность для других альтернатив (скошенных, лемановских, загрязнения), а также вопрос о сравнении рассматриваемой статистики с другими известными критериями симметрии.

В 2009 г. испанские математики Мартинес-Камблор, Корраль и Лопес [12] опубликовали статью, аналогичную работе Берраху и Луани, но для проверки однородности двух выборок. Помимо нахождения больших уклонений и вычисления бахадуровских локальных точных наклонов для альтернативы сдвига, они сравнили свой критерий со знаменитым критерием инверсий Манна-Уитни. Мы дополняем их работу вычислением бахадуровской асимптотической эффективности для других альтернатив и сравнением с другими критериями. Наша работа позволяет оценить практическую работоспособность рассматриваемых критериев в той или иной модельной ситуации.

2 Известные результаты

2.1 Проверка симметрии

Пусть $X_1, X_2, .., X_n$ - независимые одинаково распределённые наблюдения с генеральной функцией распределения (ф.р.) F и плотностью f. Задачей о проверке гипотезы симметрии H_0 называется проверка с помощью данной выборки следующего утверждения:

$$1 - F(x) - F(-x) = 0, \qquad \forall x \in \mathbb{R}^1.$$

В качестве альтернативы выдвигается утверждение, что это равенство нарушается хотя бы для одного x, либо параметрическая альтернатива, состоящая в том, что генеральная ф.р. - известная абсолютно непрерывная ф.р. $G(x, \theta)$, $x \in \mathbb{R}^1, \theta \in \Theta$, где Θ - это некоторая правосторонняя окрестность 0 в $\mathbb{R}^1$. При этом предполагается, что $1 - G(x, \theta) - G(-x, \theta) = 0 \;\; \forall x \in \mathbb{R}^1$ выполнено лишь в случае $\theta = 0$. Далее для краткости обозначаем для любой ф.р. F, заданной на $\mathbb{R}$, $\Delta F(x) = 1 - F(x) - F(-x)$, а если ф.р. F задана на $[0, 1]$, то полагаем $\Delta F(x) = 1 - F(x) - F(1 - x)$.

Существует множество критериев для проверки данной гипотезы. Например, один из самых известных и старых критериев, известный с начала XVIII в. [1], - это критерий знаков, основанный на статистике

$$S_n = n^{-1} \sum_{i=1}^{n} \left[\mathbf{1}_{\{X_i > 0\}} - \frac{1}{2} \right].$$

Опишем еще несколько хорошо известных и детально изученных статистик для проверки симметрии.

Смирнов [17] ввёл в рассмотрение аналог статистики Колмогорова:

$$I_n = \sup_{-\infty < x < \infty} |\Delta F_n(x)|,$$

где $F_n(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i < t\}$, $t \in \mathbb{R}$ - обычная эмпирическая ф.р.

Ченцов [18] рассматривал статистику, аналогичную статистике ω_n^2:

$$R_n^2 = \int_{-\infty}^{\infty} (\Delta F_n(x))^2 \, dF_n(x)$$

Хилл и Рао [10] рассматривали статистику, аналогичную статистике Ватсона:

$$N_n^2 = \int\limits_{-\infty}^{\infty} \left(\Delta F_n(x) - \int\limits_{-\infty}^{\infty} \Delta F_n(y) dF_n(y) \right)^2 dF_n(x)$$

Аббакумов [13] в конце 80-х годов прошлого века ввёл в рассмотрение аналоги статистик Ватсона-Дарлинга:

$$H_n = \sup_{-\infty<x<\infty} \left| \Delta F_n(x) - \int\limits_{-\infty}^{\infty} \Delta F_n(y) dF_n(y) \right|,$$

$$H_n^+ = \sup_{-\infty<x<\infty} \left[\Delta F_n(x) - \int\limits_{-\infty}^{\infty} \Delta F_n(y) dF_n(y) \right],$$

$$H_n^- = \sup_{-\infty<x<\infty} \left[\int\limits_{-\infty}^{\infty} \Delta F_n(y) dF_n(y) - \Delta F_n(x) \right].$$

Также можно рассмотреть обобщение статистики R_n^2 [16] вида

$$R_{n,q}^k = \int\limits_{-\infty}^{\infty} (\Delta F_n(x))^k \, q(F_n(x)) dF_n(x),$$

где $k \in \mathbb{N}$, а q - неотрицательная весовая функция на $[0,1]$, симметричная относительно точки $\frac{1}{2}$ и удовлетворяющая определённым условиям гладкости.

Еще один вариант статистики для проверки симметрии -

$$BH_n = \sqrt{n} \sup_{x \geq 0} |D_n(x) - G_n(x)|,$$

статистика, введённая Барингхаузом и Хенце [4], где D_n - эмпирическая ф.р. $|X_j|, 1 \leq j \leq n$, а G_n - U-эмпирическая ф.р. C_n^2 случайных величин $|\max(X_j, X_k)|$, $1 \leq j \leq k \leq n$.

2.2 Локальная бахадуровская эффективность

Для сравнения двух последовательностей статистик, построенных по выборке объёма n и предназначенных для проверки гипотезы H_0 против альтернативы H_1, вводится относительная эффективность. Введём обозначение:$N_T(\alpha, \beta, \theta)$ - это объём выборки, необходимый для того, чтобы последовательность T_n при уровне значимости α и альтернативном значении параметра θ достигла мощности β (точно так же определяется $N_V(\alpha, \beta, \theta)$). Относительной эффективностью последовательности статистик T_n и V_n в таком случае называют величину

$$e_{T,V}(\alpha, \beta, \theta) = \frac{N_V(\alpha, \beta, \theta)}{N_T(\alpha, \beta, \theta)}.$$

Это классическое определение восходит к известному статистику Э.Питмену [14].

Достоинства относительной эффективности общепризнаны, но главный недостаток заключается в том, что её практически невозможно вычислить даже для самых простых последовательностей статистик. Эту трудность обходят разными путями; в частности, Бахадур [5], [6] предложил вычислять асимптотическую относительную эффективность (АОЭ), вычисляя предельное значение $e_{T,V}(\alpha, \beta, \theta)$ при стремлении к нулю уровня значимости: $\alpha \to 0$. Есть и другие определения АОЭ, принадлежащие самому Питмену, Чернову, Ходжесу и Леману и другим [16].

Величину

$$e^B_{T,V}(\beta, \theta) := \lim_{\alpha \to 0} e_{T,V}(\alpha, \beta, \theta)$$

называют АОЭ по Бахадуру. Вычислить её можно по формуле

$$e^B_{T,V}(\beta, \theta) = \frac{c_T(\theta)}{c_V(\theta)},$$

где $c_T(\theta)$ - положительная неслучайная функция параметра θ, называемая точным наклоном последовательности T_n по Бахадуру. Сам же Бахадур и предложил наиболее удобный способ вычисления точных наклонов.

Пусть Θ - некоторая правосторонняя окрестность нуля в $\mathbb{R}^1$, $\Theta_0 \subset \Theta$, $\Theta_1 = \Theta \setminus \Theta_0$. Введём гипотезу, альтернативу и некоторые обозначения:

$$H_0 : \theta \in \Theta_0, \qquad H_1 : \theta \in \Theta_1,$$

$$F_n(t, \theta) := \mathbb{P}_\theta(T_n < t), \, t \in \mathbb{R},$$

$$G_n(t) := \inf\{F_n(t, \theta) : \theta \in \Theta_0\}.$$

Тогда имеет место [5] следующая теорема.

***Теорема Бахадура**: пусть последовательность $\{T_n\}$ удовлетворяет следующим двум условиям:*

1) $\{T_n\} \to b(\theta)$ по P_θ-вероятности, где $-\infty < b(\theta) < \infty$ для $\forall\theta$;

2) $\lim\limits_{n\to\infty} n^{-1}\ln[1-G_n(t)] = -r(t)$ для любого t из открытого интервала I, где r непрерывна на I и $\{b(\theta), \theta \in \Theta_0\} \in I$. Тогда для любого $\theta \in \Theta_0$ справедливо

$$c_T(\theta) = 2r(b_T(\theta)).$$

Отметим, что вычисление $r(t)$ - функции уклонений при H_0 - задача нетривиальная, связанная с вычислением логарифмической асимптотики вероятностей больших уклонений при основной гипотезе.

Бахадуру и Рагавачари также удалось вывести неравенство, являющееся своего рода аналогом неравенства Рао-Крамера в теории оценивания. Оно имеет следующий вид:

$$c_T(\theta) \leq 2K(\theta), \tag{1}$$

где $K(\theta) = \inf\big(\int\limits_{-\infty}^{\infty} \ln\frac{g(x,\theta)}{h(x)} g(x,\theta)dx : h \in \mathfrak{H}\big)$ - информация Кульбака-Лейблера, $g(x,\theta)$ - альтернативная плотность, а $\mathfrak{H}$ есть класс всех плотностей, отвечающих нулевой гипотезе.

Теперь мы можем ввести локальную бахадуровскую эффективность последовательности статистик $\{T_n\}$:

$$e_T^B := \lim_{\theta\to 0}\frac{c_T(\theta)}{2K(\theta)}.$$

В целом, информация Кульбака-Лейблера сложна для вычисления, но благодаря работам других авторов у нас имеются её асимптотики для некоторых случаев.

2.3 Критерий, основанный на ядерных оценках плотности

В поисках новых критериев симметрии статистики стали исследовать критерии симметрии, основанные не на эмпирической ф.р., а на (ядерных) оценках плотности. Можно рассмотреть нестрогую проверку симметрии, а именно: $H : \int|f(x) - f(-x)|dx = 0$ против альтернативы $A : \int|f(x) - f(-x)|dx > 0$.

Нестрогость выражается в том, что у плотности f может иметься множество "несимметричных" точек лебеговой меры 0, но на практике считают, что она эквивалентна рассматривавшейся ранее гипотезе H_0. Рассмотрим оценку функции плотности f, именуемую ядерной, она же оценка Парзена-Розенблатта:

$$f_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{a_n}\right),$$

где $\{a_n\}$ - последовательность положительных величин, $a_n \to 0$. Последовательность $_n$ часто называют шириной окна, а функцию K - ядром. Для проверки гипотезы H вводится следующая статистика, впервые рассмотренная в статье [7]:

$$V_n = \int |f_n(x) - f_n(-x)| dx$$

Благодаря статье [8] имеется следующая теорема:

Теорема Берраху-Луани*: пусть выполняются следующие условия:*

1) $\int K(z)dz = 1$.

2) Ядро K *симметрично п.в.*

3) Существует последовательность r_n *положительных чисел, стремящихся к бесконечности:* $\lim\limits_{n\to\infty} \frac{r_n}{na_n} = 0$, *таких, что* $\mathbb{P}(|X| \geq r_n) \leq w_n$, *и* $\forall \tau > 0$ *выполнено* $\lim\limits_{n\to\infty} \frac{w_n \exp(n\tau)}{n} = 0$.

Тогда, если $a_n \to 0$ *и* $na_n \to \infty$ *при* $n \to \infty$, *то выполнено:*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}(V_n > \lambda) = -g(\lambda),$$

где $g(\epsilon) = \inf(\Gamma_a(\epsilon), 0 \leq a \leq 1)$, *и, в свою очередь,*

$$\Gamma_a(\epsilon) = \min(\Gamma_a^+, \Gamma_a^-),$$

причем

$$\Gamma_a^+ = \left(a + \frac{\epsilon}{2}\right) \ln\left(1 + \frac{\epsilon}{2a}\right) + \left(1 - a - \frac{\epsilon}{2}\right) \ln\left(1 - \frac{\epsilon}{2(1-a)}\right), 0 < \epsilon < 2 - 2a;$$

$$\Gamma_a^- = \left(a - \frac{\epsilon}{2}\right) \ln\left(1 - \frac{\epsilon}{2a}\right) + \left(1 - a + \frac{\epsilon}{2}\right) \ln\left(1 + \frac{\epsilon}{2(1-a)}\right), 0 < \epsilon < 2a.$$

Из этой теоремы следует, что точный наклон по Бахадуру для статистики V_n эквивалентен выражению $2g(\int |f(x) - f(-x)|dx)$.

Если ввести альтернативу сдвига, $F_0(x-\theta)$ для $\theta > 0$, где F_0 обозначает симметричную ф.р., то при ней точный наклон по Бахадуру будет иметь вид:

$$c_V(\theta) \sim 2g(2\theta \int |f_0'(x)|dx),\ \theta \to 0.$$

Луани вывел асимптотику для g, а именно $g(\lambda) \sim \frac{1}{8}\lambda^2$, $\lambda \to 0$, следовательно, можно преобразовать выражение выше, и мы получаем:

$$c_V(\theta) \sim \theta^2 \left(\int |f_0'(x)|dx \right)^2, \text{ при } \quad \theta \to 0.$$

Определяя локальные индексы как коэффициенты при θ^2 при $\theta \to 0$ в локальных точных наклонах по Бахадуру, Берраху и Луани получили следующую таблицу локальных индексов для трех классических распределений выборки:

Статистика	Гаусс	Логистическое	Лаплас
V_n	0.637	0.25	1
BH_n	0.764	0.25	0.422
H_n	0.955	0.333	0.75
R_n^2	0.907	0.329	0.822
N_n^2	0.486	0.219	0.822

Для сравнения локальных точных наклонов по Бахадуру с их максимально возможным значением понадобится неравенство Хо [11], которое является частным случаем неравенства (1):

$$c(\theta) \leq 2K(\theta) := 2 \int f_\theta(x) \ln \frac{2f_\theta(x)}{f_\theta(x) + f_\theta(-x)} dx.$$

В типичных случаях информация Кульбака-Лейблера $K(\theta)$ имеет локальное асимптотическое поведение

$$K(\theta) \sim \frac{1}{2} I(f_0)\theta^2, \theta \to 0,$$

где $I(f_0) = \int\limits_{-\infty}^{\infty} \frac{f_0'^2(x)}{f_0(x)} dx$ - информация Фишера.

Обозначим через Φ класс абсолютно непрерывных плотностей f_0 на $\mathbb{R}$, удовлетворяющих асимптотике, выписанной выше с $0 < I(f_0) < \infty$.

Если при альтернативе наблюдения имеют плотность, отвечающую $f_0(x-\theta)$, $f_0 \in \Phi$, то условие локальной оптимальности по Бахадуру выполнено, когда выполнено равенство

$$\left(\int\limits_{-\infty}^{\infty} |f_0'(x)| \, dx \right)^2 = \int\limits_{-\infty}^{\infty} \frac{f_0'^2(x)}{f_0(x)} dx$$

Из этого получаем, по неравенству Коши-Буняковского-Шварца, что критерий на основе ядерных плотностей локально оптимален (в частности) для плотности Лапласа и для некоторого более широкого класса плотностей, определяемого следующей теоремой [8]:

Теорема*: Статистика V_n локально оптимальна по Бахадуру в классе распределений Φ для плотностей выборки, имеющих вид $cf_0(cx)$, $c > 0$, где*

$$f_0(x) = \begin{cases} \sum\limits_{i=0}^{n-1} \{c_i \exp((-1)^i x)\} \mathbf{1}_{[(-1)^i \ln(\frac{b_i}{c_i}), (-1)^i \ln(\frac{b_{i+1}}{c_i})]}(x) & \text{для } x \in (-\infty, 0] \\ f_0(-x) & \text{для } x \in [0, \infty), \end{cases}$$

где $0 = a_0 < a_1 < ... < a_n = \frac{1}{2}$ - разбиение интервала $[0, \frac{1}{2}]$,

$$b_0 = 0,\ b_i = (-1)^i a_i + 2\sum_{j=0}^{i} (-1)^{j+1} a_j > 0, \text{ для } 1 \le i \le n,\ c_i = \frac{\prod\limits_{0 \le p \le [\frac{n-(i+1)}{2}]} b_{i+2p+1}^2}{b_n \prod\limits_{1 \le p \le [\frac{n-(i+1)}{2}]} b_{i+2p}^2}.$$

Далее нас интересует рассмотрение других альтернатив:

1) скошенной альтернативы с плотностью $2f(x)F(\theta x)$;

2) лемановской альтернативы с ф.р. $G(x,\theta) = F^{1+\theta}(x), \theta > 0$;

3) альтернативы загрязнения с ф.р. $G(x,\theta) = (1-\theta)F(x) + \theta F^{1+r}(x)$, $\theta > 0, r > 0$.

3 Альтернатива сдвига

Начнём с того, что дополним таблицу 1 для *альтернативы сдвига*, вычислив индексы ещё для нескольких симметричных распределений выборки. В работе Берраху и Луани [8] имеются следующие выражения для локальных точных наклонов:

$c_V(\theta) = \left(\int |f_0'(y)|dy\right)^2 \cdot \theta^2;$

$c_{BH}(\theta) = \left(27 \sup\limits_x [f_0(x) \int\limits_{-x}^{x} f_0(y)dy]^2\right) \cdot \theta^2;$

$c_H(\theta) = \left(12 \sup\limits_x (f_0(x) - \int f_0^2(y)dy)^2\right) \cdot \theta^2;$

$c_{R^2}(\theta) = \left(\pi^2 \int f_0^3(y)dy\right) \cdot \theta^2;$

$c_{N^2}(\theta) = \left(4\pi^2 \left[\int f_0^3(y)dy - \left(\int f_0^2(y)dy\right)^2\right]\right) \cdot \theta^2.$

Вычислим сначала локальные индексы для распределения Коши $C(0,1)$:

$$f_0(x) = \frac{1}{\pi}\frac{1}{x^2+1}.$$

Возьмём производную для вычисления первого локального индекса:

$$f_0'(x) = \frac{-2\pi x}{\pi^2(x^2+1)^2} = \frac{-2x}{\pi(x^2+1)^2}.$$

Будем обозначать локальные индексы $LI_V, LI_{BH}, LI_H, LI_R, LI_N$ для $V_n, BH_n, H_n, R_n^2, N_n^2$ соответственно. Получим

$$LI_V = \left(\int \frac{2x}{\pi(x^2+1)^2}dx\right)^2 = \frac{16}{\pi^2}\left(\int\limits_0^\infty \frac{x}{(x^2+1)^2}dx\right)^2 =$$

$$= \frac{16}{\pi^2}\left(\frac{1}{2(x^2+1)}\Bigg|_0^\infty\right)^2 = \frac{4}{\pi^2} \approx \mathbf{0.405}$$

$$LI_{BH} = 27\sup_x\left[\frac{1}{\pi(x^2+1)}\int\limits_{-x}^{x}\frac{1}{\pi(y^2+1)}dy\right]^2 =$$

$$= \frac{27}{\pi^4}\sup_x\left[\frac{1}{x^2+1}(\operatorname{arctg}(x) - \operatorname{arctg}(-x))\right]^2 = \frac{108}{\pi^4}\sup_x\left[\frac{\operatorname{arctg}(x)}{x^2+1}\right]^2.$$

Но

$$\left(\left[\frac{\operatorname{arctg}(x)}{x^2+1}\right]^2\right)' = 2\cdot\frac{\operatorname{arctg}(x)}{x^2+1}\cdot\left(\frac{1-2x\operatorname{arctg}(x)}{(x^2+1)^2}\right) = \frac{2\operatorname{arctg}(x) - 4x\operatorname{arctg}^2(x)}{(x^2+1)^3} = 0$$

Производная равна 0 $\Leftrightarrow$ или $arctg(x) = 0$, или $2x \operatorname{arctg}(x) = 1$.

В первом случае наше выражение будет нулём - очевидно, это не супремум. Во втором случае $x \approx \pm 0.765$ и, соответственно, $LI_2 \approx \mathbf{0.188}$.

Далее, рассмотрим

$$LI_H = 12 \sup_x \left(\frac{1}{\pi(x^2+1)} - \int \frac{1}{\pi^2(y^2+1)^2} dy \right)^2 =$$

$$= \frac{12}{\pi^2} \sup_x \left(\frac{1}{x^2+1} - \frac{1}{\pi} \int \frac{1}{(y^2+1)^2} dy \right)^2 =$$

$$= \frac{12}{\pi^2} \sup_x \left(\frac{1}{x^2+1} - \frac{1}{2\pi} \left(\frac{y}{y^2+1} + \operatorname{arctg}(y) \right) \Big|_{-\infty}^{\infty} \right)^2 =$$

$$= \frac{12}{\pi^2} \sup_x \left(\frac{1}{x^2+1} - \frac{1}{2} \right)^2 = \frac{3}{\pi^2} \approx \mathbf{0.304}.$$

В данном случае достаточно очевидно, что супремум достигается в нуле (левая дробь заключена между 0 и 1, разность - между -1/2 и 1/2, достигающихся на $\pm\infty$ и 0 соответственно), особых вычислений тут не нужно.

Переходим к статистике R_n^2, здесь получаем

$$LI_R = \pi^2 \int \frac{1}{\pi^3} \frac{1}{(x^2+1)^3} dx = \frac{1}{8\pi} \left(\frac{x(3x^2+5)}{(x^2+1)^2} + 3 \operatorname{arctg}(x) \right) \Big|_{-\infty}^{\infty} = \mathbf{0.375.}$$

Рассмотрим еще последовательность статистик N_n^2:

$$LI_N = 4\pi^2 \left[\int \frac{1}{\pi^3} \frac{1}{(y^2+1)^3} dy - \left(\int \frac{1}{\pi^2(y^2+1)^2} dy \right)^2 \right] =$$

$$= 4 \cdot LI_R - \frac{4}{\pi^2} \frac{\pi^2}{4} = 1.5 - 1 = \mathbf{0.5}.$$

Интересные вычисления связаны с плотностью гиперболического секанса:

$$f_0(x) = \frac{1}{\pi \cosh(x)} = \frac{2}{\pi(e^x + e^{-x})}.$$

Возьмём производную для вычисления первого локального индекса:

$$f_0'(x) = -\frac{\sinh(x)}{\pi \cosh^2(x)}.$$

Тогда

$$LI_V = \left(\int |-\frac{\sinh(x)}{\pi\cosh^2(x)}|dx\right)^2 = \frac{4}{\pi^2} \approx \mathbf{0.405};$$

$$LI_{BH} = 27\sup_x \left[\frac{1}{\pi\cosh(x)}\int\limits_{-x}^{x}\frac{1}{\pi\cosh(y)}dy\right]^2 \approx \mathbf{0.349};$$

$$LI_H = 12\sup_x \left[\frac{1}{\pi\cosh(x)} - \int\frac{1}{\pi^2\cosh^2(y)}dy\right]^2 \approx \mathbf{0.493};$$

$$LI_{R^2} = \pi^2\int\frac{1}{\pi^3\cosh^3(y)}dy = \mathbf{0.5};$$

$$LI_{N^2} = 4\pi^2\left[\int\frac{1}{\pi^3\cosh^3(y)}dy - \left(\frac{1}{\pi^2\cosh^2(y)}dy\right)^2\right] \approx \mathbf{0.379};$$

Ввиду наличия асимптотики для информации Кульбака-Лейблера [8]

$$K(\theta) \sim \frac{1}{2}I(f_0)\theta^2, \theta \to 0$$

и наличия нер-ва Хо ($c(\theta) \leq 2K(\theta)$), информация Фишера является верхней возможной границей для локальных индексов. При этом имеем следующие значения информации Фишера для альтернативы сдвига:

1) Коши: $I(\theta) = \frac{1}{2}$.

2) Гиперболический секанс: $I(\theta) = \frac{1}{2}$.

3) Гаусс: $I(\theta) = 1$.

4) Логистическое: $I(\theta) = \frac{1}{3}$.

5) Лаплас: $I(\theta) = 1$.

Отсюда получаем следующую таблицу локальных эффективностей по Бахадуру при альтернативе сдвига:

Статистика	Коши	Гиперб. секанс	Гаусс	Логистич.	Лаплас
V_n	0.810	0.810	0.637	0.750	1
BH_n	0.377	0.698	0.764	0.750	0.422
H_n	0.608	0.996	0.955	1	0.750
R_n^2	0.750	1	0.907	0.987	0.822
N_n^2	1	0.758	0.486	0.657	0.822

Мы видим, что рассматриваемый критерий симметрии вполне конкурентоспособен в сравнении с другими ходовыми критериями, причем для более широкого набора распределений выборки, чем это было у Берраху-Луани.

4 Скошенная альтернатива

Эта альтернатива появилась, по-видимому, век назад в малоизвестных работах в Италии, но получила известность в 80-х годах после работы Аззалини [2] и приобрела большую популярность. Недавно исследования по этой альтернативе были собраны в монографии [3]. Если F - симметричная ф.р. с плотностью f, то скошенной называется плотность

$$h(x,\theta) = 2f(x)F(\theta x),$$

а через $H(x,\theta)$ мы будем обозначать соответствующую ф.р.

При $\theta \to 0$ разложение в ряд Тейлора даёт

$$H(x,\theta) \sim F(x) + 2\theta f(0) \int\limits_{-\infty}^{x} uf(u)du.$$

Введём основную гипотезу и альтернативу аналогично предыдущим параграфам:

$$H_0 : \theta = 0, \quad H_1 : \theta > 0.$$

Чтобы определить точный наклон по Бахадуру для V_n и, соответственно, определить локальный индекс, для начала нужно определить некоторые характеристики для новой альтернативы.

Из статьи Louani [7] мы знаем, что $\lim\limits_{n\to\infty} n^{-1} \ln P(V_n \geq \epsilon) = -g(\epsilon)$, где $g(\epsilon) = \inf(\Gamma_a(\epsilon), 0 \leq a \leq 1)$, а, в свою очередь,

$$\Gamma_a^+ = \left(a + \frac{\epsilon}{2}\right)\ln\left(1 + \frac{\epsilon}{2a}\right) + \left(1 - a - \frac{\epsilon}{2}\right)\ln\left(1 - \frac{\epsilon}{2(1-a)}\right), 0 < \epsilon < 2 - 2a;$$

$$\Gamma_a^- = \left(a - \frac{\epsilon}{2}\right)\ln\left(1 - \frac{\epsilon}{2a}\right) + \left(1 - a + \frac{\epsilon}{2}\right)\ln\left(1 + \frac{\epsilon}{2(1-a)}\right), 0 < \epsilon < 2a.$$

Как можно видеть, при $\epsilon \to 0$

$$\Gamma_a^+ \sim \left(a+\frac{\epsilon}{2}\right)\frac{\epsilon}{2a}+\left(1-a-\frac{\epsilon}{2}\right)\left(-\frac{\epsilon}{2(1-a)}\right)=\frac{\epsilon^2}{4a(1-a)};$$

$$\Gamma_a^- \sim \left(a-\frac{\epsilon}{2}\right)\left(-\frac{\epsilon}{2a}\right)+\left(1-a+\frac{\epsilon}{2}\right)\frac{\epsilon}{2(1-a)}=\frac{\epsilon^2}{4a(1-a)}.$$

Следовательно,

$$r(V,\epsilon)\sim \inf\left(\frac{\epsilon^2}{4a(1-a)}, 0\le a\le 1\right)=\epsilon^2.$$

В свою очередь, при $\theta \to 0$, в обозначениях

$$g(x)=\int\limits_{-\infty}^{x} uf(u)du, \quad L=\int\limits_{\mathbb{R}} f(t)|g(t)|dt$$

и в условиях регулярности $f(0)>0, f'(0)=0$ при $\theta\to 0$, имеем

$$H(x,\theta)-F(x)\sim 2\theta f(0)\int\limits_{-\infty}^{x} uf(u)du.$$

Поэтому

$$b(V,f,\theta)=\int\limits_{\mathbb{R}}|H(t)-F(t)|dF(t)\sim\int\limits_{\mathbb{R}}\left|2\theta f(0)\int\limits_{-\infty}^{t} uf(u)du\right| f(t)dt=$$

$$=2\theta f(0)\int\limits_{\mathbb{R}}\left|\int\limits_{-\infty}^{t} uf(u)du\right|\cdot f(t)dt=2\theta f(0)\int\limits_{\mathbb{R}} f(t)|g(t)|dt=2\theta f(0)L.$$

Тогда получаем, что

$$c_V(f,\theta)=2r(V,b(V,f,\theta))=8\theta^2 f^2(0)L^2.$$

Так как $c_V(f,\theta)\sim LI(V,f)\cdot\theta^2$ при $\theta\to 0$, то отсюда следует, что

$$LI(V,f)=8f^2(0)L^2.$$

Кроме того, благодаря [9] мы имеем асимптотику для информации Кульбака-Лейблера:

$$K(f,\theta)\sim 2f^2(0)\int\limits_{\mathbb{R}^1} x^2 f(x)dx\cdot\theta^2.$$

Зная локальные индексы и информацию Кульбака-Лейблера, можно найти локальную эффективность по Бахадуру. Рассчитаем значения локальных индексов и информацию Кульбака-Лейблера для различных распределений.

Начнём с нормального: $f_1(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$.

Здесь получается, что

$$LI(V, f_1) = \frac{8}{2\pi} \left(\int\limits_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left| \int\limits_{-\infty}^{t} u \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right| dt \right)^2 =$$

$$= \frac{8}{2\pi} \left(\int\limits_{\mathbb{R}} \frac{1}{2\pi} e^{-t^2} \right)^2 = \frac{8}{2\pi} \frac{1}{4\pi} = \frac{1}{\pi^2} \approx \mathbf{0.101}.$$

$$K(f_1, \theta) \sim 2 \frac{1}{2\pi} \int\limits_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \theta^2 = \frac{1}{\pi} \theta^2.$$

Логистическое распределение: $f_2(x) = \frac{e^x}{(1+e^x)^2}$

$$LI(V, f_2) = 8 \cdot \frac{1}{16} \cdot \left(\int\limits_{\mathbb{R}} \frac{e^t}{(1+e^t)^2} \left| \int\limits_{-\infty}^{t} u \frac{e^u}{(1+e^u)^2} du \right| dt \right)^2 = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} = \mathbf{0.125}.$$

$$K(f_2, \theta) \sim 2 \cdot \frac{1}{16} \int\limits_{\mathbb{R}} x^2 \frac{e^x}{(1+e^x)^2} dx \cdot \theta^2 = \frac{\pi^2}{24} \theta^2.$$

Распределение арксинуса: $f_3(x) = \frac{1}{\pi\sqrt{1-x^2}} \mathbf{1}(-1 < x < 1)$.

$$LI(V, f_3) = 8 \cdot \frac{1}{\pi^2} \left[\int\limits_{-1}^{1} \frac{1}{\pi\sqrt{1-t^2}} \left| \int\limits_{-1}^{t} u \frac{1}{\pi\sqrt{1-u^2}} du \right| dt \right]^2 = \frac{8}{\pi^2} \left[\int\limits_{-1}^{1} \frac{1}{\pi^2} dt \right]^2 = \frac{32}{\pi^6} \approx \mathbf{0.033}.$$

$$K(f_3, \theta) \sim 2 \cdot \frac{1}{\pi^2} \int\limits_{-1}^{1} x^2 \frac{1}{\pi\sqrt{1-x^2}} dx \cdot \theta^2 = \frac{1}{\pi^2} \theta^2.$$

Равномерное распределение: $f_4(x) = \frac{1}{2} \mathbf{1}(-1 < x < 1)$.

$$LI(V, f_4) = 8 \cdot \frac{1}{4} \cdot \left[\int\limits_{-1}^{1} \frac{1}{2} \left| \int\limits_{-1}^{t} \frac{u}{2} du \right| dt \right]^2 = 2 \cdot \left[\int\limits_{-1}^{1} \frac{1-t^2}{8} dt \right]^2 = \frac{1}{18} = \mathbf{0.056}.$$

$$K(f_4,\theta) \sim 2 \cdot \frac{1}{4} \int\limits_{-1}^{1} x^2 \frac{1}{2} dx \cdot \theta^2 = \frac{1}{6}\theta^2.$$

Далее, хотелось бы рассмотреть распределение Коши, но ввиду того, что информация Кульбака-Лейблера в данном случае для него равна бесконечности (поскольку распределение Коши не имеет дисперсии), рассматриваем его модификацию:

$$f_5(x) = \frac{8}{3\pi} \cdot \frac{1}{(1+x^2)^3}.$$

Здесь

$$LI(V,f_5) = 8 \cdot \frac{64}{9\pi^2} \cdot \left[\int\limits_{\mathbb{R}} \frac{8}{3\pi} \cdot \frac{1}{(1+t^2)^3} \left| \int\limits_{-\infty}^{t} u \frac{8}{3\pi} \cdot \frac{1}{(1+u^2)^3} du \right| dt \right]^2 =$$

$$= \frac{8^3}{9\pi^2} \left[\int\limits_{\mathbb{R}} \frac{16}{9\pi^2} \cdot \frac{1}{(1+t^2)^5} dt \right]^2 \approx \mathbf{0.138}.$$

Здесь

$$K(f_5,\theta) \sim 2 \cdot \frac{64}{9\pi^2} \int\limits_{\mathbb{R}} x^2 \frac{8}{3\pi} \cdot \frac{1}{(1+x^2)^3} dx \cdot \theta^2 = \frac{128}{27\pi^2}\theta^2.$$

Рассмотрим, кроме того, следующие статистики:

$$S_n = n^{-1} \sum_{i=1}^{n} \left[\mathbf{1}_{\{X_i>0\}} - \frac{1}{2} \right].$$

$$W_n = n^{-1} \sum_{i=1}^{n} a_n \left(\frac{R_i}{n+1} \right) \mathbf{1}(X_i > 0)$$

где $(R_1,..,R_n)$ - вектор рангов выборки $|X_1|,..,|X_n|$, а a_n - ф-я, определённая на $[0,1]$, постоянная на интервалах $(\frac{i-1}{n}, \frac{i}{n}], 1 \le i \le n$ и $a_n(u) \to \sqrt{12}u, n \to \infty$.

Для этих статистик локальные эффективности были просчитаны в [9]. Отсюда получаем таблицу локальных эффективностей по Бахадуру при скошенной альтернативе:

Статистика	Гаусс	Логистическое	Арксинус	Равномерное	f_5
V_n	0.318	0.304	0.329	0.333	0.287
S_n	0.637	0.584	0.812	0.750	0.540
W_n	0.955	0.912	0.986	1	0.862

Даже немногих взятых критериев хватает, чтобы понять, что эффективность в данном случае у V_n довольно мала и ее трудно рекомендовать для проверки симметрии для скошенных альтернатив.

5 Лемановская альтернатива

Функция распределения имеет вид $G(x, \theta) = F^{1+\theta}(x), \theta > 0$.

В данном случае аналогично $r(V, \epsilon) \sim \epsilon^2$ при $\epsilon \to 0$. Нужно найти b из теоремы Бахадура для получения выражения точного наклона. Лемановская альтернатива - чисто непараметрическая, потому ответ не должен зависеть от вида F. Имеем

$$F^{1+\theta} = \exp[(1+\theta)\ln F] \sim F + F \ln F \cdot \theta + O(\theta^2), \theta \to 0$$

.

Отсюда получаем выражение для $b(\theta)$:

$$b(V, f, \theta) \sim \theta \cdot \int\limits_{\mathbb{R}} -F(t) \ln F(t) dF(t) = -\theta \cdot \int_0^1 t \ln t dt.$$

Следовательно, локальный точный наклон по Бахадуру получается следующим:

$$c_V(\theta) = 2r(V, b(V, f, \theta)) = 2\theta^2 \left(\int_0^1 t \ln t dt \right)^2 = \theta^2 \cdot \frac{1}{8}.$$

Информация Кульбака-Лейблера имеет следующую асимптотику (как показано в [15]):

$$K(f, \theta) \sim \theta^2 \cdot \frac{\pi^2}{24}.$$

Значит, локальная эффективность критерия V_n при лемановской альтернативе составляет:

$$e_V^B(\theta) = \frac{\frac{1}{8}}{\frac{\pi^2}{12}} \approx \mathbf{0.152}.$$

Это снова маленькое значение, стало быть, статистика V_n мало пригодна и для этой альтернативы.

6 Альтернатива загрязнения

При этой альтернативе ф.р. имеет вид $G(x,\theta) = (1-\theta)F(x)+\theta F^{1+r}(x)$, $\theta > 0, r > 0$.

Имеем следующую асимптотику для информации Кульбака-Лейблера [15] при $\theta \to 0$:

$$K(\theta) \sim (r+1)^2 \left(\frac{1}{4(2r+1)} - \frac{\sqrt{\pi}\Gamma(1+r)}{2^{2r+3}\Gamma(\frac{3}{2}+r)} \right) \cdot \theta^2.$$

Найдём значение b из теоремы Бахадура:

$$b(V,f,\theta) = \int\limits_{\mathbb{R}} \theta \cdot \left| F^{1+r}(x) - F(x) \right| dF(x) = \theta \int\limits_0^1 \left(x - x^{1+r} \right) dx = \theta \cdot \frac{r}{2(r+2)}.$$

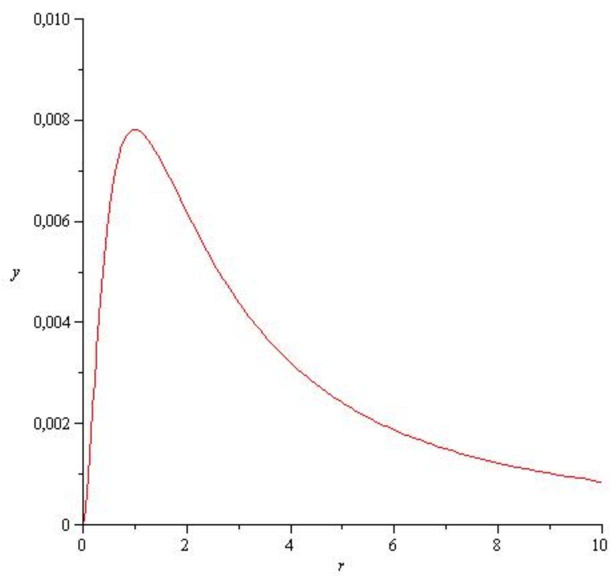
Следовательно, точный наклон имеет вид:

$$c_V(\theta) \sim \frac{r^2}{4(r+2)^2} \cdot \theta^2, \theta \to 0.$$

Тогда получаем, что локальная эффективность при альтернативе загрязнения имеет следующее выражение:

$$e_V^B = \frac{r^2}{8(r+2)^2(r+1)^2 \left(\frac{1}{4(2r+1)} - \frac{\sqrt{\pi}\Gamma(1+r)}{2^{2r+3}\Gamma(\frac{3}{2}+r)} \right)}.$$

График эффективности критерия V_n при альтернативе загрязнения при разных значениях r:



Как можно видеть, максимум эффективности достигается примерно при $r = 1$, и даже этот максимум составляет достаточно малое значение (около 0.008). Следовательно, в данном случае статистика V_n снова практически непригодна на практике.

7 Проверка однородности

7.1 Альтернатива сдвига

Уже после работы Луани и Берраху испанский статистик Мартинес-Камблор с соавторами [12] рассмотрел двухвыборочный вариант L_1-критерия Луани для проверки однородности. Общая постановка задачи: пусть $X = \{X_1, ..., X_{n_1}\}$ и $Y = \{Y_1, ..., Y_{n_2}\}$ - две независимые случайные выборки с непрерывными генеральными ф.р. F_1 и F_2 и функциями плотности f_1 и f_2 соответственно. Проверяется основная гипотеза однородности

$$H_0 : F_1 = F_2.$$

В качестве альтернативы предполагается, что $F_1(t) = R_1(t), F_2(t) = R_2(t)$, где R_1, R_2 - некоторые непрерывные ф.р. на $\mathbb{R}^1$, причём $R_1 \neq R_2$. В качестве альтернатив целесообразно рассматривать простые параметрические гипотезы.

Понимаем под Θ множество всех пар непрерывных ф.р. на $\mathbb{R}^1$, под Θ_0 - подмножество пар из Θ с совпадающими компонентами, $\Theta_1 = \Theta \setminus \Theta_0$.

Пусть f_{n_1} и f_{n_2} - соответствующие ядерные оценки плотности, построенные по первой и второй выборке. Для проверки гипотезы однородности введём статистику

$$V_n = \int |f_{n_1}(X, t) - f_{n_2}(Y, t)| dt.$$

Благодаря статье [12] у нас имеется следующая теорема о больших уклонениях V_n.

***Теорема**: пусть выполнены следующие условия:*

1) $\lim\limits_{n_1 \to \infty, n_2 \to \infty} \frac{n_1}{n_2} = 1$;

2) Функция ядра K является непрерывной симметричной функцией плотности;

3) $\lim\limits_{n_1 \to \infty} a_{n_1} = 0,\ \lim\limits_{n_2 \to \infty} a_{n_2} = 0,\ \lim\limits_{n_1 \to \infty} n_1 a_{n_1} = \infty,\ \lim\limits_{n_2 \to \infty} n_2 a_{n_2} = \infty.$

Тогда, если λ - неотрицательная величина, стремящаяся к нулю, выполнено

$$\lim_{n_1\to\infty, n_2\to\infty} \frac{n_1+n_2}{2n_1n_2} \ln \mathbb{P}(\int |f_{n_1}(X,t) - f_{n_2}(Y,t)|dt > \lambda) = -\frac{\lambda^2}{4}(1+o(1)).$$

Отсюда следует, что наклон по Бахадуру для V_n есть величина

$$\frac{1}{2}(\int |f_1 - f_2|)^2(1+o(1)).$$

Замечание. В этой теореме предполагается, что $\lim \frac{n_1}{n_2} = 1$, что ощутимо сужает общность результата и не позволяет сравнивать выборки существенно неравных объемов.

Если ввести альтернативу сдвига, $F_1(x) = F(x)$, $F_2(x) = F(x-\theta)$ для $\theta > 0$, где F обозначает непрерывную ф.р., f - соответствующая функция плотности, то точный наклон по Бахадуру будет иметь вид:

$$c_V(\theta) = \frac{1}{2}\theta^2 \left(\int |f'(x)|dx\right)^2.$$

Далее, нас интересует верхняя граница для наклона. Пользуясь [16], убеждаемся в том, что в случае альтернативы сдвига она имеет следующий вид (см. (3.3.7) в [16]):

$$c_V(\theta) \leq \ 2K^*(\theta) \sim \frac{1}{2}\theta^2 I(f),$$

где $I(f)$ есть информация Фишера.

Таким образом, локальная бахадуровская эффективность в случае альтернативы сдвига есть

$$e_V^B = \frac{\frac{1}{2}\theta^2(\int |f'(x)|dx)^2}{\frac{1}{2}\theta^2 I(f)} = \frac{(\int |f'(x)|dx)^2}{I(f)},$$

что абсолютно аналогично случаю с проверкой симметрии, следовательно, результаты переносятся сюда, и мы имеем локальную эффективность для распределения Лапласа и более широкого класса распределений, определяемых завершающей теоремой пункта 2.3.

7.2 Лемановская альтернатива

Если ввести лемановскую альтернативу, $F_1(x) = F(x)$, $F_2(x) = F^{1+\theta}(x)$ для $\theta > 0$, где F обозначает непрерывную ф.р., f и f_θ - соответствующие F и $F^{1+\theta}$ функции плотности, то, пользуясь собственными результатами из пункта 5 касательно значений из теоремы Бахадура ($b(\theta) = \frac{\theta}{4}$), получаем, что точный наклон по Бахадуру в этом случае имеет вид:

$$c_V(\theta) = \frac{\theta^2}{16} \cdot \frac{1}{4} = \frac{\theta^2}{64}.$$

Далее, чтобы найти верхнюю границу для наклона, воспользуемся [16]: в нашем случае

$$c_V(\theta) \leq K\left(f_1, \frac{1}{2}(f_1 + f_2)\right) + K\left(f_2, \frac{1}{2}(f_1 + f_2)\right),$$

где, соответственно,

$$K\left(f_1, \frac{1}{2}(f_1 + f_2)\right) = \int\limits_{-\infty}^{\infty} \ln\left[\frac{2f_\theta(x)}{f_\theta(x) + f(x)}\right] \cdot f_\theta(x)dx \sim \frac{\theta^2}{8}, \theta \to 0$$

$$K\left(f_2, \frac{1}{2}(f_1 + f_2)\right) = \int\limits_{-\infty}^{\infty} \ln\left[\frac{2f(x)}{f_\theta(x) + f(x)}\right] \cdot f(x)dx \sim \frac{\theta^2}{8}, \theta \to 0.$$

Т.о., получаем, что локальная эффективность критерия V_n при лемановской альтернативе составляет

$$e_V^B = \frac{\frac{\theta^2}{64}}{\frac{\theta^2}{4}} = \frac{1}{16} = \mathbf{0.0625}.$$

7.3 Скошенная альтернатива

Если ввести скошенную альтернативу, $f_1(x) = f(x)$, $f_2(x) = 2f(x)F(\theta x)$, то, пользуясь собственными результатами из пункта 4 касательно значений из теоремы Бахадура ($b(f, \theta) = 2\theta f(0) \int\limits_{\mathbb{R}} f(t)|g(t)|dt$), получаем, что точный наклон по Бахадуру в данном случае имеет вид:

$$c_V(\theta) = 2\theta^2 f^2(0) \left(\int\limits_{\mathbb{R}} f(t)|g(t)|dt\right)^2.$$

Информация Кульбака-Лейблера, необходимая нам для получения верхней границы, при отсутствии известной асимптотики достаточно сложна для вычисления даже самых простых распределений. Однако, в случае равномерного распределения, при помощи нескольких вычислительных средств одновременно, можно получить следующее:

$$K(f_1, \frac{1}{2}(f_1+f_2)) \sim \frac{\theta^2}{24}, \theta \to 0$$
$$K(f_2, \frac{1}{2}(f_1+f_2)) \sim \frac{\theta^2}{12}, \theta \to 0$$

Формула для точного наклона была приведена выше, и он равняется

$$c_V(\theta) = 2\theta^2 \frac{1}{4} \left[\int_{-1}^{1} \frac{1}{2} \left| \int_{-\infty}^{t} \frac{u}{2} du \right| dt \right]^2 = \frac{\theta^2}{2} \cdot \frac{1}{36} = \frac{\theta^2}{72}$$

Соответственно, локальная эффективность в данном случае составляет

$$e_V^B = \frac{\frac{\theta^2}{72}}{\frac{\theta^2}{8}} = \frac{1}{9} \approx \mathbf{0.111}.$$

Мы видим, что и здесь эффективность нашего критерия однородности невелика.

8 Заключение

Вычислив локальные бахадуровские эффективности последовательности статистик V_n для различных альтернатив и распределений, мы пришли к выводу, что данный критерий является весьма работоспособным и перспективным тестом не только для проверки симметрии, но и для проверки однородности, но *при альтернативе сдвига*, что должно позволить использовать его в прикладных целях. В остальных рассмотренных случаях, при других типах альтернатив, эффективность не была столь велика. Однако даже полученных результатов должно хватить, чтобы вызвать определённый интерес к дальнейшему изучению критерия симметрии, основанного на статистике V_n, ибо ранее, кроме работы Берраху-Луани [8], где он был введён, он не исследовался. То же относится и к аналогичному критерию однородности.

Список литературы

[1] Arbuthnot J.(1712). An argument for divine providence, taken from the constant regularity observ'd in the births of both sexes. By Dr. John Arbuthnott, Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society. Philos. Trans. Roy. Soc. of London. 27,N 328, 186-190.

[2] A. Azzalini (1985). A class of distributions which includes the normal ones. Scand. J. Statist. 12, 171–178.

[3] A. Azzalini with the collaboration of A. Capitanio(2013). The Skew-Normal and Related Families. Cambridge University Press, Cambridge.

[4] Baringhaus, L., Henze, N.(1992). A characterization of and new consistent tests for symmetry. Comm. Statist. Theory Methods, 6, 1555–1566.

[5] Bahadur R.R.(1967). Rates of convergence of estimates and test statistics. Ann.Math.Statist., 38, N 2, 303-324.

[6] Bahadur R.R.(1971). Some limit theorems in statistics. SIAM: Philadelphia.

[7] Louani, D.(2000). Large deviations for the L_1-distance in kernel density estimation. J. Statist. plan. Infer., 90, 177–182.

[8] Berrahou N., Louani D.(2006). Efficiency of some tests when testing symmetry hypothesis. Journal of Nonparametric Statistics, 18, N 7-8, 465-482.

[9] Durio A. , Nikitin Ya.Yu.(2002). Asympotic efficiency of signed - rank symmetry tests under skew alternatives. ICER Working Papers, 12-2002.

[10] Hill D.L., Rao P.V.(1977). Tests of symmetry based on Cramer–von Mises statistics. Biometrika, 64, N 3, 484-494.

[11] Ho N.V.(1973). Asymptotic efficiency in the Bahadur sense for the signed rank tests. Proc. Prague Sympos. on asympt. Statist, 12, 127–156.

[12] Martinez-Camblor P., Corral N., Lopez T.(2009). Cramer-Chernoff Theorem for L_1-norm in Kernel Density Estimator for Two Independent Samples. Revista Colombiana de Estadistica, 32, N 2, 289-299.

[13] Аббакумов В.Л.(1987). Асимптотическая эффективность непараметрических критериев симметрии. Дис. канд. физ.-мат. наук.Л.

[14] Кендалл, М., Стьюарт, А.(1973). Статистические выводы и связи. Наука. Гл. ред. физ.-мат. лит., М.

[15] Литвинова В.В.(2004). Асимптотические свойства критериев симметрии и согласия, основанных на характеризациях. Дис. канд. физ.-мат. наук.СПб.

[16] Никитин Я.Ю.(1995). Асимптотическая эффективность непараметрических критериев. М., Наука.

[17] Смирнов Н.В.(1947). О критерии симметрии закона распределения случайной величины. ДАН СССР, 56, №1, 13-16.

[18] Ченцов Н.Н.(1958). Обоснование статистических критериев методами теории случайных процессов. Дис. канд. физ.-мат. наук.М.