На правах рукописи

ВОЛКОВА КСЕНИЯ ЮРЬЕВНА

АСИМПТОТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ КРИТЕРИЕВ СОГЛАСИЯ, ОСНОВАННЫХ НА ХАРАКТЕРИЗАЦИОННЫХ СВОЙСТВАХ РАСПРЕДЕЛЕНИЙ

01.01.05 — теория вероятностей и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата физико-математических наук

Санкт-Петербург — 2011

Работа выполнена на кафедре теории вероятностей и математической статистики математико-механического факультета Санкт-Петербургского государственного университета.

Научный руководитель	доктор физико-математических наук, профессор **Никитин Яков Юрьевич**
Официальные оппоненты	доктор физико-математических наук, **Ингстер Юрий Измайлович**
	кандидат физико-математических наук, доцент **Подкорытова Ольга Анатольевна**
Ведущая организация	Московский государственный университет имени М. В. Ломоносова

Защита состоится "____" ____________ 2012 года в ____ часов на заседании диссертационного совета Д 002.202.01 в Санкт-Петербургском отделении Математического института им. В. А. Стеклова РАН по адресу 191023, Санкт-Петербург, наб. р. Фонтанки, д. 27, к. 311.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского отделения Математического института им. В. А. Стеклова РАН по адресу 191023, Санкт-Петербург, наб. р. Фонтанки, д. 27.

Автореферат разослан "____" ____________ 2012 года.

Ученый секретарь
диссертационного совета Д 002.202.01
доктор физико-математических наук А. Ю. Зайцев

Общая характеристика работы

Актуальность темы. Математическая теория проверки гипотез была создана Ю. Нейманом и Э. Пирсоном около 80 лет назад, и с тех пор построение статистических критериев и изучение их асимптотических свойств является одним из важнейших разделов математической статистики. В настоящий момент существует огромное множество критериев для проверки статистических гипотез.

Большое многообразие критериев обусловлено тем, что для проверки сложных гипотез даже в классических параметрических моделях лишь в редких случаях удается построить критерий, который является наилучшим, оптимальным (равномерно наиболее мощным) критерием. Поэтому многие тесты предлагаются их авторами из эвристических соображений, примерами могут служить знаменитые критерии хи-квадрат, Колмогорова и омега-квадрат.

При проверке непараметрических гипотез число возможных критериев увеличивается еще больше. Поэтому возникает интерес к выработке единой меры сравнения критериев, чтобы систематизировать их и дать количественную оценку их чувствительности для выработки рекомендаций по их использованию на практике. Такой мерой в последние десятилетия служит их асимптотическая относительная эффективность (АОЭ).

В начале 50-х годов XX века Ю.В. Линник предложил общий принцип построения новых критериев согласия на основе характеризационных теорем для семейств вероятностных распределений, связанных со свойством равнораспределенности статистик. Таких теорем существует сравнительно немного, и приспособить их для проверки гипотез нелегко, ввиду многочисленных трудностей технического характера. Возможно, поэтому критериев согласия «характеризационного» типа разработано очень мало, а литература в этой области весьма невелика. Здесь можно упомянуть работы Л. Барингхауза, В.В. Литвиновой, С. Мейнтаниса, П. Мульере, Я.Ю. Никитина и Н. Хенце, опубликованные в последние 15 лет.

Однако этот подход далеко не исчерпал свои возможности и представляется актуальным и перспективным для исследований, что демонстрируется в нашей работе. В ней строятся новые критерии согласия, основанные на характеризациях распределений свойством равнораспределенности статистик, изучаются их асимптотические свойства, а также вычисляется их бахадуровская АОЭ для ряда альтернатив.

Цель работы. Диссертация посвящена построению и исследованию ряда новых критериев согласия, основанных на характеризациях распределений, для проверки экспоненциальности, нормальности, для проверки согласия со степенным законом и законом арксинуса, а также вычислению их асимптотической относительной эффективности.

Методы исследований. В диссертационной работе используется идея построения критериев на основе характеризационных свойств распределений. Критериальные статистики изучаются методами теории эмпирических процессов и теории U- и V-статистик, используются предельные теоремы для U- и V-статистик, теория больших уклонений для U- и V-статистик и функционалов от U-эмпирических распределений, а также теория бахадуровской асимптотической эффективности.

Основные результаты.

1. Построены основанные на характеризациях новые критерии согласия для таких семейств вероятностных распределений как экспоненциальное, нормальное, степенное распределения, а также распределение арксинуса.
2. Описаны предельные распределения и вероятности больших уклонений построенных статистик при основной гипотезе.
3. Вычислена локальная бахадуровская эффективность новых критериев для ряда альтернатив, изучены условия их локальной асимптотической оптимальности и даны рекомендации по использованию на практике.

Научная новизна. Все основные результаты диссертации являются новыми. В ней построено 14 новых критериев согласия с нормальным, экспоненциальным, степенным законами и законом арксинуса, изучены их асимптотические свойства. Также впервые получены значения локальной бахадуровской эффективности новых критериев для ряда естественных альтернатив и изучены условия локальной асимптотической оптимальности рассматриваемых критериев.

Теоретическая и практическая ценность. Диссертация носит в основном теоретический характер. В ней разработаны и исследованы новые статистические критерии. Эффективности новых статистик для проверки экспоненциальности, нормальности, согласия со степенным законом и законом арксинуса превышают эффективности многих известных статистик и поэтому могут составить конкуренцию при разнообразных применениях проверки статистических гипотез в прикладных областях. Полученные результаты могут использоваться статистиками-практиками для обоснованного выбора статистического критерия в прикладных задачах проверки гипотез.

Апробация работы. Результаты диссертации докладывались на Шестнадцатой Всероссийской школе-коллоквиуме по стохастическим методам (Санкт-Петербург, 19–24 мая 2009 г.), на Шестом международном симпозиуме по моделированию (Санкт-Петербург, 28 июня–4 июля 2009 г.), на Втором международном симпозиуме по вычислениям и статистике ERCIM Working Group (Лимассол, Кипр, 29–31 октября 2009 г.), на Втором Северном трехстороннем (финско-шведско-российском) семинаре (Стокгольм, 15–17 марта 2010 г.), на Десятой международной вильнюсской конференции по теории вероятностей и математической статистике (Вильнюс, 28 июня–2 июля 2010 г.), на Втором международном симпозиуме по обратным задачам, анализу данных, математической статистике и экологии (Хельсинки, 25–27 августа 2010 г.), на санкт-петербургском городском семинаре по теории вероятностей и математической статистике под руководством академика РАН И.А. Ибрагимова (в сентябре 2011 г.) и на международной конференции "Analytical Methods in Statistics" (Amistat 2011, Прага, 28–30 октября 2011 г.)

Публикации. Основные результаты диссертации опубликованы в работах [П1]–[П9]. Из них четыре работы [П1]–[П4] опубликованы в журналах, рекомендованных ВАК (работа [П1] опубликована в журнале, удовлетворяющем достаточному условию включения в перечень ВАК: его переводная версия “Journal of Mathematical Sciences” входит в систему цитирования SCOPUS). Работы [П5]–[П9] – это тезисы совместных докладов на международных конференциях на общую тему, где представлены как результаты автора, так и его научного руководителя.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав, заключения и списка литературы, содержащего 91 наименование. Общий объем работы составляет 144 страницы.

Содержание работы

В **первой главе** вводятся необходимые вспомогательные сведения из теории U- и V-статистик, основные определения и теоремы теории больших уклонений и теории Бахадура.

В качестве метода вычисления АОЭ критериев нами выбран подход Бахадура по следующим соображениям [3]:

- статистика колмогоровского типа D_n не является асимптотически нормальной, поэтому подход Питмена к ней не применим. В случае асимптотически нормальных интегральных статистик типа I_n, локальная АОЭ по Бахадуру и питменовская эффективность совпадают.

- АОЭ по Бахадуру различает статистики, имеющие одинаковую питменовскую эффективность.

- при использовании эффективности по Ходжесу–Леману, все двусторонние критерии оказываются асимптотически оптимальными, а в случае односторонних критериев АОЭ по Ходжесу–Леману локально совпадает с АОЭ по Бахадуру. Тем самым АОЭ по Ходжесу–Леману оказывается менее удобным и содержательным средством сравнения критериев, чем бахадуровская АОЭ.

- развитие теории больших уклонений, в частности работы [4] и [15] позволяют преодолеть трудности по вычислению асимптотики больших уклонений U-статистик и функционалов от U-эмпирических распределений.

Вторая глава посвящена построению и анализу критериев для проверки экспоненциальности. Общая постановка задачи проверки экспоненциальности выглядит так: пусть $X_1, \ldots, X_n$ – 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 $f(x) = \lambda e^{-\lambda x}, x \geq 0$, где $\lambda > 0$ – некоторый неизвестный параметр.

В качестве альтернатив для рассматриваемых характеризаций мы используем стандартные альтернативы к гипотезе экспоненциальности:

i) альтернативу Вейбулла с плотностью

$$(1+\theta)x^{\theta}\exp(-x^{1+\theta}),\ \theta \geq 0,\ x \geq 0;$$

ii) альтернативу Макегама с плотностью

$$(1+\theta(1-e^{-x}))\exp(-x-\theta(e^{-x}-1+x)),\ \theta \geq 0,\ x \geq 0;$$

iii) альтернативу линейности функции интенсивности отказов с плотностью

$$(1+\theta x)e^{-x-\frac{1}{2}\theta x^2},\ \theta \geq 0,\ x \geq 0;$$

iv) гамма-альтернативу с плотностью

$$\frac{x^{\theta}}{\Gamma(\theta+1)}e^{-x},\ \theta \geq 0,\ x \geq 0.$$

Мы рассмотрим критерии экспоненциальности, основанные на следующих характеризациях:

1. Характеризация Россберга [17, 18].

Будем обозначать через $X_{s,n}$ s-ю порядковую статистику, $1 \leq s \leq n$, в выборке $X_1, \dots, X_n$.

Пусть $X_1, \dots, X_n$ – неотрицательные н.о.р.с.в. Если для некоторого j статистики $X_{j+s,n} - X_{j,n}$ и $X_{s,n-j}$ одинаково распределены, то выборка имеет экспоненциальный закон распределения.

Рассмотрим частный случай этой характеризации при $n = 3, s = 1$ и $j = 1$. Тогда рассматриваемая характеризация примет вид: *если статистики $X_{2,3} - X_{1,3}$ и $\min(X_1, X_2)$ одинаково распределены, то выборка имеет экспоненциальный закон распределения.*

В соответствии с этой характеризацией строятся две U-эмпирические ф.р.

$$H_n^R(t) = \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} \mathbf{1}\{X_{2,\{i,j,k\}} - X_{1,\{i,j,k\}} < t\}, \quad t \geq 0,$$

$$G_n^R(t) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{1}\{\min(X_i, X_j) < t\}, \quad t \geq 0,$$

здесь и далее под выражением $X_{s,\{i,j,k\}}$, $s = 1, 2$, понимается s-я порядковая статистика выборки X_i, X_j, X_k.

На основе этих функций строятся инвариантные к параметру масштаба λ статистики:

$$I_n^R = \int_0^\infty \left(H_n^R(t) - G_n^R(t)\right) dF_n(t),$$
$$D_n^R = \sup_{t \geq 0} \mid H_n^R(t) - G_n^R(t) \mid .$$

Статистика I_n^R интегрального типа, простейшие статистики такого вида уже изучались в [2]. Доказывается, что она имеет в пределе нормальное распределение, а именно: при $n \to \infty$

$$\sqrt{n} I_n^R \xrightarrow{d} \mathcal{N}\left(0, \frac{52}{1125}\right).$$

Вторая статистика принадлежит к колмогоровскому типу. Явное предельное распределение статистики D_n^R неизвестно. Опираясь на методы работы [22], можно показать, что U-эмпирический процесс

$$\eta_n(t) = \sqrt{n}\left(H_n^R(t) - G_n^R(t)\right),\ t \geq 0,$$

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

Далее для статистик I_n^R и D_n^R изучаются вероятности их больших уклонений при нулевой гипотезе, вычисляется локальная бахадуровская АОЭ. АОЭ для статистик D_n^R исследуется впервые, поскольку метод изучения их больших уклонений был разработан лишь в работе [15] 2010 г., впрочем, асимптотическая эффективность более простых статистик колмогоровского типа уже изучалась в литературе, см. [9], [13], [3].

Следует заметить, что статистики интегрального типа, как правило, имеют более высокую АОЭ по сравнению со статистиками колмогоровского типа, зато последние являются состоятельными против любых альтернатив.

2. Характеризация Ахсануллаха.

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 $t \geq 0$.

Ахсануллах доказал в [5] следующую характеризацию свойствами порядковых статистик в пределах 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 $1 \leq i < j < k$, то выборка имеет экспоненциальный закон распределения.

Задачи 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.

Мы рассмотрим два частных случая характеризации:

а) Случай $k = 2$, $i = 0$ 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 $\mathcal{F}$.

Строятся две V-статистические ф.р.

$$H_n^{Ah1}(t) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{1}\{|X_i - X_j| < t\}, \quad t \geq 0,$$

$$G_n^{Ah1}(t) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{1}\{2\min(X_i, X_j) < t\}, \quad t \geq 0.$$

Критерии для проверки гипотезы экспоненциальности основываются на инвариантных к параметру масштаба λ статистиках:

$$I_n^{Ah1} = \int_0^\infty \left(H_n^{Ah}(t) - G_n^{Ah}(t)\right) dF_n(t),$$

$$D_n^{Ah1} = \sup_{t \geq 0} \mid H_n^{Ah}(t) - G_n^{Ah}(t) \mid .$$

Предельное распределение статистики I_n^{Ah1} нормально, а именно при $n \to \infty$

$$\sqrt{n} I_n^{Ah1} \xrightarrow{d} \mathcal{N}\left(0, \frac{647}{4725}\right),$$

а распределение статистики D_n^{Ah1} неизвестно.

б) При $k = 3$, $i = 1$ и $j = 2$ характеризация формулируется так:

Пусть X_1, X_2 и X_3 неотрицательные н.о.р.с.в. с ф.р. из класса $\mathcal{F}$. Если статистики $X_{3,3} - X_{2,3}$ и $2(X_{2,3} - X_{1,3})$ одинаково распределены, то выборка имеет экспоненциальный закон распределения.

Снова в соответствии с характеризацией построим две V-статистические ф.р.

$$H_n^{Ah2}(t) = \frac{1}{n^3} \sum_{i,j,k=1}^{n} \mathbf{1}\{X_{3,\{i,j,k\}} - X_{2,\{i,j,k\}} < t\}, \quad t \geq 0,$$

$$G_n^{Ah2}(t) = \frac{1}{n^3} \sum_{i,j,k=1}^{n} \mathbf{1}\{2(X_{2,\{i,j,k\}} - X_{1,\{i,j,k\}}) < t\}, \quad t \geq 0.$$

Мы предлагаем основывать критерии против альтернатив из класса $\mathcal{F}$ на следующих инвариантных к параметру масштаба λ статистиках:

$$I_n^{Ah2} = \int_0^\infty \left(H_n^{Ah2}(t) - G_n^{Ah2}(t)\right) dF_n(t),$$

$$D_n^{Ah2} = \sup_{t \geq 0} \mid H_n^{Ah2}(t) - G_n^{Ah2}(t) \mid .$$

Доказывается, что статистика I_n^{Ah1} имеет в пределе нормальное распределение, а именно при $n \to \infty$

$$\sqrt{n} I_n^{Ah2} \xrightarrow{d} \mathcal{N}\left(0, \frac{38}{525}\right),$$

а распределение статистики D_n^{Ah2}, как было описано ранее, неизвестно. Как уже отмечалось выше, мы предлагаем искать критические значения статистик D_n^{Ah1} и D_n^{Ah2} методом статистического моделирования.

3. Характеризация Шимидзу.

Одна из наиболее известных характеризаций экспоненциального распределения принадлежит Дезу [10]: *Пусть X и Y – неотрицательные н.о.р.с.в. с ф.р. F. Тогда X и $2\min(X, Y)$ одинаково распределены тогда и* *только тогда, когда выборка имеет экспоненциальный закон распределения.* Критерии экспоненциальности, основанные на характеризации Дезу, изучались в [2] и [15].

Мы рассмотрим критерии для проверки экспоненциальности, основанные на обобщении характеризации Дезу. В литературе нет устоявшегося названия данной характеризации, однако в более общем виде она была сформулирована Шимидзу [21], см. также [11], [1], [19]:

Пусть $X_1, \ldots, X_n$ *- неотрицательные н.о.р.с.в. и пусть натуральные* k *и* m *таковы, что число* $\frac{\log k}{\log m}$ *иррационально. Тогда статистики* $k \min(X_1, \ldots, X_k)$ *и* $m \min(X_1, \ldots, X_m)$ *одинаково распределены тогда и только тогда, когда выборка имеет экспоненциальный закон распределения.*

В соответствии с этой характеризацией строятся для любого натурального $l \geq 1$ V-эмпирическая ф.р. $G_l(t)$ и U-эмпирическая ф.р. $H_l(t)$.

$$G_l^{Sz}(t) = n^{-l} \sum_{i_1,\ldots,i_l=1}^{n} \mathbf{1}\{l \min(X_{i_1}, \ldots, X_{i_l}) < t\}, \quad t \geq 0,$$

$$H_l^{Sz}(t) = \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} \mathbf{1}\{l \min(X_{i_1}, \ldots, X_{i_l}) < t\}, \quad t \geq 0.$$

Критерии для проверки гипотезы экспоненциальности могут быть основаны на инвариантных к параметру масштаба λ статистиках:

$$I_{k,m}^{Sz} = \int_0^\infty \left(G_k^{Sz}(t) - G_m^{Sz}(t)\right) dF_n(t),$$

$$D_{k,m}^{Sz} = \sup_{t \geq 0} \mid H_k^{Sz}(t) - H_m^{Sz}(t) \mid .$$

Построенные статистики являются обобщением статистик, рассмотренных в [2] и [15].

Доказывается, что для статистики $I_{k,m}^{Sz}$ выполнена следующая сходимость по распределению при $n \to \infty$:

$$\sqrt{n} I_{k,m}^{Sz} \xrightarrow{d} \mathcal{N}(0, (m+1)^2 \Delta_{Sz}^2(k,m)),$$

где

$$\Delta_{Sz}^2(k,m) = \frac{(m-k)^2(8mk - 2k - 2m + 1)}{4(2k+2m-1)(4m-1)(4k-1)(m+1)^2}.$$

Как и прежде, предельное распределение статистики $D_{k,m}^{Sz}$ неизвестно, но в принципе может быть найдено на основе методов работы [22].

Для всех построенных во второй главе статистик найдены вероятности больших уклонений при нулевой гипотезе, вычислена их локальная бахадуровская эффективность при указанных альтернативах.

Результаты вычислений эффективности предложенных нами критериев приведены в таблице 1. Показывается, что для статистики $I_{k,m}^{Sz}$ значение локальной бахадуровской эффективности для стандартных альтернатив будет наилучшим для случая, рассмотренного Дезу ($k = 1$ и $m = 2$), так что использование других k и m вряд ли оправдано. Для статистик $D_{k,m}^{Sz}$ мы можем улучшать значения эффективности путем выбора подходящих k и m.

Таблица 1: Значения локальной АОЭ по Бахадуру для различных статистик

Статистика	Плотность Вейбулла	Плотность Макегама	Плотность Лин. инт. отк.	Гамма Плотность
Статистики интегрального типа				
I_n^R	0.650	0.450	0.119	0.757
I_n^{Ah1}	0.795	0.692	0.257	0.819
I_n^{Ah2}	0.445	0.737	0.486	0.277
$I_{1,2}^{Sz}$	0.697	0.509	0.149	0.790
Статистики колмогоровского типа				
D_n^R	0.320	0.207	0.047	0.414
D_n^{Ah1}	0.450	0.470	0.187	0.482
D_n^{Ah2}	0.280	0.513	0.362	0.174
$\max_{k,m} D_{k,m}^{Sz}$	0.261	0.218	0.073	0.366

Кроме того, изучены условия локальной асимптотической оптимальности построенных статистик, то есть описаны семейства альтернатив, для которых изучаемая статистика является локально наилучшей в бахадуровском смысле.

В **третьей главе** рассматриваются критерии для проверки нормальности, основанные на характеризации свойством Шеппа. В 1964 г. Шепп показал [20], что если X и Y – две независимые нормальные случайные величины со средним нуль и некоторой дисперсией $\tau^2 > 0$, то и случайная величина $2XY/\sqrt{X^2+Y^2}$ имеет снова распределение $\mathcal{N}(0, \tau^2)$. Это утверждение обычно называют *свойством Шеппа.* В 2003 г. было доказано [12], что с указанным свойством связана характеризация нормального закона в широком классе распределений.

Рассмотрим ф.р. F из класса $\mathfrak{F}$, определяемого условиями $0 < F(0) < 1$ и тем, что $F(x) - F(-x)$ регулярно меняется в нуле с показателем 1. В [12] доказано следующее утверждение. *Пусть X и Y независимые случайные величины с общей ф.р. F из класса $\mathfrak{F}$. Тогда равенство по распределению*

$$2XY/\sqrt{X^2+Y^2} \stackrel{d}{=} X$$

имеет место в том и только том случае, когда $X \in \mathcal{N}(0, \tau^2)$ для некоторой дисперсии $\tau^2 > 0$. Эта характеризация позволяет нам построить новые критерии нормальности, основанные на свойстве Шеппа.

Пусть $X_1, \ldots, X_n$ – независимые наблюдения с нулевым средним и ф.р. F, а F_n – обычная эмпирическая ф.р., построенная по этим наблюдениям. Нас интересует проверка сложной гипотезы $H_0 : F \in \mathcal{N}(0, \tau^2)$ при некоторой неизвестной дисперсии τ^2 против альтернативы H_1, при которой $F \in \mathfrak{F}$, но гипотеза H_0 неверна. Обозначим для краткости $k(x, y) = 2xy/\sqrt{x^2+y^2}$ и построим V-эмпирическую ф.р. G_n^{Sh} и U-эмпирическую ф.р. H_n^{Sh} с помощью формул

$$H_n^{Sh}(t) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbf{1}\{k(X_i, X_j) < t\}, \quad t \in R^1,$$

$$G_n^{Sh}(t) = \frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{1}\{k(X_i, X_j) < t\}, \quad t \in R^1.$$

Строятся новые инвариантные к параметру масштаба τ статистики для проверки гипотезы нормальности:

$$I_n^{Sh} = \int_{R^1} (G_n^{Sh}(t) - F_n(t)) dF_n(t),$$

$$D_n^{Sh} = \sup_{t \in R^1} \mid H_n^{Sh}(t) - F_n(t) \mid .$$

Для статистики I_n^{Sh} предельное распределение при $n \to \infty$ нормально

$$\sqrt{n} I_n^{Sh} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{12}\right).$$

Предельное распределение статистики D_n^{Sh} неизвестно, но критические точки могут быть найдены с помощью моделирования.

Затем описываются вероятности больших уклонений рассматриваемых статистик при нулевой гипотезе, вычисляется локальная бахадуровская АОЭ для следующих параметрических альтернатив при $\theta \geq 0$ и $x \in R^1$:

– альтернативы сдвига с ф.р. $F(x-\theta)$;

– скошенной альтернативы Аззалини [8] с плотностью $2f(x)F(\theta x)$;

– альтернативы загрязнения с ф.р. $(1-\theta)F(x) + \theta F^2(x)$.

Результаты вычислений приведены в таблице 2. Оказывается, что эффективности обеих статистик довольно высоки.

Таблица 2: Значения локальной АОЭ по Бахадуру для статистик I_n^{Sh} и D_n^{Sh}

Статистика	Альтернатива сдвига	Альтернатива Аззалини	Альтернатива загрязнения
I_n^{Sh}	0.955	0.955	0.417
D_n^{Sh}	0.637	0.637	0.313

В **четвертой главе** строятся критерии согласия для проверки сложной гипотезы $H_0 : F$ – ф.р. степенного закона с плотностью $f(x) = \mu x^{\mu-1}, x \in (0,1]$ с неизвестным показателем $\mu > 0$ против альтернативы, состоящей в том, что F – ф.р. не степенного закона. В отличие от экспоненциального и нормального семейств распределений, где существует множество критериев согласия, критерии согласия для проверки принадлежности степенному семейству весьма немногочисленны, единственным исключением является работа [14], причем критерии согласия для степенного закона, основанные на характеризационных свойствах, в литературе неизвестны.

Рассмотрим критерий согласия со степенным законом, основанный на следующей характеризации Пури и Рубина [16]: *Пусть X и Y – независимые неотрицательные случайные величины с ф.р. F. Тогда равенство по* *распределению X и $\min(\frac{X}{Y}, \frac{Y}{X})$ имеет место в том и только том случае, когда X имеет степенное распределение с ф.р. $F(x) = x^\mu, x \in (0,1], \mu > 0$.*

Строится U-эмпирическая ф.р.

$$H_n^{PR}(t) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbf{1}\left\{\min\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) < t\right\}, \quad t \in (0,1],$$

и новые статистики для проверки гипотезы согласия со степенным законом:

$$I_n^{PR} = \int_0^1 (H_n(t) - F_n(t))\, dF_n(t),$$
$$D_n^{PR} = \sup_{t \in (0,1]} \mid H_n(t) - F_n(t) \mid .$$

В качестве альтернатив рассматриваются следующие параметрические альтернативы для $x \in (0,1]$:

- альтернативу загрязнения с ф.р. $(1-\theta)F(x) + \theta F^r(x)$, $\theta \ge 0,\ r > 1$;
- вторую альтернативу с ф.р. $F(x) - \theta \sin(\pi F(x)) e^{F(x)}$, $-\frac{1}{e\pi} < \theta < \frac{1}{e\sqrt{\pi^2+1}}$;
- третью альтернативу с ф.р. $\frac{1}{1+\theta/2}(F(x) + \frac{\theta}{2}F^2(x))$, $\theta \ge 0$.

Для построенных статистик мы описываем предельное распределение и вероятности их больших уклонений при нулевой гипотезе, вычисляем локальную бахадуровскую АОЭ и изучаем условия их локальной асимптотической оптимальности. Оказывается, что эффективности обеих статистик довольно высоки.

Таблица 3: Значения локальной АОЭ по Бахадуру для статистик I_n^{PR} и D_n^{PR}

Статистика	Альтернатива загрязнения	Альтернатива вторая	Альтернатива третья
I_n^{PR}	0.968	0.968	0.800
D_n^{PR}	0.635	0.649	0.473

В заключительной **пятой главе** мы строим критерии согласия для закона арксинуса. Нас интересует проверка простой гипотезы $H_0 : F$ – ф.р.

закона арксинуса с плотностью $f(x) = \frac{1}{\pi\sqrt{1-x^2}}, -1 \leq x \leq 1$, против общей альтернативы, состоящей в том, что F – ф.р. имеет иной закон распределения.

Проверяемая нами гипотеза – простая, и потому может проверяться с помощью других критериев, например, с помощью критерия хи-квадрат или критерия Колмогорова. Однако нам кажется интересным исследовать возможности U-эмпирических критериев, основанных на характеризациях, применительно к этой задаче, поскольку в литературе отсутствуют какие-либо критерии согласия, основанные на характеризациях арксинуса (к тому же, такие характеризации вообще крайне немногочисленны).

Рассмотрим критерий согласия с законом арксинуса, основанный на следующей характеризации Арнольда–Гроневельда [7]: *Пусть X – симметричная случайная величина с ф.р. F. Тогда равенство по распределению X^2 и $\frac{1+X}{2}$ имеет место в том и только том случае, когда X имеет распределение арксинуса с ф.р. $F(x) = \frac{1}{2} + \frac{\arcsin x}{\pi}$, $x \in [-1, 1]$.*

В соответствии с характеризацией построим две эмпирические ф.р.

$$H_n^{AG}(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i^2 < t\}, \quad t \in [0, 1],$$

$$G_n^{AG}(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}\{\frac{1+X_i}{2} < t\}, \quad t \in [0, 1].$$

Мы предлагаем основывать критерии для проверки гипотезы согласия с законом арксинуса на следующих статистиках:

$$I_n^{AG} = \int_0^1 (H_n(t) - G_n(t))\, dF_n(t), \quad D_n^{AG} = \sup_{t \in [0,1]} \mid H_n(t) - G_n(t) \mid .$$

В качестве альтернатив рассмотрим скошенную альтернативу, альтернативу загрязнения, а также альтернативу Лемана с ф.р. $F^{1+\theta}(x)$, $\theta \geq 0$, $x \in [-1, 1]$.

Для рассмотренных статистик снова описывается предельное распределение и изучаются вероятности их больших уклонений при справедливости нулевой гипотезе, вычисляется локальная бахадуровская АОЭ при указанных альтернативах и исследуются условия их локальной асимптотической оптимальности.

Представление об асимптотической эффективности критериев согласия, основанных на предлагаемых статистиках, дает следующая таблица.

Таблица 4: Значения локальной АОЭ по Бахадуру для статистик I_n^{AG} и D_n^{AG}

Статистика	Альтернатива Аззалини	Альтернатива Лемана	Альтернатива загрязнения
I_n^{AG}	0.575	0.776	0.578
D_n^{AG}	0.304	0.347	0.281

В **заключении** подведены итоги проведенных исследований, обсуждаются достоинства и недостатки построенных критериев, а также их асимптотические эффективности, даются рекомендации по использованию предлагаемых критериев согласия на практике.

Список литературы

[1] Азларов Т.А., Володин Н.А. Характеризационные задачи, связанные с экспоненциальным распределением. — Ташкент: Фан, 1982.

[2] Литвинова В.В. Асимптотические свойства критериев симметрии и согласия, основанных на характеризациях. — Кандидатская диссертация. СПбГУ. 2004.

[3] Никитин Я.Ю. Асимптотическая эффективность непараметрических статистических критериев. — М.: Наука, 1995.

[4] Никитин Я.Ю., Поникаров Е.В. Грубая асимптотика вероятностей больших уклонений черновского типа для функционалов Мизеса и U-статистик. // Труды Санкт-Петербургского математического общества. 1999. Т. 7. С. 23–47.

[5] Ahsanullah M. On a characterization of the exponential distribution by spacings. // Ann. Inst. Statist. Math. 1978. A30. P. 163–166.

[6] Arnold B.C. Two characterizations of the exponential distribution using order statistics. // Preprint. Iowa State University. Ames. 1971.

[7] Arnold B.C., Groeneveld R.A. Some properties of the arcsine distribution. // J. Amer. Statist. Association. 1980. V. 75, № 369. P. 173–175 .

[8] Azzalini A. A class of distributions which includes the normal ones. // Scand. J. Statist. 1985. V. 12. P. 171–178.

[9] Bahadur R.R. Some limit theorems in statistics. — Philadelphia: SIAM, 1971.

[10] Desu M.M. A characterization of the exponential distribution by order statistics. // Ann. Math. Statist. 1971. V. 42, № 2. P. 837–838.

[11] Galambos J., Kotz S. Characterizations of probability distributions. // Lecture Notes in Math. New York: Springer. 1978. V. 675.

[12] Galambos J., Simonelli I. Comments on a recent limit theorem of Quine. // Statist. Probab. Lett. 2003. V. 63. P. 89–95.

[13] Groeneboom P., Oosterhoff J. Bahadur efficiency and probabilities of large deviations. // Statist. Neerlandica. 1977. V. 31, № 1. P. 1–24.

[14] Martynov G. Cramér–von Mises test for the Weibull and Pareto distributions. // In: Proceedings of Dobrushin International Conference, July 15–20. 2009. Moscow. P. 117–122.

[15] Nikitin Ya.Yu. Large deviations of U-empirical Kolmogorov–Smirnov tests, and their efficiency. // J. Nonparam. Statist. 2010. V. 22. P. 649–668.

[16] Puri P.S., Rubin H.A. A characterization based on the absolute difference of two i.i.d. random variables. // Ann. Math. Statist. 1970. V. 41. P. 2113–2122.

[17] Riedel M. , Rossberg H.J. Characterization of the exponential distribution function by properties of the difference $X_{k+s:n} - X_{k:n}$ of order statistics. // Metrika. 1994. V. 41. P. 1–19.

[18] Rossberg H.J. Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which differences and quotients of order statistics are subjected to. // Math. Operationsforsch Statist. 1972. V. 3. P. 207–216.

[19] Sethuraman J. On a characterization of three limiting types of extremes. // Sankhya. 1965. A27. P. 357–364.

[20] Shepp L. Normal functions of normal random variables. // SIAM Rev. 1964. V. 6. P. 459–460.

[21] Shimizu R. A characterization of the exponential distribution. // Ann. Inst. Statist. Math. 1979. V. 31, № 3. P. 367–372.

[22] Silverman B.W. Convergence of a class of empirical distribution functions of dependent random variables. // Ann. Probab. 1983. V. 11. P. 745–751.

Публикации автора по теме диссертации

Статьи в журналах, рекомендованных ВАК:

[П1] Волкова К.Ю. Об асимптотической эффективности критериев экспоненциальности, основанных на характеризации Россберга. // Зап. научн. семин. ПОМИ. 2009. Т. 368. С. 95–109.

[П2] Волкова К.Ю., Никитин Я.Ю. Об асимптотической эффективности критериев нормальности, основанных на свойстве Шеппа. // Вестник СПБГУ. 2009. Сер. 1, вып. 4. С. 13–19.

[П3] Волкова К.Ю., Никитин Я.Ю. Критерии нормальности, основанные на характеризации Галамбоша–Симонелли. // Обозрение прикл. и промышл. матем. 2009. Т. 16, № 2. С. 256.

[П4] Nikitin Ya.Yu., Volkova K.Yu. Asymptotic efficiency of exponentiality tests based on order statistics characterization. // Georgian Math. J. 2010. V. 17. P. 749–763.

Другие публикации:

[П5] Nikitin Ya.Yu., Volkova X. Tests of normality based on Shepp property, and their efficiencies. — Abstracts of the 2th Workshop of the ERCIM Working Group on Computing and Statistics. 2009. P. 87.

[П6] Volkova K. Tests of the exponentiality based on properties of order statistics. — Proceedings of the 6th St.Petersburg Workshop on Simulation. 2009. V. 2. P. 761–764.

[П7] Volkova K.Yu. On asymptotic efficiency of exponentiality tests based on properties of order statistics. — Abstracts of the 10th International Vilnius Conference on Probability Theory and Mathematical Statistics. 2010. P. 288–289.

[П8] Volkova K.Yu. Goodness-of-fit tests based on distribution characterizations, and their efficiencies. — Proceedings of Workshops on Inverse Problems, Data, Mathematical Statistics and Ecology, Linköping University. 2011. P. 129–133.

[П9] Nikitin Ya.Yu., Litvinova V.V., Volkova K.Yu. U-empirical tests of fit based on characterizations, and their efficiencies. — Book of Abstracts and Program of the Conference "Analytical Methods in Statistics" (Amistat 2011), October 28–30, Prague, Czech Republic. P. 6–7.