

Сибирский государственный университет
телекоммуникаций и информатики

**ОБРАБОТКА ИНФОРМАЦИИ
И
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ**

**РОССИЙСКАЯ
НАУЧНО-ТЕХНИЧЕСКАЯ
КОНФЕРЕНЦИЯ**

МАТЕРИАЛЫ КОНФЕРЕНЦИИ

Новосибирск
2020

978-5-91434-054-1

© ФГБОУ ВО «Сибирский государственный университет телекоммуникаций и информатики» 2020
© Авторы 2020

СОДЕРЖАНИЕ

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Дошинский И.В. Оптимизация методов определения местоположения объектов внутри помещений в сети Wi-Fi.	5
Елизов А.И. Сравнение алгоритмов повторных запросов в HTTP-балансировке с использованием инструментов моделирования.	10
Ефимов А.В., Поляков А.Ю., Крамаренко К.Е., Бочкарев Б.В. Исследование проблемы оптимизации доступа к информации о среде исполнения параллельных вычислительных задач.	16
Крамаренко К.Е., Бочкарев Б.В. Инструментальные средства дешифрации синдрома вычислительных систем.	22
Курносов М.Г., Ткачева Т.А., Токмашева Е.И. Анализ временных характеристик конвейерных алгоритмов операции Bcast для многопроцессорных узлов с общей памятью.	26
Курносов М.Г. Учет топологии многопроцессорного NUMA-узла при реализации широковещательной передачи стандарта MPI.	32
Пименов Е.С. Обзор методов нечеткого словарного поиска в Web и мобильных приложениях.	38

ИНФОРМАТИКА И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Баргуев С.Г. Доказательство существования решения одной начально-краевой задачи путем разложения по собственным функциям.	42
Баргуев С.Г. Доказательство существования решения начально-краевой задачи о колебаниях системы упругого стержня с твердым телом методом Галеркина.	49
Баргуев С.Г. О полноте гильбертовых пространств в одной начально-краевой задаче.	54
Гайдук П.А., Крюкова Л.П. Метод для оценки и сравнения эффективности алгоритмов сборки мусора.	58
Гвоздев С.Е. Эффективный приближенный алгоритм нахождения остовного леса в неориентированном графе, каждому ребру которого приписаны два веса.	63
Епифанцева К.Г., Полетайкин А.Н. Применение алгоритмов классификации в системе организации образовательного контента.	68
Захарова Т.Э. Методика определения функциональных зависимостей и параметров определяющих уравнений ползучести и повреждаемости.	73
Казначеев Д.А., Казначеева Н.В. Использование логистической регрессии в прогнозировании поведения абитуриентов.	78
Латкин В.И., Утепов С.Д. Задача о распределении оптимальной нагрузки.	83
Лемешко Б.Ю., Блинов П.Ю. Руководство по применению критериев проверки экспоненциальности.	92
Лемешко Б.Ю., Лемешко С.Б., Веретельникова И.В. Расширение области применения критериев однородности.	100
Ляхов О.А. Сетевое планирование пограничных операций.	106
Мачикина Е.П. Исследование статистических свойств импульсов в волоконно-оптических линиях связи.	117
Морозова К.И., Ракитский А.А. Разработка и исследование метода на основе нейронных сетей для восстановления аудиосигналов.	122
Нестеров А.С., Матвеев А.П. Применение алгоритмов нечеткого поиска в задаче устранения искажений в текстовых данных при их ручном вводе.	128
Осинцева Е.А., Чимитова Е.В. Процедура поиска оптимального плана эксперимента для винеровской деградиационной модели с учётом объясняющих переменных.	135
Павлова У.В., Ракитский А.А. Разработка и исследование метода прогнозирования временных рядов на основе конечных автоматов.	142

Панасенко И.А., Полетайкин А.Н., Сони́на С.Д. Математическая модель учёта и анализа геоданных о транспортных потоках крупного города.	146
Попов А.А. Идентификация локально адаптивных регрессионных моделей.	155
Ракитский А.А., Шубин М.О. Разработка и исследование стратегии управления ставками для автоматизации процессов принятия решений.	161
Тимофеев В.С., Исаева Е.В., Малышкина Е.Д., Слободчикова А.Э. Оценивание функции плотности распределения с использованием вейвлета Литлвуда–Пэли.	165
Тимофеев В.С., Санина А.А. Программа для решения задач классификации с применением моделей бинарного выбора.	171
Устинова К.А., Полетайкин А.Н., Кунц Е.Ю. Информационная модель планирования профессиональной траектории ИТ-специалистов.	179
Филимонова Н.А., Колпаков А.Г. Экспериментально-численный подход к моделированию трафика от служб Интернет в локальных сетях ЭВМ.	189
Чирихин К.С. Реализация алгоритма прогнозирования временных рядов на основе методов сжатия данных и искусственного интеллекта.	195

Расширение области применения критериев однородности

Б. Ю. Лемешко, С. Б. Лемешко, И. В. Веретельникова¹
Новосибирский государственный технический университет

В работе рассмотрено применение критериев однородности в качестве аналога критериев, используемых для проверки гипотез об отсутствии тренда. Рассмотрено также применение критериев однородности для проверки гипотезы о симметричности наблюдаемого закона. Соответствующие возможности применения реализованы в рамках развиваемой программной системы.

Ключевые слова: проверка гипотез, проверка симметричности, критерии однородности, статистическое моделирование

1. Введение

В процессе подготовки руководства по критериям однородности [1] и дальнейших исследований [2–7] в развиваемую программную систему [8] были встроены возможности применения и исследования множества критериев проверки гипотез об однородности законов, об однородности математических ожиданий анализируемых выборок и об однородности дисперсий.

Ранее при исследовании свойств некоторых критериев [9–11], используемых при проверке гипотез об отсутствии тренда в наблюдаемых временных рядах, было отмечено, что в этих задачах много общего с проверкой однородности. Более того, одними и теми же авторами были предложены похожие критерии для проверки однородности и для проверки гипотез об отсутствии тренда.

В этой связи естественно использовать реализованные в [8] возможности для применения различных критериев однородности при проверке гипотез об отсутствии тренда.

Вторая идея связана с использованием критериев однородности законов для проверки гипотезы о симметричности наблюдаемого закона. С критерием симметричности всё ясно при выполнении предположения о нормальности закона. В случае нарушения этого предположения аналогичная проверка уже не является очевидной. Идея использования критериев однородности законов для проверки симметричности естественна и, вероятно, не нова, но нам о таких реализациях неизвестно. Возможно, причина этого кроется в недостаточной разработанности и недостаточной известности аппарата, применяемого для анализа однородности k выборок.

Задача проверки гипотезы об однородности законов, соответствующих k выборкам, формулируется следующим образом. Имеется k выборок

$$x_{11}, x_{12}, \dots, x_{1n_1}, x_{21}, x_{22}, \dots, x_{2n_2}, \dots, x_{k1}, x_{k2}, \dots, x_{kn_k},$$

¹ Работа выполнена при поддержке Министерства образования и науки РФ в рамках государственной работы «Обеспечение проведения научных исследований» (№ 1.4574.2017/6.7) и проектной части государственного задания (№ 1.1009.2017/4.6).

где n_i – объём i -й выборки, $i = \overline{1, k}$. Проверяется гипотеза о том, что все выборки извлечены из одной и той же генеральной совокупности, т. е. $H_0: F_1(x) = F_2(x) = \dots = F_k(x) = F(x)$ при любом x . Конкурирующая гипотеза может иметь вид $H_1: F_i(x) \neq F_j(x)$ для некоторых $i \neq j$, $i, j \leq k$. Как правило, критерии проверки гипотезы об однородности законов непараметрические.

В системе [8] реализованы критерии Смирнова [12], Лемана–Розенблатта [13, 14], Андерсона–Дарлинга–Петита [15], k -выборочный вариант критерия Андерсона–Дарлинга [16], 3 k -выборочных критерия Жанга [17–19] и предложенные нами в [6] k -выборочные критерии, опирающиеся на использование двухвыборочных.

При проверке гипотезы об однородности средних (о равенстве математических ожиданий) предполагается, что анализируемые выборки принадлежат какому-то одному закону: неизвестному в случае непараметрических критериев и известному в случае параметрических. Проверяемая гипотеза имеет вид $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ при конкурирующей гипотезе $H_1: \mu_{i_1} \neq \mu_{i_2}$, где неравенство выполняется хотя бы для некоторой пары индексов i_1, i_2 .

Для проверки гипотез о равенстве математических ожиданий может использоваться ряд параметрических критериев, применение которых, как правило, опирается на стандартное предположение о принадлежности анализируемых выборок нормальному закону, а также непараметрические критерии, свободные от этого предположения. В [8] реализовано в общей сложности 10 параметрических и непараметрических критериев однородности средних.

В критериях проверки однородности дисперсий проверяемая гипотеза о постоянстве дисперсий k выборок имеет вид $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$, а конкурирующая с ней гипотеза $H_1: \sigma_{i_1}^2 \neq \sigma_{i_2}^2$, где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2 .

Для проверки такого рода гипотез может использоваться значительный перечень классических параметрических критериев. Обоснованное применение этих критериев требует выполнения стандартного предположения о принадлежности анализируемых выборок нормальному закону. В системе [8] реализовано порядка 20 параметрических и непараметрических критериев, применяемых для проверки гипотезы об однородности дисперсий.

2. Применение критериев однородности для проверки отсутствия тренда

В критериях проверки гипотез об отсутствии тренда в качестве проверяемой гипотезы H_0 рассматривается отсутствие среди элементов выборки каких-либо вероятностных или детерминированных зависимостей, то есть элементы выборки представляют собой независимые одинаково распределённые величины. Отклонение H_0 ещё не говорит о том, что присутствует тренд, и тем более ничего не говорит о виде возможного тренда. Определение характера тренда является целью уже другой задачи.

Разбив анализируемую выборку на части, можно проверять гипотезу о принадлежности этих частей одному и тому же закону распределения (проверять гипотезу об однородности законов), или проверять гипотезу о равенстве математических ожиданий этих частей (проверять гипотезу об однородности средних), или проверять гипотезу о равенстве дисперсий (проверять гипотезу об однородности дисперсий), используя для этого соответствующие критерии однородности. Если гипотеза H_0 не отклоняется, то это может говорить об отсутствии тренда. При отклонении H_0 можно подозревать наличие какого-то тренда в математическом ожидании, в дисперсии или в математическом ожидании и в дисперсии. Судить об этом можно по виду применяемого критерия однородности.

Анализируемую выборку можно разбить на k частей отдельно (вне рамок системы [8]), а затем к этим k выборкам применить соответствующие критерии однородности. Можно ана-

лизируемую выборку разбить на k частей прямо в системе [8], а к анализу этих частей привлечь нужные критерии. Во втором случае в системе откроется форма, позволяющая для анализа этих частей выбрать, например, определённые критерии однородности законов (см. рис.1), или критерии однородности дисперсий (рис. 2), или критерии однородности средних.

Однородность распределений

Загрузите выборки для анализа

☒ ☐ Выделить все выборки

Название выборки	n
☒ Первая часть выборки	25000
☒ Вторая часть выборки	75000

Уровень значимости
☐ 0.1 ☐ 0.05 ☒ 0.01

Выберите критерии

- ☐ Критерий однородности Хи-квадрат
- ☒ Критерий однородности Андерсона-Дарлинга-Петита
- ☐ К-выборочный критерий Андерсона-Дарлинга
- ☒ Критерий однородности Лемана-Розенблатта
- ☒ Критерий однородности Смирнова
- ☐ Критерий однородности Смирнова (модиф.)
- ☐ К-выборочный max Андерсона-Дарлинга-Петита
- ☐ К-выборочный max Лемана-Розенблатта
- ☐ К-выборочный max Смирнова (модиф.)
- ☐ Критерий Жанга Z_a (K выборок)
- ☐ Критерий Жанга Z_c (K выборок)
- ☐ Критерий Жанга Z_k (K выборок)

Проверить

Рис. 1. Форма для выбора используемых критериев однородности законов

Однородность дисперсий

Загрузите выборки для анализа

☒ ☐ Выделить все выборки

Название выборки	n
☒ Первая часть выборки	25000
☒ Вторая часть выборки	75000

Если предполагаемый закон распределения, которому соответствуют анализируемые выборки, отличается от нормального, задайте его в меню 'Параметры'.

Выберите критерии

Параметрические критерии

- ☒ Критерий Бартлетта
- ☐ Критерий Блисса-Кокрена-Тьюки
- ☐ Критерий Кэдуэлла-Лесли-Брауна
- ☐ Критерий Кокрена
- ☐ F-критерий Фишера
- ☒ Критерий Хартли
- ☐ Критерий Лайарда
- ☐ Критерий Лайарда Мод.
- ☒ Критерий Левене (со средним)
- ☐ Критерий Левене (с медианой)

Непараметрические критерии

- ☒ Критерий Ансари-Бредли
- ☐ Критерий Ансари-Бредли (Нормированный)
- ☐ Критерий Кейпена
- ☐ Критерий Кейпена (Нормированный)
- ☐ Критерий Коновера
- ☐ Критерий Флигнера-Киллина
- ☒ Критерий Клотца
- ☐ Критерий Клотца (Нормированный)
- ☐ Критерий Муда
- ☐ Критерий Муда (Нормированный)

Параметры Проверить

Рис. 2. Форма для выбора используемых критериев однородности дисперсий

При использовании критериев однородности законов достигнутые уровни значимости p_{value} по значениям статистик для критериев Андерсона–Дарлинга, Лемана–Розенблатта и Смирнова находятся по известным асимптотическим распределениям, для других критериев по построенным нами моделям предельных распределений [1, 6, 7]. Распределения статистик критериев Жанга [17–19] и оценки p_{value} в системе [8] находятся по результатам проводимого статистического моделирования.

При использовании параметрических критериев однородности дисперсий в условиях нарушения стандартного предположения о нормальности распределения статистик этих критериев и оценки p_{value} в системе [8] также находятся на основании статистического моделирования (при конкретных предполагаемых законах).

3. Проверка симметричности закона

Проверка симметричности с использованием совокупности критериев однородности законов осуществляется следующим образом. Анализируемая выборка сортируется по возрастанию $x_1 \leq x_2 \leq \dots \leq x_n$. Находится оценка медианы, которая вычитается из всех элементов x_i выборки. Затем полученная выборка делится на 2 выборки объёмом $n/2$ при четном n , и $(n-1)/2$ – при нечетном n (значение, совпадающее с медианой из исходной выборки исключается). Все элементы первой выборки берутся по модулю. Далее проверяется гипотеза об однородности законов распределения, соответствующих этим двум выборкам. При симметричности закона распределения гипотеза об однородности законов полученных выборок не должна отклоняться.

Проверка симметричности закона распределения

Загрузите выборки для анализа

☐ Выделить все выборки

Название выборки	n
☒ Нормальное с масштабом 1.0000 со сдвигом 0.0000	300

Выберите критерии

- ☐ Критерий однородности Хи-квадрат
- ☒ Критерий однородности Андерсона-Дарлинга-Петита
- ☐ К-выборочный критерий Андерсона-Дарлинга
- ☒ Критерий однородности Лемана-Розенблатта
- ☐ Критерий однородности Смирнова
- ☒ Критерий однородности Смирнова (модиф.)
- ☐ К-выборочный max Андерсона-Дарлинга-Петита
- ☐ К-выборочный max Лемана-Розенблатта
- ☐ К-выборочный max Смирнова (модиф.)
- ☐ Критерий Жанга Za (K выборок)
- ☐ Критерий Жанга Zc (K выборок)
- ☐ Критерий Жанга Zk (K выборок)

Уровень значимости: 0.1

Результаты проверки гипотезы

Заданный уровень значимости: 0.01

Критерий	Гипотеза	Статистика	p-value
Критерий Андерсона-Дарлинга-Петита	НЕ ОТКЛОНЯЕТСЯ	0.459619	0.788222
Критерий Лемана-Розенблатта	НЕ ОТКЛОНЯЕТСЯ	0.0802	0.690699
Критерий Смирнова (модиф.)	НЕ ОТКЛОНЯЕТСЯ	0.660187	0.776067

Требуется моделирование для следующих критериев:

- ☒ Критерий однородности Андерсона-Дарлинга-Петита
- ☒ Критерий однородности Лемана-Розенблатта
- ☒ Критерий однородности Смирнова (модиф.)

Моделировать Отмена

Рис. 3. Форма при проверке симметричности закона

Именно такая процедура проверки симметричности закона, опирающаяся на реализованные критерии однородности, применяется в системе [8]. На рис. 3 показана форма с выбором критериев однородности, используемых при проверке симметричности, и вкладка с результатами проверки симметричности.

4. Заключение

В работе предложено для проверки гипотез об отсутствии тренда и гипотез о симметричности наблюдаемого закона применять критерии проверки гипотез об однородности законов, об однородности средних и дисперсий. Такая стратегия реализована в развиваемом программном обеспечении [8].

В результате исследований расширено множество критериев, которые могут эффективно и корректно использоваться в приложениях для выявления тренда в анализируемых временных рядах.

Использование непараметрических критериев однородности позволяет осуществлять корректные статистические выводы при проверке симметричности в случае произвольных законов распределения.

Полученные результаты расширяют сферу применения совокупности критериев однородности в различных приложениях.

Литература

1. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению. М.: ИНФРА-М, 2017. – 208 с. DOI: 10.12737/22368
2. Лемешко Б.Ю., Сатаева Т.С. Применение и мощность параметрических критериев проверки однородности дисперсий. Ч. 3 // Измерительная техника. 2017. № 1. – С. 8-13.
3. Лемешко Б.Ю., Сатаева Т.С. Применение и мощность параметрических критериев проверки однородности дисперсий. Ч. 4 // Измерительная техника. 2017. № 5. – С. 12-17.
4. Лемешко Б.Ю., Лемешко С.Б., Веретельникова И.В. О применении критериев проверки однородности законов распределения // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2017. № 41. – С. 24-31. DOI: 10.17223/19988605/41/3
5. Лемешко Б.Ю. О применении критериев проверки однородности средних / Б. Ю. Лемешко, С.Б. Лемешко, И.В. Веретельникова, А.Ю. Новикова // Вестник СибГУТИ. 2018. № 1. – С. 41-55.
6. Лемешко Б.Ю., Веретельникова И.В. Мощность k -выборочных критериев проверки однородности законов // Измерительная техника. 2018. № 7. – С. 3-7. DOI: 10.32446/0368-1025it-2018-7-3-7
7. Lemeshko B., Veretelnikova I. On application of k -samples homogeneity tests // Applied Methods of Statistical Analysis. Statistical Computation and Simulation - AMSA'2019, Novosibirsk, Russia, 18-20 September, 2019: Proceedings of the International Workshop. - Novosibirsk: NSTU publisher, 2019. - P.138-151.
8. ISW – Программная система статистического анализа одномерных наблюдений. <https://ami.nstu.ru/~headrd/ISW.htm>. (дата обр. 12.04.2020)
9. Лемешко Б.Ю., Комиссарова А.С., Щеглов А.Е. Применение некоторых критериев проверки гипотез случайности и отсутствия тренда // Метрология. 2010. № 12. – С. 3-25.
10. Лемешко Б.Ю., Комиссарова А.С., Щеглов А.Е. Свойства и мощность некоторых критериев случайности и отсутствия тренда // Научный вестник НГТУ. 2012. № 1(46). – С. 53-66.

11. *Veretel'nikova I.V., Lemeshko B.Yu.* Criteria of Test against Absence of Trend in Dispersion Characteristics // Proceedings 2016 11th International Forum on Strategic Technology (IFOST), June 1-3, 2016, Novosibirsk, Russia. Part 1. – P. 333-337. DOI: 10.1109/IFOST.2016.7884121
12. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
13. *Lehmann E.L.* Consistency and unbiasedness of certain nonparametric tests // Ann. Math. Statist. 1951. Vol. 22 No. 1. – P. 165–179.
14. *Rosenblatt M.* Limit theorems associated with variants of the von Mises statistic // Ann. Math. Statist. 1952. Vol. 23. – P. 617–623.
15. *Pettitt A.N.* A two-sample Anderson-Darling rank statistic // Biometrika. 1976. Vol. 63. No.1. P. 161-168.
16. *Scholz F.W., Stephens M.A.* K-Sample Anderson–Darling Tests // Journal of the American Statistical Association. 1987. Vol. 82. No. 399. – P. 918-924.
17. *Zhang J.* Powerful goodness-of-fit and multi-sample tests / J. Zhang // PhD Thesis. York University, Toronto. 2001. – 113 p. URL: <http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf> (дата обращения 28.03.2020)
18. *Zhang J.* Powerful Two-Sample Tests Based on the Likelihood Ratio // Technometrics. 2006. Vol. 48. No. 1. – P.95-103. DOI 10.1198/004017005000000328
19. *Zhang J., Wu Y.* k-Sample tests based on the likelihood ratio // Computational Statistics & Data Analysis. 2007. V. 51. No. 9. – P. 4682-4691.

Лемешко Борис Юрьевич

Профессор кафедры прикладной и теоретической информатики НГТУ, д.т.н., профессор (630073, Новосибирск, пр. Карла Маркса, 20), тел. (383) 346-06-00, e-mail: Lemeshko@ami.nstu.ru, <http://www.ami.nstu.ru/~headrd/>

Лемешко Станислав Борисович

С.н.с. кафедры прикладной и теоретической информатики НГТУ, к.т.н., e-mail: skyer@mail.ru.

Веретельникова Ирина Викторовна

М.н.с. кафедры прикладной и теоретической информатики НГТУ, к.т.н., e-mail: ira-veterok@mail.ru.

Expanding the area of the homogeneity tests

B. Yu. Lemeshko, S. B., Lemeshko, I. V. Veretelnikova

The paper considers the application of homogeneity tests as an analogue of the criteria used to test hypotheses about the absence of a trend. The application of homogeneity tests to test the hypothesis of symmetry of the observed law is also considered. Corresponding application possibilities are realized within the framework of the developed software system.

Keywords: hypothesis testing, symmetry checking, homogeneity tests, statistical simulating.