

БЕЛКОВА Анна Андреевна

Выпускная квалификационная работа

**ИССЛЕДОВАНИЕ МОЩНОСТИ ПЕРЕСТАНОВОЧНЫХ ТЕСТОВ ДЛЯ
ПРОВЕРКИ ГИПОТЕЗ О РАВЕНСТВЕ ДВУХ РАСПРЕДЕЛЕНИЙ**

Уровень образования: бакалавриат

Направление 01.03.02 «Прикладная математика и информатика»

Основная образовательная программа СВ.5004.2017 «Прикладная математика и
информатика»

Профиль «Вычислительная стохастика и статистические модели»

Научный руководитель:

Профессор, кафедра статистического
моделирования,
д. ф.-м. н., профессор В. Б. Мелас

Рецензент:

Профессор, кафедра математического
обеспечения и применения ЭВМ,
Санкт-Петербургский государственный
электротехнический университет,
д. т. н., профессор Ю. Д. Григорьев

Saint Petersburg State University
Applied Mathematics and Computer Science
Computational Stochastics and Statistical Models

BELKOVA Anna Andreevna

Graduation Project

**INVESTIGATION OF PERMUTATION TESTS POWER FOR TESTING
HYPOTHESES OF TWO DISTRIBUTIONS EQUALITY**

Scientific Supervisor:

Professor, Department of Statistical

Modelling,

doctor of physical and mathematical sciences

V. B. Melas

Reviewer:

Professor, St. Petersburg State

Electrotechnical University,

doctor of technology Y. D. Grigoriev

Saint Petersburg

2021

Оглавление

Введение	4
Глава 1. Новый критерий и его реализация с помощью метода перестановок	5
1.1. Новый тест для проверки гипотез	5
1.2. Использование перестановочных методов	6
Глава 2. Теоретические результаты	8
2.1. Теорема о виде распределения статистики критерия и о его мощности	8
2.2. Продолжение исследования теста. Оценки J_0 , $J^*(h)$, \bar{b} и a для нескольких видов распределений	9
2.2.1. Оценка через разложение в ряд Тейлора. Случай нормального распределения	10
2.2.2. Оценка через минимизацию при фиксированном объеме выборки. Случай нормального распределения	11
2.2.3. Оценка через минимизацию при фиксированном объеме выборки. Случай распределения Коши	13
2.2.4. Оценка через минимизацию при фиксированном объеме выборки. Случай распределения Лапласа	14
Глава 3. Результаты моделирования	16
3.1. Описание программы	16
3.2. Результаты работы программы. Таблицы	16
3.3. Графики	22
Заключение	25
Список литературы	26

Введение

Недавно В. Б. Меласом и Д. И. Сальниковым был предложен новый тест для проверки гипотезы о равенстве двух генеральных совокупностей [1]. Естественно, существует достаточно много тестов для проверки подобных гипотез, однако, с одной стороны, это могут быть универсальные тесты, как тест Колмогорова – Смирнова, которые маломощны в силу своей универсальности и показывают слабые результаты для распределений, отличающихся параметром масштаба, или более мощные тесты, но направленные прицельно на конкретный вид распределений, как, к примеру, тест Андерсона – Дарлингга. Идеей авторов статьи было предложить мощный тест, применимый для широкого класса распределений.

Исследование, положенное в основу выпускной квалификационной работы, было начато в предыдущем семестре. Целью ВКР является численное исследование нового метода. В ходе работы необходимо было разобраться в сущности метода, а также в аналитических выкладках. В процессе работы при проведении численных исследований был написан код на языке R, который с помощью методов стохастического моделирования позволяет получить значение мощности теста для случая различных нулевых гипотез и альтернатив.

Глава 1

Новый критерий и его реализация с помощью метода перестановок

1.1. Новый тест для проверки гипотез

Сформулируем гипотезу и опишем статистику нового теста. Ставится задача проверки гипотезы о принадлежности двух выборок одному распределению. Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — две независимые выборки с функциями распределения соответственно F и G . В дальнейшем для удобства будем считать, что $m = n$, однако равенство объемов выборки не имеет принципиального значения. Допустим также, что эти распределения таковы, что они относятся к классу распределений, таких что принадлежащие ему случайные величины ξ обладают свойством

$$E(\ln^2(1 + \xi^2)) < \infty. \quad (1.1)$$

Большинство распределений (в частности, распределение Коши и нормальное распределение) обладают этим свойством. Зафиксируем нулевую гипотезу

$$H_0 : F(x) = G(x) \text{ для любого } x \in \mathbb{R} \quad (1.2)$$

и альтернативу

$$H_1 : F(x) \neq G(x) \text{ для хотя бы одного } x \in \mathbb{R}. \quad (1.3)$$

Статистика теста имеет вид

$$\begin{aligned} T_n &= \Phi_A + \Phi_B + \Phi_{AB}, \\ \Phi_A &= -\frac{1}{n^2} \sum_{i < j}^n g(|X_i - X_j|), \\ \Phi_B &= -\frac{1}{n^2} \sum_{i < j}^n g(|Y_i - Y_j|), \\ \Phi_{AB} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n g(|X_i - Y_j|), \end{aligned}$$

где $g(|u|) = \ln(1 + |u|^2)$.

Основная идея нового теста — сравнение распределений общего вида, отличающихся сдвигом и (или) масштабом [1].

Необходимо отметить, что похожий тест был независимо предложен ранее в работе Аслана и Цеха [2], однако тест исследовался только эмпирически и применялся для набора примеров, в том числе многомерных, но подобранных без ясно сформулированного принципа. Аналитических результатов в статье [2] нет.

1.2. Использование перестановочных методов

В практической части работы использован перестановочный метод нахождения доверительной и критической области значений статистики критерия. Перестановочными тестами называют статистические методы, используемые для определения достигнутого уровня значимости теста, основанные на алгоритме с перестановками внутри выборок. Идея была введена Фишером в 1930-е гг., однако являлась скорее теоретическим аргументом в поддержку t -критерия Стьюдента, нежели непосредственно полезным статистическим методом. Сейчас же современные вычислительные мощности позволяют использование перестановочных тестов на рутинной основе. Опишем суть метода.

Пусть мы имеем:

- Две независимые выборки: $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_n)$;
- Гипотезу о равенстве двух распределений, например:

$$H_0 : F(x) = G(x) \text{ для любого } x \in \mathbb{R},$$

$$H_1 : F(x) \neq G(x) \text{ для хотя бы одного } x \in \mathbb{R};$$
- Статистику критерия $\varphi = \varphi(X, Y)$.

Ставится задача найти доверительную и критическую области значений статистики критерия.

Идея метода заключается в том, что если H_0 верна, то элементы выборок X и Y можно воспринимать как элементы из одной выборки. В этом случае если поменять в выборках X и Y некоторые элементы X_i и Y_j местами, получатся новые выборки X' и Y' , статистика критерия от которых не будет значительно отличаться от $\varphi(X, Y)$.

Исходя из этого можно сформулировать алгоритм [3] :

- Объединим X и Y в единую выборку Z ;

- Найдем K случайных разбиений Z на $(X^{(i)}, Y^{(i)})$, таких что: $\#X^{(i)} = \#Y^{(i)}$, $X^{(i)} \cup Y^{(i)} = Z$, $X^{(i)} \cap Y^{(i)} = \emptyset$;
- Найдем $ASL_{perm} = \frac{\#\{\varphi(X, Y) \leq \varphi(X^{(i)}, Y^{(i)}), i=1, \dots, K\}}{K}$ — достигнутый уровень значимости теста;
- Сравним заранее выбранный уровень значимости α с ASL_{perm} .

Глава 2

Теоретические результаты

2.1. Теорема о виде распределения статистики критерия и о его мощности

Мелас и Сальников в работе [1] поставили своей основной задачей аналитическое исследование теста — получение таких результатов как, к примеру, асимптотические формулы для значения мощности теста. Ключевой результат статьи Меласа и Сальникова — формулировка и доказательство теоремы, демонстрирующей предельное распределение статистики критерия и содержащей асимптотическую формулу для мощности нового теста. Перед формулировкой теоремы зафиксируем некоторые условные обозначения.

Через $C(v, u)$ обозначим распределение Коши с плотностью

$$\frac{v}{\pi(v^2 - (x - u)^2)}.$$

Пусть $f_1(x)$ и $f_2(x)$ плотности, соответствующие функциям распределения F_1 и F_2 , где $F_2 = F_1(x - \theta)$, $\theta = h/\sqrt{n}$, $h, n \in \mathbb{R}$. Обозначим

$$J_h = \int_R g(x - y) f_1(x) f_2(y) dx dy = \int_R g(x - y - |h|/\sqrt{n}) f_1(x) f_1(y) dx dy,$$

$$J^*(h) = \lim_{n \rightarrow \infty} n(J_h - J_0). \quad (2.1)$$

Существование предела (2.1) для симметричных распределений доказано в [1].

Через \bar{b} обозначим выражение $\sqrt{J^*(h)/h^2}$.

Далее в разделе 2.2 будет показано, что $J^*(h)$ для конкретного распределения можно получить через разложение в ряд Тейлора, и также будет приведен альтернативный способ нахождения \bar{b} .

Теорема 1 ([1]) *Рассмотрим задачу проверки гипотезы о равенстве двух распределений (1.2)–(1.3), причем пусть оба распределения симметричны и обладают свойством (1.1). В этом случае:*

(1) *При верной H_0 и $n \rightarrow \infty$ распределение nT_n стремится к распределению с.в.*

$$(aZ)^2,$$

где $Z \sim N(0, 1)$; $a \in \mathbb{R}$, $a > 0$;

(2) Положим $F_1 = F(x)$, $F_2 = F(x + \theta)$, где F функция распределения, которая соответствует некоторому симметричному распределению, обладающему свойством (1.1), $\theta = h/\sqrt{n}$, h — произвольное число. Тогда распределение nT_n при $n \rightarrow \infty$ стремится к распределению случайной величины

$$(aZ + b)^2,$$

где $Z \sim N(0, 1)$; $a, b \in \mathbb{R}$, $b = 0$, если верна H_0 и $b = \bar{b}h$ при верной H_1 . И в этом случае мощность критерия T_n при уровне значимости α асимптотически эквивалентна величине, получаемой по формуле

$$Pr\{Z \geq z_{1-\alpha/2} - \bar{b}h/a\} + Pr\{Z \leq -z_{1-\alpha/2} - \bar{b}h/a\},$$

где $z_{1-\alpha/2}$ — $(1 - \alpha/2)$ -квантиль нормального распределения, то есть

$$Pr\{Z \geq z_{1-\alpha/2}\} = \alpha/2.$$

Если $F_1 = C(\nu, 1)$ и $F_2 = C(\nu + \theta, 1)$, то $b = h/3$.

Важно отметить, что аналитическое представление коэффициента a является пока сложной нерешенной задачей. Также отметим, что формулу для мощности можно записать в более удобном для практических целей виде как

$$\begin{aligned} Pr\{Z \geq z_{1-\alpha/2} - \bar{b}h/a\} + Pr\{Z \leq -z_{1-\alpha/2} - \bar{b}h/a\} = \\ = 1 - \Phi(z_{1-\alpha/2} - \bar{b}h/a) + \Phi(-z_{1-\alpha/2} - \bar{b}h/a), \end{aligned}$$

где Φ — функция распределения $N(0, 1)$. Доказательство теоремы базируется на работе с интегралами J_h и использовании закона больших чисел и центральной предельной теоремы в случае U -статистик [4, 5]. Значение коэффициента b в случае распределения Коши получается из-за возможности в явном виде сосчитать интегралы J_h и предел $J^*(h)$.

2.2. Продолжение исследования теста. Оценки J_0 , $J^*(h)$, \bar{b} и a для нескольких видов распределений

Рассмотрим сначала случай нормальных распределений, отличающихся параметром сдвига и имеющих параметр масштаба равный 1.

2.2.1. Оценка через разложение в ряд Тейлора. Случай нормального распределения

В случае нормального распределения посчитать аналитически интегралы J_h , как это было сделано в случае распределения Коши, сложнее, поэтому одной из задач ВКР было численно получить соответствующие оценки. Интеграл J_h в этом случае принимает вид

$$\begin{aligned} J_h &= E \ln(1 + (X - Y - \theta)^2) = \int_R \int_R \frac{\ln(1 + (x - y - \theta)^2) e^{-\frac{(x^2+y^2)}{2}}}{2\pi} dx dy = \\ &= \int_R \int_R u(x, y, \theta) dx dy, \end{aligned}$$

$$J^*(h) = \lim_{n \rightarrow \infty} n(J_h - J_0).$$

Численно получить оценку для интеграла J_0 не составляет труда, в данном случае для J_0 было получено значение $J_0 \approx 0.810113$. Однако J_h — интеграл с параметром, а оценка предела $J^*(h)$ должна была быть получена в виде функции от h . Поэтому приблизим нужную нам функцию с помощью разложения в ряд Тейлора. Тогда для $J^*(h)$ может быть использована оценка

$$J^*(h) \approx \int_R \int_R \frac{1}{2} u''_{\theta}(x, y, 0) dx dy \cdot h^2 \approx 0.2271795 h^2 \approx \frac{h^2}{4.4},$$

то есть интеграл от половины второй производной подынтегральной функции J_h по параметру θ в нуле. В результате получим значение коэффициента близкое к $1/4.4$.

Полученный результат для приближения предела $J^*(h)$ можно проверить другим численным методом — посчитав оценки для значений \hat{J}_0 , \hat{J}_h и $n(\hat{J}_h - \hat{J}_0)$ по методу Монте-Карло. В данном случае вычисления проведены с достаточно большой погрешностью, однако и они не противоречат приближению $J^*(h) \approx \frac{h^2}{4.4}$. В таблице 2.1 показаны результаты оценок.

Таким образом, при $J^*(h) \approx \frac{h^2}{4.4}$ получаем $\hat{b} = \sqrt{J^*(h)/h^2} \approx 1/(\sqrt{4.4}) \approx 0.477$ и $b = \bar{b}h \approx 0.477h$.

Таблица 2.1. Результаты оценки интегралов J_0 , J_h и предела $J^*(h)$ при $n = 10^4$

h	\hat{J}_0	\hat{J}_h	$n(\hat{J}_h - \hat{J}_0)$	$h^2/4.4$
1	0.810113	0.810136	0.23	0.227
2	0.810113	0.810204	0.91	0.909
4	0.810113	0.810477	3.64	3.636
6	0.810113	0.810931	8.18	8.18
8	0.810113	0.811567	14.54	14.545
10	0.810113	0.812384	22.71	22.727

2.2.2. Оценка через минимизацию при фиксированном объеме выборки.

Случай нормального распределения

С другой стороны, важно помнить, что формула для мощности является асимптотической, то есть иллюстрирует ситуацию, при которой объем выборки стремится к бесконечности. Это полезное знание, однако при практическом использовании теста мы всегда будем находиться в условиях некоторого фиксированного объема выборки. Поэтому также было принято решение определенным образом подобрать коэффициент \bar{b}/a в случае конкретного объема выборки — например, при $n = 100$.

Схема получения коэффициента \bar{b}/a путем моделирования и минимизации:

1. Запишем формулу для мощности как функцию

$$Q_{h_i}(\bar{b}/a) = 1 - \Phi(z_{1-\alpha/2} - \bar{b}h/a) + \Phi(-z_{1-\alpha/2} - \bar{b}h/a)$$

и получим эмпирические значения для мощности теста $\tilde{Q}_{h_i}(\bar{b}/a)$ в случае распределений $N(0, 1)$ и $N(h/\sqrt{n}, 1)$ при объеме выборки $n = 100$ (обычным образом, см. подробнее о численных исследованиях теста в главе 3). Результаты представлены в таблице 2.2.

2. Запрограммируем отдельной функцией формулу для Q из пункта 1. Код на языке R:

```
power_formula <- function(b, h)
{
  return(1 - pnorm(qnorm(1 - 0.05/2) - b*h) + pnorm(-qnorm(1 - 0.05/2) - b*h))
}
```

3. Для нахождения оптимального значения \bar{b}/a будем минимизировать сумму квадратов разницы между $Q_{h_i}(\bar{b}/a)$ и $\tilde{Q}_{h_i}(\bar{b}/a)$, то есть искать

$$\min_{\bar{b}/a} \sum_{i=1}^{12} (Q_{h_i}(\bar{b}/a) - \tilde{Q}_{h_i}(\bar{b}/a))^2.$$

4. Будем рассматривать разумные значения \bar{b}/a (возьмем \bar{b}/a на отрезке $[-5, 5]$ с шагом 0.01) и найдем оценку \bar{b}/a как

$$\widehat{\bar{b}/a} = \arg \min_{\bar{b}/a} \sum_{i=1}^{12} (Q_{h_i}(\bar{b}/a) - \tilde{Q}_{h_i}(\bar{b}/a))^2.$$

График оптимизируемой функции приведен на рис. 2.1, на рассматриваемом отрезке функция имеет два минимума в симметричных точках.

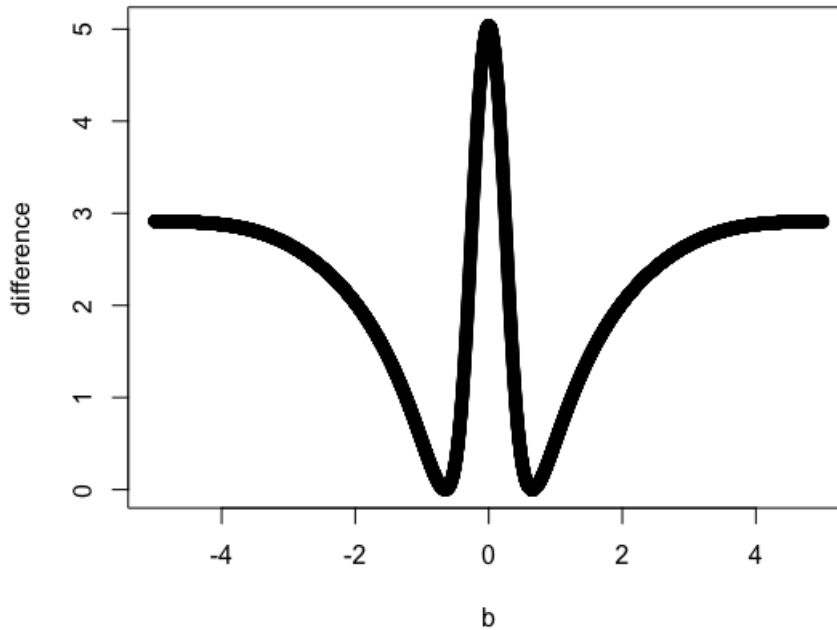


Рис. 2.1. Сумма квадратов разницы между $Q_{h_i}(\bar{b}/a)$ и $\tilde{Q}_{h_i}(\bar{b}/a)$ в зависимости от \bar{b}/a , нормальное распределение

Соответственно, полученный результат это $\widehat{\bar{b}/a} = 0.651$ (и $\widehat{\bar{b}/a} = -0.651$). С учетом полученного в разделе 2.2.1 значения оценки \bar{b} как 0.477, можем получить отдельно оценку для $\hat{a} = b/(\widehat{\bar{b}/a}) = 0.477/0.651 \approx 0.733$.

Таблица 2.2. Значение эмпирической мощности $\tilde{Q}_{h_i}(\bar{b}/a)$ и мощности $Q_{h_i}(\widehat{\bar{b}/a})$, посчитанной по формуле при $\widehat{\bar{b}/a} = 0.651$

h_i	$\tilde{Q}_{h_i}(\bar{b}/a)$	$Q_{h_i}(\widehat{\bar{b}/a})$
1.0	0.100	0.09468162
1.5	0.190	0.16265221
2.0	0.264	0.25574695
2.5	0.372	0.37065111
3.0	0.487	0.49840064
3.5	0.626	0.62628310
4.0	0.726	0.74151838
4.5	0.861	0.83498055
5.0	0.892	0.90320567
5.5	0.943	0.94802900
6.0	0.969	0.97453304
6.5	0.986	0.98863778

2.2.3. Оценка через минимизацию при фиксированном объеме выборки.

Случай распределения Коши

Получим приближение коэффициента \bar{b}/a для случая выборок из $Cauchy(0, 1)$ и $Cauchy(h/\sqrt{n}, 1)$ при $n = 100$.

Результат: $\widehat{\bar{b}/a} = 0.339$.

Поскольку в случае этого распределения значение для b получено в явном виде (см. Теорему 1) и равно $h/3$, можем получить отдельно значение для \bar{b} и оценку a . Тогда $\bar{b} = 1/3$ и $\hat{a} = \bar{b}/(\widehat{\bar{b}/a}) = (1/3)/0.339 \approx 0.983$.

Таблица 2.3. Значение эмпирической мощности $\tilde{Q}_{h_i}(\bar{b}/a)$ и мощности $Q_{h_i}(\widehat{\bar{b}/a})$, посчитанной по формуле при $\widehat{\bar{b}/a} = 0.339$

i	h_i	$\tilde{Q}_{h_i}(\bar{b}/a)$	$Q_{h_i}(\widehat{\bar{b}/a})$
1	1.0	0.060	0.06326617
2	1.5	0.076	0.08011007
3	2.0	0.092	0.10409791
4	2.5	0.130	0.13546602
5	3.0	0.152	0.17430525
6	3.5	0.211	0.22045021
7	4.0	0.265	0.27339043
8	4.5	0.351	0.33222141
9	5.0	0.378	0.39564721
10	5.5	0.461	0.46203868
11	6.0	0.544	0.52954163
12	6.5	0.605	0.59622058

2.2.4. Оценка через минимизацию при фиксированном объеме выборки.

Случай распределения Лапласа

Получим приближение коэффициента \bar{b}/a для случая выборок из $Laplace(0, 1)$ и $Laplace(h/\sqrt{n}, 1)$ при $n = 100$.

Результат: $\widehat{\bar{b}/a} = 0.558$.

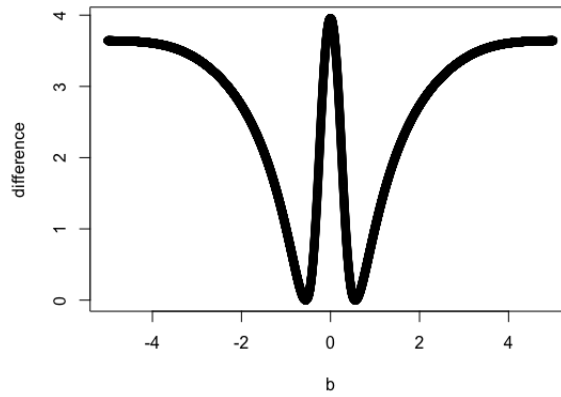


Рис. 2.2. Сумма квадратов разницы между $Q_{h_i}(\bar{b}/a)$ и $\tilde{Q}_{h_i}(\bar{b}/a)$ в зависимости от \bar{b}/a , распределение Лапласа

Таблица 2.4. Значение эмпирической мощности $\tilde{Q}_{h_i}(\bar{b}/a)$ и мощности $Q_{h_i}(\widehat{\bar{b}/a})$, посчитанной по формуле при $\widehat{\bar{b}/a} = 0.558$

i	h_i	$\tilde{Q}_{h_i}(\bar{b}/a)$	$Q_{h_i}(\widehat{\bar{b}/a})$
1	1.0	0.096	0.08636477
2	1.5	0.133	0.13330566
3	2.0	0.205	0.20039388
4	2.5	0.282	0.28644601
5	3.0	0.401	0.38759240
6	3.5	0.507	0.49726738
7	4.0	0.577	0.60721666
8	4.5	0.715	0.70919940
9	5.0	0.794	0.79674181
10	5.5	0.876	0.86629292
11	6.0	0.923	0.91743702
12	6.5	0.954	0.95224639

Глава 3

Результаты моделирования

Поскольку основная задача ВКР — численное исследование нового метода, в текущей главе будет приведено описание численных экспериментов и их результаты.

3.1. Описание программы

В рамках работы была создана программа на языке R, которая с помощью методов стохастического моделирования воспроизводит проверку гипотезы о равенстве двух распределений. Опишем основные этапы работы части программы, позволяющей получить численную оценку мощности теста:

- Сначала программа проверяет путем моделирования ситуации, соответствующей нулевой гипотезе, что ошибка первого рода приблизительно равна уровню значимости;
- В течение работы программы N раз генерируются 2 выборки заданных распределений (например, $Cauchy(0, 1)$ и $Cauchy(0, 1 + h/\sqrt{n})$);
- Для каждой пары считается значение нового критерия и других критериев и сохраняется информация о том, отвергается ли гипотеза о равенстве распределений;
- После выполнения N итераций подсчитывается мощность каждого из тестов против заданной альтернативы. Примеры результатов работы программы приведены в разделе 3.2 и 3.3.

3.2. Результаты работы программы. Таблицы

Следующие таблицы представляют результаты работы программы для нескольких случаев. При моделировании ситуации, соответствующей нулевой гипотезе (для обоснования применимости теста), в каждом случае была получена вероятность ошибки первого рода приблизительно равная выбранному уровню значимости $\alpha = 0.05$, что позволяет далее анализировать мощность теста.

Для моделирования были выбраны распределение Коши как популярное, но неудобное при проверке гипотез об однородности — распределение с тяжелыми хвостами, нормальное распределение как наиболее классическое и некоторые промежуточные варианты — распределение Лапласа и логистическое распределение.

Мощность нового критерия сравнивалась с мощностью нескольких широко используемых и хорошо себя зарекомендовавших критериев — критерия Андерсона – Дарлинга, Колмогорова – Смирнова и Вилкоксона – Манна – Уитни.

Приведем краткое описание этих методов. Критерий Колмогорова – Смирнова был предложен в работе [6]. Этот критерий подходит для случая непрерывных распределений, а его статистика измеряет расстояние между эмпирическими функциями распределения G_m и F_n , которые строятся по выборкам,

$$D_{m,n} = \sup_x |G_m(x) - F_n(x)|.$$

Критерий Андерсона – Дарлинга описан в работах [7, 8]; его статистика имеет вид

$$A^2 = \frac{mn}{m+n} \int_{-\infty}^{+\infty} \frac{(G_m(x) - F_n(x))^2}{(1 - H_{m+n}(x))H_{m+n}(x)},$$

где $H_{m+n}(x) = \frac{nF_n(x) + mG_m(x)}{n+m}$. Критерий в силу знаменателя $(1 - H_{m+n}(x))H_{m+n}(x)$ хорошо различает разницу на хвостах распределений и более предназначен для отслеживания различий в параметре масштаба, чем критерий Колмогорова – Смирнова.

Критерий Уилкоксона – Манна – Уитни [9] преимущественно используется для определения различий в параметре сдвига. Для построения статистики критерия независимые выборки объединяют в одну объема $n + m$ и упорядочивают, и по упорядоченной выборке определяют сумму рангов первой и второй выборки — R_n и R_m . Далее определяют

$$U_n = nm + n(n+1)/2 - R_n,$$

$$U_m = nm + m(m+1)/2 - R_m,$$

$$U = \min(U_n, U_m).$$

Наконец, U-статистику нормализуют, и при проверке гипотез используют

$$z = \frac{U - nm/2}{\sqrt{nm(n+m+1)/12}}.$$

Приведенные ниже таблицы позволяют сравнить мощность нового теста с мощностью других тестов. В таблицах показаны результаты моделирования при объеме выборок $n = 100$, числе итераций $N = 1000$ и количестве перестановок $K = 700$.

Таблицы 3.1 и 3.2 демонстрирует значения мощности тестов в случае соответственно распределений Коши и нормальных распределений, отличающихся параметром масштаба. В данном случае видно, что значения мощности нового теста во всех случаях превосходят значения мощности других тестов, особенно заметна разница в случае распределения Коши.

Таблица 3.1. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $Cauchy(0, 1)$ и $Cauchy(0, 1 + h/\sqrt{n})$

h	Power(T_n)	Power(K-S test)	Power(A-D test)	Power(W-M-W test)
0	0.052	0.045	0.048	0.052
2	0.104	0.059	0.066	0.054
4	0.208	0.098	0.124	0.06
5	0.403	0.115	0.146	0.061
6	0.394	0.159	0.232	0.065
8	0.692	0.261	0.367	0.071
10	0.899	0.365	0.528	0.064

Таблица 3.2. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $N(0, 1)$ и $N(0, 1 + h/\sqrt{n})$

h	Power(T_n)	Power(K-S test)	Power(A-D test)	Power(W-M-W test)
0	0.051	0.039	0.048	0.05
4	0.461	0.174	0.399	0.06
6	0.803	0.354	0.77	0.062
8	0.983	0.609	0.94	0.051
10	0.999	0.8	0.994	0.066

В случае распределений Коши, отличающихся сдвигом (Таблица 3.3), можно сравнить полученную численно мощность с мощностью, получаемой по формуле в теореме

1, где в качестве \bar{b}/a использованы оценки, полученные в разделе 2.2.3. В случае отличия только в сдвиге новый тест показывает близкие результаты к другим тестам, однако все же более слабые.

Таблица 3.3. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $Cauchy(0, 1)$ и $Cauchy(h/\sqrt{n}, 1)$

h	Power(T_n)	Power by formula	Power(K-S test)	Power(A-D test)	Power(W-M-W test)
0	0.045	0.05	0.05	0.049	0.052
2	0.084	0.1040979	0.101	0.118	0.108
4	0.274	0.2733904	0.335	0.384	0.342
6	0.544	0.5295416	0.694	0.623	0.639
8	0.81	0.7739868	0.928	0.888	0.851
10	0.941	0.9236467	0.98	0.972	0.953

Отличие в параметрах и сдвига, и масштаба должно быть «комбинацией» результатов значений для случая отличия в одном из параметров, однако для наглядности проиллюстрируем этот случай тоже (см. таблицу 3.4). В этом случае новый тест показывает наиболее мощные результаты, однако близок к результатам критерия Андерсона – Дарлинга.

Также полезно проверить мощность теста на разных но похожих распределениях с одинаковыми параметрами сдвига и масштаба. Известно, что нормальное и логистическое распределения с одинаковыми параметрами близки [10], поэтому рассмотрим теперь ситуацию, когда одна выборка соответствует стандартному нормальному закону, а другая — логистическому. Результаты представлены в таблице 3.5. В данном случае продемонстрированы результаты для разных уровней значимости, поскольку параметр h здесь не используется. На выборках из таких распределений новый тест показывает значительно более хорошие результаты, чем тест Колмогорова – Смирнова, и более хорошие, чем критерий Андерсона – Дарлинга.

На самом деле, еще хуже тесты различают логистический закон и распределение Коши (с одинаковыми параметрами сдвига и масштаба). В этом случае новый тест показывает принципиально более хорошие результаты, чем другие (таблица 3.6), даже

Таблица 3.4. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $Cauchy(0, 1)$ и $Cauchy(h/\sqrt{n}, 1 + h/\sqrt{n})$

h	Power(T_n)	Power(K-S test)	Power(A-D test)	Power(W-M-W test)
0	0.049	0.047	0.05	0.052
2	0.137	0.105	0.133	0.119
4	0.381	0.363	0.355	0.25
6	0.673	0.638	0.647	0.42
8	0.86	0.816	0.847	0.588
10	0.954	0.924	0.943	0.718

Таблица 3.5. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$. Для моделирования соответствующей H_1 ситуации использовались распределения $N(0, 1)$ и $Logistic(0, 1)$

α	Power(T_n)	Power(K-S test)	Power(A-D test)
0.01	0.71	0.118	0.461
0.02	0.802	0.25	0.645
0.05	0.901	0.365	0.843
0.1	0.956	0.623	0.954

в сравнении с критерием Андерсона – Дарлинга его мощность заметно выше.

Таблица 3.6. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$. Для моделирования соответствующей H_1 ситуации использовались распределения $Cauchy(0, 1)$ и $Logistic(0, 1)$

α	Power(T_n)	Power(K-S test)	Power(A-D test)
0.01	0.137	0.014	0.015
0.02	0.189	0.03	0.039
0.05	0.363	0.058	0.115
0.1	0.497	0.114	0.23

Также для наглядности «испортим» немного логистический закон с параметрами

0, 1 и будем рассматривать выборки из $N(0, 1)$ и $Logistic(0, 1 + h/\sqrt{n})$ (таблица 3.7). Здесь важно отметить, что случай $h = 0$ больше не соответствует нулевой гипотезе, как было в предыдущих таблицах, поскольку в этом случае мы рассматриваем выборки из разных распределений (хотя с одинаковыми параметрами). В этом случае новый тест также более эффективен.

Таблица 3.7. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $Cauchy(0, 1)$ и $Logistic(0, 1 + h/\sqrt{n})$

h	Power(T_n)	Power(K-S test)	Power(A-D test)
0	0.31	0.062	0.101
2	0.29	0.099	0.117
4	0.426	0.182	0.181
6	0.586	0.308	0.333
8	0.766	0.458	0.528
10	0.889	0.559	0.687

Поскольку тест показал особенно хорошие результаты на распределении Коши, которое обладает тяжелыми хвостами, рассмотрим также случай другого распределения с таким же свойством — распределения Лапласа (таблица 3.9). Новый тест в этом случае также показывает более хорошие результаты, в сравнении с тестом Колмогорова – Смирнова мощность нового теста больше более чем на 0.4, в случае теста Андерсона – Дарлинга — на 0.1.

Таблица 3.8. Число итераций $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, уровень значимости $\alpha = 0.05$. Для моделирования соответствующей H_1 ситуации использовались распределения $Laplace(0, 1)$ и $Laplace(0, 1 + h/\sqrt{n})$

h	Power(T_n)	Power(K-S test)	Power(A-D test)
0	0.05	0.045	0.05
2	0.103	0.058	0.078
4	0.326	0.092	0.212
6	0.624	0.215	0.454
8	0.81	0.335	0.69
10	0.954	0.497	0.876

3.3. Графики

Результаты численных исследований нового критерия могут быть представлены в более наглядном виде с помощью графиков. График на рис. 3.1 показывает кривые зависимости мощности тестов от параметра h (для h от 0 до 16 с шагом 1). То, что новый тест работает лучше других в случае распределений Коши, отличающихся параметром масштаба, видно из того, что кривая мощности нового теста расположена выше, чем кривые мощностей тестов Колмогорова – Смирнова и Андерсона – Дарлингга.

Также было проведено исследование теста при фиксированном параметре h и при разных объемах выборки. На графике на рис. 3.2 показаны значения мощности трех тестов для объемов выборки $n = 10, 50, 100, 200, 500$. В данном случае также видно превосходство нового теста.

Таблица 3.9. Сравнение мощностей в зависимости от объема выборок. $N = 1000$, параметр $h = 5$, число перестановок $K = 700$, $\alpha = 0.05$. $H_1: X \sim Cauchy(0, 1)$ и $Y \sim Cauchy(0, 1 + h/\sqrt{n})$

n	Power(T_n)	Power(K-S test)	Power(A-D test)
10	0.205	0.03	0.09
50	0.227	0.112	0.136
100	0.271	0.121	0.146
200	0.359	0.15	0.18
500	0.403	0.186	0.223

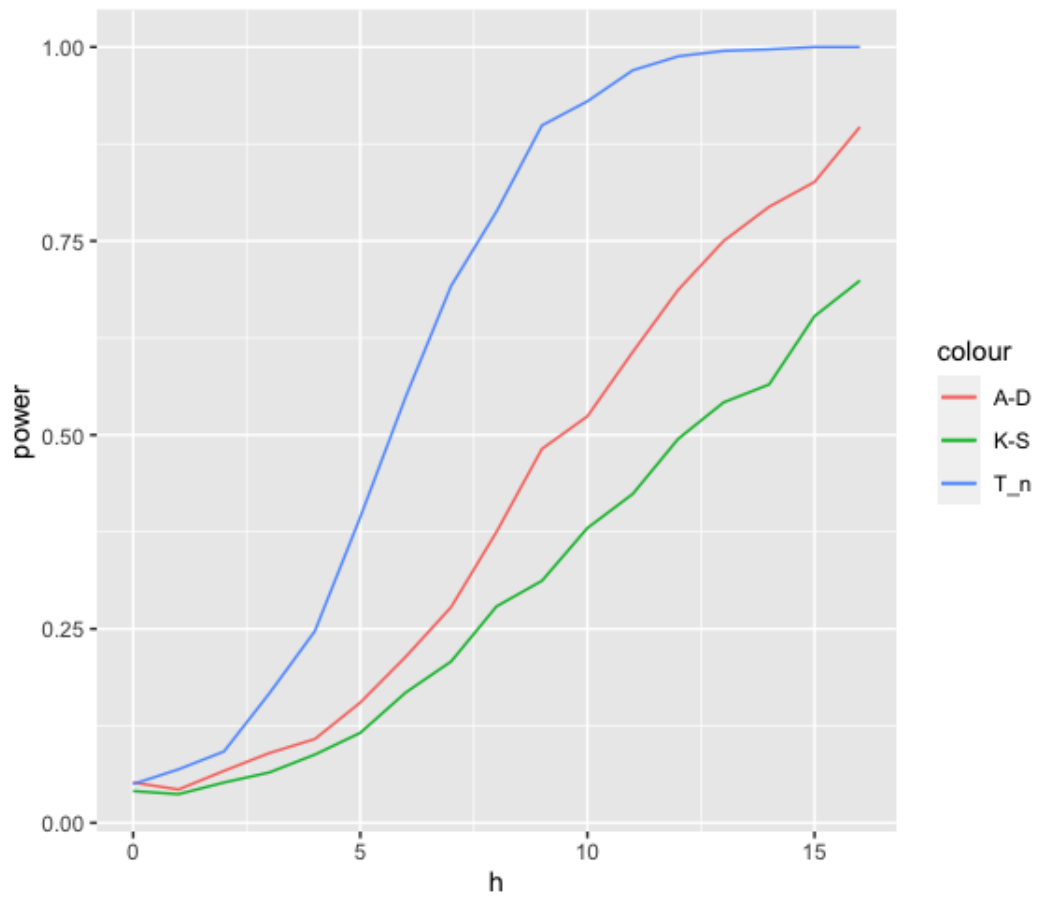


Рис. 3.1. Сравнение мощностей в зависимости от параметра h . $N = 1000$, объем выборок $n = 100$, число перестановок $K = 700$, $\alpha = 0.05$. $H_1: X \sim Cauchy(0, 1)$ и $Y \sim Cauchy(0, 1 + h/\sqrt{n})$

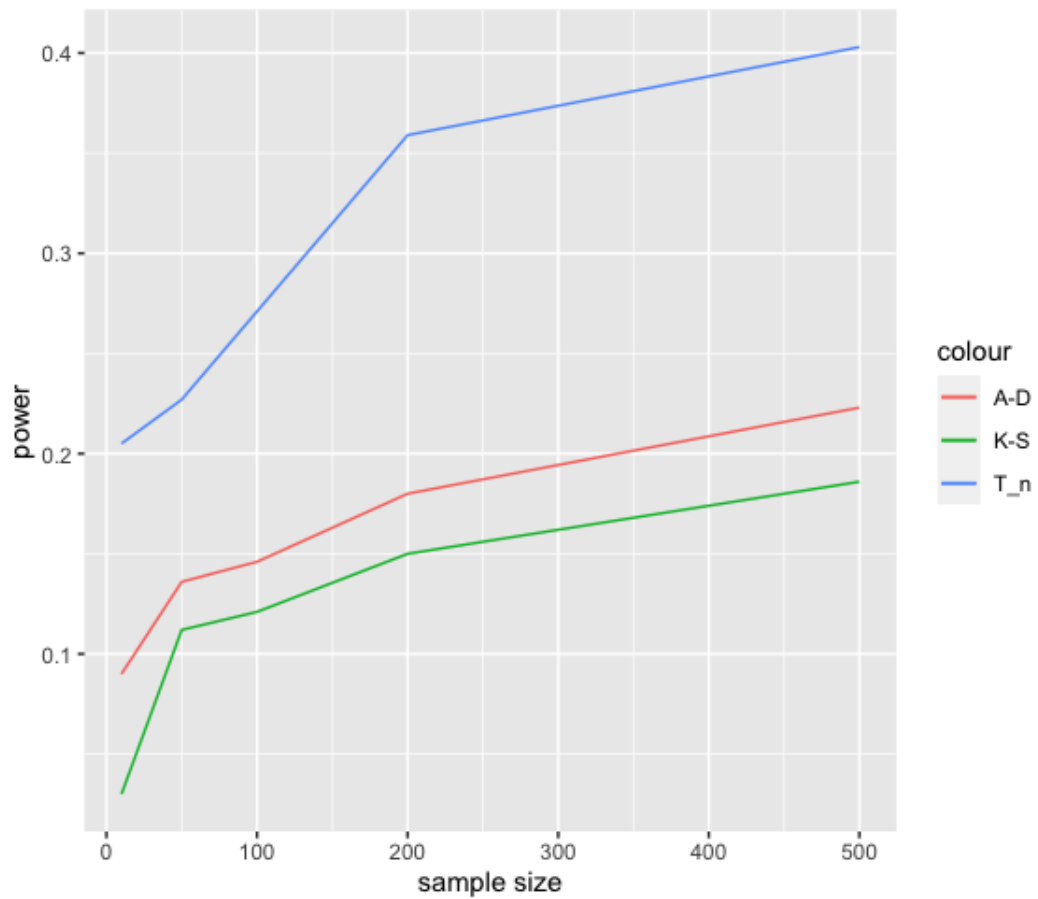


Рис. 3.2. Сравнение мощностей в зависимости от объема выборок. $N = 1000$, параметр $h = 5$, число перестановок $K = 700$, $\alpha = 0.05$. $H_1: X \sim Cauchy(0, 1)$ и $Y \sim Cauchy(0, 1 + h/\sqrt{n})$

Заключение

В результате работы были изучены перестановочные методы нахождения критической и доверительной области значений статистики критерия и статьи, посвященные новому тесту. Получены оценки параметра \bar{b}/a для случаев нормального распределения и распределения Коши и реализована программа на R, позволяющая эмпирически проверить аналитические результаты исследования мощности нового теста, собраны результаты численных исследований теста. В результате численных исследований сделан вывод о том, что наиболее эффективен тест в отношении распределений, отличающихся параметром масштаба, а также в отношении распределений с тяжелыми хвостами. Численные результаты подтверждают высокую эффективность метода, обнаружено, что в рассмотренных примерах он либо является значительно более мощным, чем альтернативные критерии, либо (только в случае отличия в сдвиге) показывает чуть меньшие, но близкие к другим тестам результаты, что представляет значительный практический и теоретический интерес.

Список литературы

1. Melas V. Salnikov D. New test for equality of two disributions // The 5th International Conference on Stochastic Methods (ICSM-5): Proceedings of the international scientific conference. — 2020. — P. 125–129.
2. Aslan B., Zech G. New test for the multivariate two-sample problem based on the concept of minimum energy // Journal of Statistical Computation and Simulation. — 2005. — Vol. 75, no. 2. — P. 109–119.
3. Efron B., Tibshirani R.J. An Introduction to the Bootstrap. — Springer Science+Business Media Dordrecht, 1993. — P. 202–218.
4. Janson S. Renewal theory for asymmetric U-statistics // Electron. J. Probab. — 2018. — P. 1–27.
5. Hoeffding W. A class of statistics with asymptotically normal distribution // Ann. Math. Statistics. — 1948. — P. 293–325.
6. Смирнов Н. В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках // Бюллетень МГУ, Серия А. — 1939. — Т. 2, № 2. — С. 3–14.
7. Pettitt A. N. A Two-Sample Anderson–Darling Rank Statistic // Biometrika. — 1976. — Vol. 63, no. 1. — P. 161–168.
8. Anderson T.W. Anderson–Darling Tests of Goodness-of-Fit // International Encyclopedia of Statistical Science. — 2011.
9. Wilcoxon F. Individual Comparisons by Ranking Methods // Biometrics Bulletin. — 1945. — Vol. 1, no. 6. — P. 80–83.
10. Лемешко Б. Ю. Критерии проверки гипотез об однородности. Руководство по применению: монография. — М.: ИНФРА-М., 2017. — 207 с.