

Санкт–Петербургский государственный университет  
Прикладная математика и информатика  
Вычислительная стохастика и статистические модели

Сальников Дмитрий Игоревич

СРАВНЕНИЕ ПАРАМЕТРИЧЕСКИХ И НЕПАРАМЕТРИЧЕСКИХ ТЕСТОВ С  
ПОМОЩЬЮ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Бакалаврская работа

Научный руководитель:  
д. ф.-м. н., профессор В. Б. Мелас

Рецензент:  
к. ф.-м. н., доцент П. В. Шпилев

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

Salnikov Dmitry Igorevich

COMPARISON OF PARAMETRIC AND NONPARAMETRIC TESTS WITH THE  
AID OF STOCHASTIC SIMULATION

Bachelor's Thesis

Scientific Supervisor:  
Professor V. B. Melas

Reviewer:  
Associate Professor P. V. Shpilev

Saint Petersburg  
2016

# Оглавление

<b>Введение</b> . . . . .	4
<b>Глава 1. Постановка задачи</b> . . . . .	5
1.1. Исследование эквивалентности перестановочных критериев . . . . .	5
1.2. Постановка задачи сравнения методов проверки гипотез с помощью сто- хастического моделирования . . . . .	7
<b>Глава 2. Анализ численных результатов</b> . . . . .	9
2.1. Нормальное распределение $N(\mu, \sigma)$ . . . . .	9
2.2. Распределение Коши $(x_0, \gamma)$ . . . . .	9
2.3. Распределение Стьюдента $t(n, x_0)$ . . . . .	10
2.4. Распределение Фишера $F(d_1, d_2)$ . . . . .	11
2.5. Бета-распределение $B(\alpha, \beta)$ . . . . .	12
2.6. Гамма-распределение $G(k, \theta)$ . . . . .	13
2.7. Равномерное распределение $U(a, b)$ . . . . .	13
2.8. Распределение Вейбулла $W(k, \lambda)$ . . . . .	14
2.9. Резюме . . . . .	14
<b>Заключение</b> . . . . .	16
<b>Список литературы</b> . . . . .	17
<b>Приложение А. Табличные результаты мощности тестов</b> . . . . .	18
<b>Приложение Б. Графическое представление мощности тестов</b> . . . . .	25
<b>Приложение В. Реализация алгоритма</b> . . . . .	32

## Введение

Задача сравнения двух распределений является классической задачей математической статистики и имеет большой теоретический и практический интерес. Хорошо известно (см., например, [1]), что в случае, когда оба распределения являются нормальными и имеют одинаковые дисперсии, классический тест Стьюдента ( $t$ -тест) обладает рядом оптимальных свойств, но на практике распределения часто таковыми не являются и, вообще говоря, неизвестны. При этом сильную конкуренцию  $t$ -тесту составляют непараметрические тесты, важным классом которых являются тесты, основанные на перестановках.

В работе представлены результаты исследования мощности четырех перестановочных тестов, а так же тестов Стьюдента, Колмогорова-Смирнова и Манна-Уитни. Исследования были проведены для широкого класса стандартных распределений, которые отличались либо отдельными параметрами, либо сдвигом плотности, либо утяжелением хвостов распределения.

Неперестановочные тесты были выбраны из следующих соображений: тест Колмогорова-Смирнова — наиболее универсальный непараметрический тест, основанный, в случае проверки однородности двух выборок, на сравнении эмпирических функций распределения. Тест Манна-Уитни — непараметрический ранговый тест, позволяющий выявлять различия распределений даже для малых объемов выборок и, по сообщениям стандартных руководств, — наиболее мощный непараметрический тест в случае распределений, отличающихся только сдвигом.

Работа осуществлена в рамках темы СПбГУ 6.38.435.2015.

## Глава 1

## Постановка задачи

Рассмотрим уравнение регрессии

$$X_{ij}(t) = X_i(t) + \varepsilon_{ij}(t), \quad i = 1, 2, \quad j = 1, 2, \dots, n_i, \quad t = 1, 2, \dots, N, \quad (1.1)$$

где  $X_1(t)$  and  $X_2(t)$  — неизвестные функции и  $\varepsilon_{ij}(t)$ ,  $j = 1, \dots, n_i$  — независимые случайные величины, т.ч.

$$\mathbb{E}[\varepsilon_{ij}(t)] = 0, \quad i = 1, 2, \quad j = 1, \dots, n_i, \quad t = 1, \dots, N, \quad (1.2)$$

и где  $\varepsilon_{1j}(t)$  и  $\varepsilon_{2j}(t)$  — независимые, и  $\varepsilon_{ij}(t)$  и  $\varepsilon_{ij}(\bar{t})$  для  $t \neq \bar{t}$  могут быть как угодно зависимы.

Обозначим  $\Delta(t) = X_2(t) - X_1(t)$ . Рассмотрим следующие нулевую и альтернативную гипотезы:

$$H_0 : \sum_{t=1}^N \Delta^2(t)/N = 0, \quad (1.3)$$

$$H_1 : \sum_{t=1}^N \Delta^2(t)/N \neq 0, \quad (1.4)$$

то есть  $H_0 : F_1 = F_2$ , где  $F_1, F_2$  - функции распределения общего вида по результатам наблюдений

$$X_1 = \{X_{11}(t), \dots, X_{1n_1}(t)\}, \quad X_2 = \{X_{21}(t), \dots, X_{2n_2}(t)\}.$$

Для простоты обозначений и без потери общности предположим сбалансированность выборок, т.е.  $n_1 = n_2 = n$  (в случае несбалансированной выборки аргументы очень похожи).

В работе [2] для этой задачи были предложены критерии на основе норм  $L_1$  и  $L_2$ . В недавней работе научного руководителя ([3]) было показано, что три метода перестановок, основанные на норме  $L_2$ , эквивалентны между собой при достаточно слабых предположениях о модели. Дадим достаточно подробное изложение этих результатов в виду их важности для постановки задачи.

### 1.1. Исследование эквивалентности перестановочных критериев

Для  $t = 1, 2, \dots, N$  определим вектора

$$Z(t, \pi_0) = \{X_{11}(t), \dots, X_{1n}(t), X_{21}(t), \dots, X_{2n}(t)\}, \quad (1.5)$$

$$Z(t, \pi_k) = \{\tilde{X}_{11}(t), \dots, \tilde{X}_{1n}(t), \tilde{X}_{21}(t), \dots, \tilde{X}_{2n}(t)\}, \quad (1.6)$$

$$\begin{aligned}
\tilde{X}_{1i_l} &= X_{2j_l}, \quad l = 1, \dots, k, \\
\tilde{X}_{2i_l} &= X_{1j_l}, \quad l = 1, \dots, k, \\
\tilde{X}_{1j} &= X_{1j}, \quad j \neq i_1, \dots, i_k, \\
\tilde{X}_{2j} &= X_{2j}, \quad j \neq j_1, \dots, j_k,
\end{aligned} \tag{1.7}$$

где  $\pi_k = \pi_k(s)$ ,  $s = 1, 2, \dots, (C_n^k)^2$  – различные способы замены  $k$  элементов из первой половины на  $k$  элементов из второй половины. Обозначим через  $Z = Z(\pi_0)$  совокупность векторов (1.5) при  $t = 1, \dots, N$  и определим на этом множестве три критерия  $K_i = K_i(Z)$ ,  $i = 1, 2, 3$ ,

$$\begin{aligned}
K_1(Z) &= \sum_{t=1}^N \left( \tilde{X}_1(t) - \tilde{X}_2(t) \right)^2 / N, \\
K_2(Z) &= \sum_{i,j=1}^n \left( \sum_{t=1}^N \left( \tilde{X}_{1i}(t) - \tilde{X}_{2j}(t) \right)^2 / N \right) / n^2, \\
K_3(Z) &= nK_1(Z) / S^2(Z),
\end{aligned}$$

где

$$\begin{aligned}
S^2(Z) &= S_1^2(Z) + S_2^2(Z), \\
S_1^2(Z) &= \sum_{t=1}^N \left( \sum_{i=1}^n (X_{1i}(t) - \tilde{X}_1(t))^2 / n \right) / N, \\
S_2^2(Z) &= \sum_{t=1}^N \left( \sum_{i=1}^n (X_{2i}(t) - \tilde{X}_2(t))^2 / n \right) / N.
\end{aligned}$$

Мощность критерия  $K_1$  была изучена численными методами в работе [4]. Критерий  $K_2$  был введен в работе [2].  $K_3$  является естественным обобщением классического  $t$ -критерия и аналогичен критерию перестановок, который предложили в работах [5] и [6]. Для  $Z = Z(\pi)$ ,  $\pi = \pi_k(s)$ ,  $s = 1, \dots, (C_n^k)^2$ ,  $k = 1, 2, \dots, n$  функции  $K_1$ ,  $K_2$  и  $K_3$  определяются теми же формулами с заменой  $Z = Z(\pi_0)$  на  $Z = Z(\pi)$ . Следующий алгоритм называется *перестановочный  $K_i$ -тест*:

- пусть  $r_2 = \sum_{k=1}^{n-1} (C_n^k)^2$ , и пусть  $r_1$  — число перестановок  $\pi$ , для которых  $K_i(Z(\pi)) > K_i(Z(\pi_0))$ ;
- если  $\frac{r_1}{r_2} > \alpha$ , где  $\alpha$  заданный уровень значимости, то нулевая гипотеза не отвергается при заданном  $\alpha$ -уровне;
- если  $\frac{r_1}{r_2} \leq \alpha$ , то нулевая гипотеза отвергается в пользу альтернативной гипотезы  $H_1$ .

Следующая теорема устанавливает эквивалентность всех трех перестановочных  $K_i$ -критериев, так как каждый из них характеризуется одной и той же функцией мощности.

**Теорема 1.1** ([3]) *Для модели, заданной уравнениями (1.1)-(1.2), критерии перестановок  $K_1$ ,  $K_2$  и  $K_3$  для проверки нулевой гипотезы  $H_0$ , заданной формулой (1.3), против альтернативы  $H_1$ , заданной (1.4), эквивалентны для любой перестановки и для любого произвольно заданного уровня значимости  $\alpha$ .*

## 1.2. Постановка задачи сравнения методов проверки гипотез с помощью стохастического моделирования

Определим 4 статистики ( $\bar{X}_i$  — выборочное среднее,  $X_{imed}$  — медиана по  $i$ -той выборке)

$$\begin{aligned} K_1(Z) &= (\bar{X}_1 - \bar{X}_2)^2, \\ K_4 &= (X_{1med} - X_{2med})^2, \\ K_5 &= \left( \sum_{i=1}^n |X_{1i} - X_{1med}| - \sum_{i=1}^n |X_{2i} - X_{2med}| \right)^2, \\ K_6 &= \sum_{i,j=1}^n |X_{1i} - X_{2j}|. \end{aligned}$$

Под перестановочным  $K_i$ -тестом проверки гипотезы  $H_0$  будем понимать следующий алгоритм:

Пусть  $r_2 = \sum_{k=0}^n (C_n^k)^2$  и пусть  $r_1$  — число перестановок  $\pi_k$ , для которых  $K_i(Z(\pi_k)) > K_i(Z(\pi_0))$ , тогда, если  $\frac{r_1}{r_2} \geq \alpha$  для  $K_1$ ,  $K_4$ ,  $K_6$  и  $\frac{r_1}{r_2} \leq 1 - \alpha$  для  $K_5$ , где  $\alpha$  — заданный уровень значимости, то нулевая гипотеза не отвергается.

В качестве альтернатив будем рассматривать тест Стьюдента (t.test), тест Колмогорова-Смирнова (ks.test) и тест Манна-Уитни (wilcox.test). Тест Стьюдента рассматривается как обладающий оптимальными свойствами при сравнении нормальных распределений с одинаковыми дисперсиями. Тест Колмогорова-Смирнова — непараметрический тест, основанный на выборочной функции распределения и, поэтому, наиболее универсальный из возможных тестов. Тест Манна-Уитни — непараметрический тест, основанный на рангах, который особенно эффективен в случае распределений, отличающихся только сдвигом.

Введем следующее утверждение, взятое из готовящейся к публикации статьи научного руководителя, устанавливающее значение мощности тестов  $K_1$  и t.test для симметричных распределений с общим центром:

**Предложение 1.1** ([7]) *Для любых симметричных относительно одного и того же центра функций распределения  $F_1$ ,  $F_2$  для проверки нулевой гипотезы  $H_0$ , заданной формулой (1.3), против альтернативы  $H_1$ , заданной (1.4), мощность теста  $K_1$ , а также  $t$ -теста совпадает с уровнем значимости для любого произвольно заданного уровня значимости  $\alpha$ .*

Задача заключается в сравнительном анализе мощности рассматриваемых тестов для следующих распределений:

- Нормальное распределение  $N(\mu, \sigma)$
- Распределение Коши  $C(x_0, \gamma)$
- Распределение Стьюдента  $t(n, x_0)$
- Распределение Фишера  $F(d_1, d_2)$
- Бета-распределение  $B(\alpha, \beta)$
- Гамма-распределение  $\Gamma(k, \theta)$
- Равномерное распределение  $U(a, b)$
- Распределение Вейбулла  $W(k, \lambda)$

Всего возможно  $\sum_{k=0}^n (C_n^k)^2$  вариантов замены элементов одной выборки на элементы другой. Такое число перестановок затруднительно обработать, поэтому мы возьмем лишь 1600 случайных вариантов. В работе [8] обнаружено, что мощность перестановочных тестов при таком количестве перестановок близка к мощности тестов, использующих все перестановки.

В каждом эксперименте будем проводить по  $N = 1000$  испытаний с заданным уровнем значимости  $\alpha$  для выборок объемом  $n = \{10, 30, 100\}$ . Стандартная ошибка среднего в  $N$  испытаниях Бернулли не превосходит  $0.5/\sqrt{N}$ , в нашем случае  $0.5/\sqrt{1000} = 0.0159$ . Данная точность позволяет в полной мере оценить мощность тестов.

Введем следующее утверждение, оно понадобится нам для сравнения тестов, мощности которых достаточно близки:

**Предложение 1.2.** Пусть в  $N$  экспериментах получено среднее значение мощность  $p_1$  по первому тесту и  $p_2$  по второму. Тогда, если  $N$  достаточно велико и

$$p_2 > p_1 + 3 \frac{p_1(1-p_1)}{\sqrt{N}},$$

то с вероятностью более чем 0.99% второй тест является более мощным, чем первый.

Доказательство следует из предельной теоремы для последовательности испытаний Бернулли. На практике это предложение оказывается верным уже при  $N \geq 100$ .

## Глава 2

### Анализ численных результатов

По умолчанию будем предполагать объем выборки равным 30. При объемах выборки 10 и 100 иерархия мощностей тестов сохраняется, однако сама мощность меняется. В рассмотренных ниже экспериментах (графические результаты которых можно увидеть на рис. Б.0.1–Б.0.23 в Приложении Б) параметры подобраны так, что в точке наибольшего различия распределений мощность хотя бы одного теста была близка к единице.

#### 2.1. Нормальное распределение $N(\mu, \sigma)$

Рассмотрим изменение параметра сдвига  $\mu$  от 0 до 1 с шагом 0.2 (рис. Б.0.1), который также является теоретическим средним распределения. В этом случае следует ожидать, что  $t.test$  и его перестановочный аналог - тест  $K_1$ , будут оптимальны. Этот вывод подтвердился численным экспериментом, однако  $K_6$  не уступает им в мощности (согласно предложению 1.2, нельзя с 99% точностью утверждать, что  $K_6$  обладает меньшей мощностью, чем  $K_1$  и  $t.test$ ). Наименьшую мощность показал  $ks.test$  с разницей в одну или две десятых. Промежуточное место занимают тесты  $K_4$ ,  $K_5$  и  $wilcox.test$ , причем при значениях параметра 0.8 и 1  $wilcox.test$  мощнее  $K_4$  и  $K_5$  с 99% вероятностью.

Рассмотрим изменение параметра масштаба  $\sigma$  от 1 до 5 (рис. Б.0.2), являющегося теоретическим стандартным отклонением. В данном случае математическое ожидание не меняется и, согласно предложению 1.1,  $t.test$  и  $K_1$  бесполезны. Численный эксперимент показал, что  $wilcox.test$  также бесполезен, а тесты  $K_4$  и  $K_5$  обладают низкой мощностью. Эффективными в проверке данной гипотезы являются тесты  $K_6$  и  $ks.test$ , однако  $ks.test$  значительно (вплоть до 0.4) уступает в мощности  $K_6$ .

Теперь загрязним нормальное распределение распределением Коши (рис. Б.0.3). В результате загрязнения распределение будет выглядеть как смесь 95% нормального распределения и 5% распределения Коши. Повторим первый эксперимент с изменением первого параметра на загрязненном распределении. В результате мощности  $t.test$  и  $K_1$  снизились в сравнении с остальными тестами. Лидирующими стали  $K_6$  и  $wilcox.test$ , с 99% вероятностью они являются наиболее мощными для данных распределений. Таким образом, при небольшом загрязнении нормального распределения  $t.test$  уже не является оптимальным.

#### 2.2. Распределение Коши $(x_0, \gamma)$

Рассмотрим двухпараметрическое распределение Коши с параметром сдвига  $x_0$  и параметром масштаба  $\gamma$ . У распределения Коши не существует математического ожидания, а плотность симметрична относительно прямой  $x = x_0$ . Распределение Коши

бесконечно делимо, выборочное среднее выборки из такого распределения будет иметь распределение Коши с теми же параметрами. Как следствие, можно ожидать маломощность тестов  $K_1$  и  $t.test$ .

Рассмотрим изменение параметра сдвига(рис. Б.0.4). Согласно предложениям 1 и 3 работы [9], выборочная медиана распределения Коши сходится к нормальному распределению с центром  $x_0$  и стандартным отклонением  $\frac{\pi\gamma}{2\sqrt{n}}$ . Соответственно, при изменении параметра сдвига высокую мощность должен показывать тест  $K_4$ , сравнивающий медианы. В результате численного эксперимента установлено, что тесты  $K_4$  и  $K_5$  являются наиболее мощными, а  $K_1$  и  $t.test$  практически бесполезны. После  $K_4$  и  $K_5$  наиболее мощным является  $ks.test$ , за ним  $wilcox.test$  и, наконец,  $K_6$ . Таким образом, за исключением тестов, сравнивающих средние( $t.test$  и  $K_1$ ),  $K_6$  является наименее мощным. Эти выводы подтверждены с помощью предложения 1.2 сравнением 99% доверительных интервалов оценок мощностей.

Рассмотрим изменение параметра масштаба(рис. Б.0.5). Этот параметр отвечает за высоту пика плотности распределения. При  $\gamma = 0$  распределение вырождается в распределение Дирака, сосредоточенное в точке  $x_0$ . В качестве  $F_1$  возьмем распределение Коши с параметрами  $C(0, 1)$ , в качестве  $F_2 - C(0, shift)$ , где  $shift = \{1, 3, 5, 7, 9\}$ . Можно предположить, что в этом случае бесполезными будут тесты  $K_1$  и Стьюдента, а так же  $K_4$ , так как  $x_0$  не меняется. Численный эксперимент подтвердил бесполезность  $K_1$  и  $t.test$ , а тесты  $K_4$  и  $K_5$  показали хоть и невысокую мощность, но не были бесполезны. Самым мощным оказался  $K_6$ , за ним расположился  $ks.test$ , довольно сильно уступая ему в мощности.  $wilcox.test$  полностью бесполезен.

### 2.3. Распределение Стьюдента $t(n, x_0)$

Рассмотрим двухпараметрическое нецентральное распределение Стьюдента с числом степеней свободы  $n$  и параметром сдвига  $x_0$ . У этого распределения существуют моменты порядка  $k < n$ , где  $n$  – число степеней свободы. Плотность симметрична относительно прямой  $x = x_0$ .

Рассмотрим изменение параметра сдвига(рис. Б.0.6). Выборочное среднее является состоятельной оценкой  $x_0$ . Будем рассматривать распределение с  $n = 3$  степенями свободы, чтобы существовали первые два момента.  $x_0$  будем менять от 0 до 2 с шагом 0.5. Численный эксперимент показал, что с 99% вероятностью  $K_6$  и  $wilcox.test$  являются наиболее мощными, а  $ks.test$  – наименее мощным, однако разница мощности составляет всего 5-7%.

Рассмотрим изменение количества степеней свободы(рис. Б.0.7). При  $n \rightarrow \infty$  распределение слабо сходится к стандартному нормальному распределению, уже при  $n = 30$  его можно приближенно считать нормальным. При  $n = 1$  распределение совпадает со стандартным распределением Коши. Будем сравнивать  $t(1, 0)$  с  $t(shift, 0)$ ,  $shift = \{1, 2, 3, 4, 5, 6\}$ , так как при маленьких степенях свободы распределение различается луч-

ше всего. Будем брать выборки из ста элементов для общего увеличения мощности. В результате только  $K_6$  различает распределения, остальные тесты бесполезны. При объеме выборки 30 при сравнении распределений  $t(1, 0)$  и предельного при  $n \rightarrow \infty N(0, 1)$  мощность  $K_6$  равна приблизительно 0.5.

## 2.4. Распределение Фишера $F(d_1, d_2)$

Случайная величина, распределенная по Фишеру, определяется как отношение двух случайных величин, имеющих распределение хи-квадрат:

$$Y_i \sim \chi^2(d_i), i = 1, 2,$$

тогда

$$F = \frac{Y_1/d_1}{Y_2/d_2} \sim F(d_1, d_2).$$

$$E = \frac{d_2}{d_2 - 2}, \text{ если } d_2 > 2,$$

$$D = \frac{2d_2^2(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}, \text{ если } d_2 > 4,$$

Формально, математическое ожидание может быть любым числом, большим единицы, однако дисперсия конечна только при математическом ожидании от 1 до 2. Если устремить оба параметра к  $\infty$ , то распределение сосредоточится в точке  $x = 1$ . В численных экспериментах будем брать объем выборки  $n = 100$ , так как плотность с изменением параметров меняется слабо.

Попробуем менять второй параметр  $d_2 = \{2, 4, 6, 8, 10\}$  при зафиксированном первом  $d_1 = 10$  (рис. Б.0.8). Самыми мощными оказались  $K_1$  и  $K_6$  (с 99% вероятностью они равномощны), остальные тесты значительно уступают этим в мощности. Среди остальных тестов наиболее мощными являются  $t.test$  при  $d_2 = \{3, 4, 5\}$  и  $ks.test$  при  $d_2 = \{6\}$  (с увеличением  $d_2$  математическое ожидание сходится к 1).

Теперь будем менять первый параметр, зафиксировав второй:  $F(2, 10)$  vs  $F(shift, 10)$ ,  $shift = \{2, 4, 6, 8, 10\}$  (рис. Б.0.9). В этом случае математическое ожидание не меняется и согласно предложению 1.1  $K_1$  и  $t.test$  бесполезны. Самым мощным оказался тест Колмогорова-Смирнова (с 99% вероятностью он самый мощный для данных распределений). Среди остальных тестов наиболее мощным является  $K_6$ , уступая  $ks.test$  примерно 15% мощности.  $wilcox.test$  при  $d_1 = \{4\}$  немного мощнее  $K_6$ , однако при больших значениях параметра сильно уступает ему.  $K_4$  и  $K_5$  показывают низкую мощность, но не бесполезны.

Зафиксируем дисперсию равной 2.82 и будем менять математическое ожидание от 1.1 до 1.5 с шагом 0.1 (рис. Б.0.10). Тогда параметры будут  $d_1 = \{1, 1.46, 2.3, 4.31, 15.7\}$ ,  $d_2 = \{22, 12, 8.67, 7, 6\}$ . На графике по оси абсцисс отмечена разница средних. Самыми мощными с 99% вероятностью оказались тесты Колмогорова-Смирнова и Манна-Уитни, причем, с той же вероятностью, при разнице средних меньше 0.2  $wilcox.test$  мощнее  $ks.test$ , а при разнице больше 2 — наоборот,  $ks.test$  мощнее  $wilcox.test$ . За ними расположились перестановочные тесты  $K_4$ ,  $K_5$  и  $K_6$ , причем с 99% вероятностью при разнице

средних 0.1 и 0.2  $K_4$  и  $K_5$  мощнее  $K_6$ , а при разнице средних 0.3 и 0.4 —  $K_6$  мощнее  $K_4$  и  $K_5$ . Тесты  $K_1$  и  $t.test$  показывают низкую мощность, видимо из-за того, что средние меняются слишком медленно (напомним, что в данном случае задача стояла изменять математическое ожидание не меняя дисперсию с учетом того, что дисперсия существует, а это возможно только если математическое ожидание лежит в интервале от 1 до 2).

## 2.5. Бета-распределение $B(\alpha, \beta)$

$$E = \frac{\alpha}{\alpha + \beta}$$

$$D = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

В силу того, что дисперсия уменьшается с ростом параметров, интереснее рассматривать небольшие их значения.

Рассмотрим бета-распределение с равными параметрами  $B(\alpha, \alpha)$  (рис. Б.0.11, Б.0.12). В этом случае распределение симметрично относительно математического ожидания, при любом значении  $\alpha$  математическое ожидание равно  $\frac{1}{2}$ , причем при  $\alpha < 1$  плотность выпукла, при  $\alpha > 1$  — вогнута, при  $\alpha = 1$  — совпадает с плотностью равномерного на  $(0, 1)$  распределения. Дисперсия по прямой  $\alpha = \beta$  равна  $\frac{1}{8\alpha + 4}$  и стремится к нулю при  $\alpha \rightarrow +\infty$ . При  $\alpha \rightarrow 0$  распределение сосредотачивается в двух точках: 0 и 1. Так как математическое ожидание не меняется при изменении параметров,  $t.test$  и  $K_1$  бесполезны. Сравним мощности оставшихся тестов на распределениях  $B(1, 1)$  vs  $B(shift, shift)$ , где  $shift = \{0.1, 0.2, 0.4, 0.6, 1, 2, 4, 6, 8, 10\}$ . В результате численного эксперимента наибольшую мощность показали тесты  $K_6$  и  $ks.test$ , причем при  $shift < 1$  тест Колмогорова-Смирнова мощнее  $K_6$ . Тесты  $K_4$  и  $K_5$  показали низкую мощность,  $wilcox.test$  бесполезен.

Теперь рассмотрим несимметричный случай. Плотности бета-распределений  $B(\alpha, \beta)$  и  $B(\beta, \alpha)$  симметричны относительно прямой  $x = 0.5$ , вследствие этого мощности тестов для симметричных параметров должны совпадать. Свойство плотности бета-распределения таково, что уменьшая первый параметр мы сдвигаем плотность ближе к нулю, уменьшая второй — к единице. Поэтому уменьшение первого параметра в каком-то смысле эквивалентно увеличению второго. В силу этого рассмотрим следующую модель: будем сравнивать  $B(1, 2)$  с  $B(1 + shift, 2)$  и с  $B(1, 2 + shift)$ ,  $shift > 0$ . Первый параметр будем менять от 1 до 3 с шагом 0.5, второй — от 2 до 6 с шагом 1, математического ожидания в этих случаях будут  $\{0.33, 0.43, 0.50, 0.56, 0.60\}$  и  $\{0.33, 0.25, 0.20, 0.17, 0.14\}$  соответственно.

При изменении первого параметра (рис. Б.0.13) самыми мощными оказались тесты  $K_6$ ,  $wilcox.test$ ,  $K_1$  и  $t.test$ , наименее мощным —  $ks.test$ , уступая примерно 0.05 мощности, что дает основания считать его менее мощными с 99% вероятностью.

При изменении второго параметра (рис. Б.0.14) тесты  $K_1$ ,  $t.test$  и  $K_6$  являются наиболее мощными,  $wilcox.test$  немного уступает им. Как и в предыдущем случае,  $ks.test$

является наименее мощным с 99% вероятностью.

## 2.6. Гамма-распределение $G(k, \theta)$

Рассмотрим двухпараметрическое гамма-распределение с математическое ожиданием  $k$  и дисперсией  $k\theta^2$ . При  $\theta = 1$  распределение совпадает с экспоненциальным.

Рассмотрим изменение первого параметра (рис. Б.0.15). Будем сравнивать  $G(1, 2)$  с  $G(\text{shift}, 2)$ ,  $\text{shift} = \{1, 1.25, 1.5, 1.75, 2\}$ . Эмпирически `wilcox.test` оказался мощнее  $K_6$  и, согласно предложению 1.2, можно с 99% вероятностью утверждать, что `wilcox.test` на самом деле мощнее  $K_6$ . Наименее мощным оказался `ks.test`, однако разница мощности `ks.test` и `wilcox.test` не превышает 0.15, остальные тесты равноможны.

При изменении второго параметра (рис. Б.0.16) наибольшую мощность показывают  $K_1$ ,  $K_6$  и `t.test`, за ними идет `wilcox.test`, наименее мощный - `ks.test`. Эти выводы подтверждаются сравнением 99% доверительных интервалов на основании предложения 1.2.

Теперь зафиксируем дисперсию равной 2 и будем менять математическое ожидание от 1 до 3, тогда  $k = \{0.5, 2, 4.5, 8, 12.5\}$ ,  $\theta = \{2, 1, 0.67, 0.5, 0.4\}$  (рис. Б.0.17). Наибольшей мощностью с 99% вероятностью обладает `wilcox.test`. Тесты  $K_4$ ,  $K_5$ ,  $K_6$  и `ks.test` примерно равноможны и немного уступают `wilcox.test` в мощности,  $K_1$  и `t.test` наименее мощные.

Наконец, зафиксируем математическое ожидание равным 2 и будем менять дисперсию от 1 до 5, тогда  $k = \{4, 2, 1.33, 1, 0.8\}$ ,  $\theta = \{0.5, 1, 1.5, 2, 2.5\}$  (рис. Б.0.18). Согласно предложению 1.1,  $K_1$  и `t.test` бесполезны. В этом эксперименте наибольшую мощность показывает  $K_6$ , `ks.test` уступает ему примерно на 7-8% (что дает основания считать  $K_6$  наиболее мощным с 99% вероятностью), остальные тесты показывают значительно меньшую мощность.

## 2.7. Равномерное распределение $U(a, b)$

Распределение имеет константную плотность на носителе  $(a, b)$ . В качестве  $F_1$  возьмем  $U(0, 1)$  и проведем 3 эксперимента: будем сдвигать носитель распределения вправо, расширять его в обе стороны и расширять только вправо.

При сдвиге (рис. Б.0.19) наибольшей мощностью обладают тесты  $K_1$  и `t.test`, немного уступают им в мощности `wilcox.test` и  $K_6$ , остальные тесты имеют значительно меньшую мощность. Можно с 99% вероятностью говорить, что  $K_1$  и `t.test` в этом эксперименте являются наиболее мощными и что `wilcox.test` мощнее  $K_6$ , хотя разница в мощности чрезвычайно мала.

При расширении носителя (меняем дисперсию сохраняя математическое ожидание, рис. Б.0.20) наибольшей мощностью обладает  $K_6$ , ему уступает `ks.test` в среднем на 0.4, остальные тесты практически бесполезны.

При увеличении правой границы интервала(рис. Б.0.21) наиболее мощными являются тесты  $K_6$ ,  $K_1$  и `t.test`(приближенно их можно считать равномогными). Им уступают `ks.test` и `wilcox.test`. Наименьшей мощностью обладают тесты  $K_4$  и  $K_5$ .

## 2.8. Распределение Вейбулла $W(k, \lambda)$

Рассмотрим двухпараметрическое распределение Вейбулла с коэффициентом формы  $k$  и коэффициентом масштаба  $\lambda$ . Функция распределения выглядит как  $F(x) = 1 - e^{-(x/\lambda)^k}$ , а математическое ожидание и дисперсия как  $E = \lambda\Gamma(1 + \frac{1}{k})$ ,  $D = \lambda^2\Gamma(1 + \frac{2}{k}) - E^2$ .

При изменении первого параметра от 1 до 5 при значении второго 3(рис. Б.0.22) наиболее мощным является тест  $K_6$ , `ks.test` значительно уступает ему в мощности,  $K_1$  и `t.test` бесполезны, остальные тесты имеют низкую мощность.

При изменении второго параметра от 1 до 3 с шагом 0.5 при значении первого параметра 1(рис. Б.0.23) наиболее мощными оказались тесты  $K_1$ ,  $K_6$  и `t.test`, это подтверждают 99% доверительные интервалы. Наименее мощным является `ks.test`, уступая в мощности примерно 2 сотых. Среди оставшихся тестов `wilcox.test` оказался мощнее  $K_4$  и  $K_5$ .

## 2.9. Резюме

Согласно полученным результатам, среди рассмотренных перестановочных тестов наибольшую мощность имеет тест  $K_6$ . При сдвиге плотности он не уступает тесту  $K_1$ , а тестам  $K_4$  и  $K_5$  уступает только при изменении параметра сдвига распределения Коши(рис. Б.0.4).

$K_6$  также мощнее непостоянных тестов за исключением ряда случаев. В тех же случаях, когда  $K_6$  уступает непостоянным тестам — при изменении параметра сдвига распределения Коши(рис. Б.0.4), при изменении первого параметра распределения Фишера и при изменении математического ожидания без изменения дисперсии распределения Фишера(рис. Б.0.9, Б.0.10), при одновременном уменьшении обоих параметров бета-распределения(рис. Б.0.11), при изменении первого параметра гамма-распределения и при изменении математического ожидания без изменения дисперсии гамма-распределения(рис. Б.0.15, Б.0.17) — разница в мощности не превышает две десятых. Особенно велико преимущество теста  $K_6$  в случае симметричных распределений с общим центром(рис. Б.0.2, Б.0.5, Б.0.7, Б.0.12, Б.0.18, Б.0.20).

Табличные результаты мощности тестов вы можете увидеть в Приложении А, туда включено большинство рассмотренных выше экспериментов. Они получены для выборок объемом  $n = \{10, 30, 100\}$  для лучшего среди перестановочных тестов —  $K_6$ , и непостоянных тестов.

Графические результаты представлены в приложении Б, они иллюстрируют все описанные в этой главе эксперименты. Параметр *shift* — фиктивная переменная, обо-

значающая изменение одного из параметров распределения (смотри заголовки над графиками). Если под осью абсцисс написано *mean difference* это значит, что по оси абсцисс отложено изменение математического ожидания при неизменной дисперсии, если написано *variance difference* — изменение дисперсии при неизменном математическом ожидании.

С помощью таблиц можно посмотреть на полученные точные значения мощности тестов, в то время как графические результаты направлены на сравнительный их анализ.

## Заключение

Стохастическое моделирование является универсальным методом исследования, который позволяет оценивать эффективность статистических процедур в случаях, когда это не удается сделать аналитическими методами.

В работе были реализованы алгоритм и программа для численного исследования мощности перестановочных тестов и классических тестов Стьюдента, Колмогорова-Смирнова и Манна-Уитни для решения задачи проверки гипотезы о равенстве двух распределений. Полученные результаты представлены в виде таблиц мощностей и с помощью графиков. Таблицы мощностей позволяют увидеть точные результаты моделирования, а графики — наглядно сравнить мощности тестов.

Опираясь на полученные численные результаты можно заключить, что тест, основанный на сумме модулей разностей элементов двух выборок, в большинстве случаев превосходит по мощности все другие рассмотренные тесты. Особенно велико преимущество этого теста, если центры сравниваемых распределений совпадают.

Среди неперестановочных тестов тест Колмогорова-Смирнова является наиболее универсальным, его разумно использовать в случае, когда нет никакой априорной информации о распределениях. В случае только сдвига плотности он уступает в мощности тесту Манна-Уитни в среднем 10% мощности, однако при отсутствии сдвига преимущество теста Колмогорова-Смирнова очень велико.

## Список литературы

1. Леман Э. Проверка статистических гипотез. — М. : Наука, 1979. — 408 с.
2. Sirsky M. On the Statistical Analysis of Functional Data Arasing from Designed Experiments : Ph. D. thesis / M. Sirsky ; University of Manitoba. — 2012.
3. New insights on permutation approach for hypothesis testing on functional data / L. Corain, V. Melas, A. Pepelyshev, L. Salmaso // Advances in Data Analysis and Classification. — 2013. — P. 1–18.
4. Statistical methods for comparative phenomics using high-throughput phenotype microarrays / J. Sturino, I. Zorych, B. Mallick et al. // The International Journal of Biostatistics. — 2010. — Vol. 6. — P. 3–4.
5. Cox D., Lee J. Pointwise testing with functional data using the Westfall-Young randomization method // Biometrika. — 2008. — Vol. 95. — P. 621–634.
6. Ramsay J., Hooker G., Graves S. Functional Data Analysis with R and Matlab. — 2009.
7. Мелас В.Б., Сальников Д.И., Гудулина А.О. Численное сравнение перестановочных и классических методов проверки статистических гипотез // Вестник СПбГУ, сер.1, вып.3. — готовится к публикации.
8. Keller-McNulty S., Higgins J. Effect of tail weight and outliers on power and type-I error of robust permutation tests for location // Communications in Statistics - Simulation and Computation. — 1987. — Vol. 16. — P. 17–35.
9. Некоторые оценки параметров распределения Коши / Е.Д. Галкина, С.В. Лещева, Н.С. Лукичев, В.Е. Рыков // Труды НГТУ им. Р.Е. Алексеева №3. — 2015.
10. Edgington E. S. Approximate randomization tests // The Journal of Psychology. — 1969. — Vol. 72. — P. 143–149.
11. Good P. I. Resampling Methods: A Practical Guide to Data Analysis. — 3 edition. — Birkhauser, 2006.

## Приложение А

## Табличные результаты мощности тестов

Таблица А.1: Мощность тестов для выборок размера  $n = 10$ 

Distr	F1	F2	$K_6$	t.test	ks.test	wilcox.test
Normal	(0, 1)	(0, 1)	0.056	0.05	0.012	0.037
		(0.5, 1)	0.209	0.18	0.053	0.151
		(1, 1)	0.593	0.552	0.237	0.519
		(1.5, 1)	0.905	0.9	0.588	0.857
		(2, 1)	0.987	0.985	0.854	0.973
Normal	(0, 1)	(0, 1)	0.048	0.05	0.015	0.046
		(0, 2)	0.135	0.047	0.018	0.04
		(0, 3)	0.321	0.053	0.058	0.064
		(0, 4)	0.543	0.058	0.088	0.068
		(0, 5)	0.688	0.048	0.11	0.071
		(0, 6)	0.795	0.051	0.125	0.075
95% Normal 5% Cauchy	(0, 1)	(0, 1)	0.051	0.051	0.014	0.049
		(0.5, 1)	0.198	0.153	0.047	0.146
		(1, 1)	0.546	0.448	0.206	0.436
		(1.5, 1)	0.849	0.762	0.509	0.794
		(2, 1)	0.96	0.879	0.794	0.944
Cauchy	(0, 1)	(0, 1)	0.056	0.022	0.015	0.044
		(0.5, 1)	0.121	0.034	0.042	0.083
		(1, 1)	0.199	0.065	0.11	0.182
		(1.5, 1)	0.383	0.121	0.229	0.324
		(2, 1)	0.521	0.184	0.38	0.457
Cauchy	(0, 1)	(0, 1)	0.049	0.02	0.013	0.036
		(0, 2)	0.095	0.022	0.02	0.049
		(0, 3)	0.168	0.023	0.036	0.044
		(0, 4)	0.262	0.02	0.039	0.058
		(0, 5)	0.333	0.019	0.06	0.065
Beta	(2, 2)	(2, 2)	0.056	0.054	0.008	0.038
		(1.8, 2)	0.073	0.055	0.013	0.05
		(1.6, 2)	0.11	0.094	0.025	0.076
		(1.4, 2)	0.166	0.139	0.044	0.121
		(1.2, 2)	0.25	0.216	0.056	0.187
Beta	(2, 2)	(2, 2)	0.055	0.047	0.014	0.044
		(2, 1.8)	0.073	0.054	0.01	0.037

		(2, 1.6)	0.104	0.086	0.025	0.07
		(2, 1.4)	0.148	0.12	0.037	0.098
		(2, 1.2)	0.249	0.208	0.07	0.192
Gamma	(2, 2)	(2, 2)	0.055	0.038	0.011	0.039
		(1.8, 2)	0.069	0.051	0.023	0.047
		(1.6, 2)	0.119	0.093	0.026	0.076
		(1.4, 2)	0.19	0.146	0.052	0.153
		(1.2, 2)	0.35	0.273	0.115	0.292
Gamma	(2, 2)	(2, 2)	0.056	0.044	0.011	0.038
		(2, 1.8)	0.065	0.05	0.015	0.043
		(2, 1.6)	0.112	0.088	0.029	0.077
		(2, 1.4)	0.185	0.144	0.038	0.129
		(2, 1.2)	0.344	0.285	0.096	0.261
Fisher	(1, 4)	(1, 4)	0.05	0.02	0.014	0.041
		(2, 4)	0.09	0.027	0.029	0.068
		(3, 4)	0.111	0.042	0.044	0.108
		(4, 4)	0.128	0.056	0.068	0.152
		(5, 4)	0.12	0.041	0.05	0.135
		(6, 4)	0.138	0.056	0.081	0.167
Fisher	(1, 6)	(1, 6)	0.049	0.027	0.012	0.044
		(1, 5)	0.082	0.035	0.011	0.056
		(1, 4)	0.074	0.028	0.009	0.044
		(1, 3)	0.098	0.033	0.022	0.068
		(1, 2)	0.106	0.024	0.01	0.048
		(1, 1)	0.276	0.026	0.039	0.109
Fisher	(6, 6)	(6, 6)	0.053	0.031	0.017	0.038
		(5, 5)	0.068	0.036	0.02	0.044
		(4, 4)	0.072	0.034	0.016	0.043
		(3, 3)	0.098	0.032	0.016	0.04
		(2, 2)	0.132	0.03	0.021	0.043
		(1, 1)	0.336	0.024	0.046	0.048
Student	(1, 0)	(1, 0)	0.051	0.02	0.01	0.035
		(1, 0.2)	0.106	0.035	0.029	0.086
		(1, 0.4)	0.129	0.042	0.026	0.083
		(1, 0.6)	0.21	0.063	0.052	0.171
		(1, 0.8)	0.324	0.122	0.135	0.284
		(1, 1)	0.425	0.171	0.205	0.392
Student	(1, 0)	(1, 0)	0.047	0.029	0.012	0.044
		(2, 0)	0.098	0.03	0.014	0.049
		(3, 0)	0.114	0.029	0.008	0.039

		(4, 0)	0.078	0.022	0.012	0.036
		(5, 0)	0.114	0.028	0.017	0.052
		(6, 0)	0.116	0.03	0.017	0.058
Uniform	(0, 1)	(0, 0)	0.046	0.061	0.01	0.05
		(0.1, 0.1)	0.1	0.095	0.019	0.08
		(0.2, 0.2)	0.293	0.287	0.063	0.239
		(0.3, 0.3)	0.549	0.557	0.166	0.476
		(0.4, 0.4)	0.807	0.822	0.377	0.734
Weibull	(1, 3)	(1, 3)	0.048	0.038	0.01	0.029
		(1, 2.5)	0.074	0.047	0.013	0.043
		(1, 2)	0.162	0.126	0.04	0.103
		(1, 1.5)	0.336	0.241	0.089	0.227
		(1, 1)	0.627	0.507	0.23	0.485

Таблица А.2: Мощность тестов для выборок размера  $n = 30$ 

Distr	F1	F2	$K_6$	t.test	ks.test	wilcox.test
Normal	(0, 1)	(0, 1)	0.052	0.046	0.043	0.045
		(0.5, 1)	0.457	0.481	0.32	0.448
		(1, 1)	0.958	0.966	0.874	0.963
		(1.5, 1)	0.999	0.999	0.999	0.999
		(2, 1)	1	1	1	1
Normal	(0, 1)	(0, 1)	0.055	0.054	0.032	0.048
		(0, 1.5)	0.141	0.041	0.072	0.052
		(0, 2)	0.475	0.036	0.183	0.053
		(0, 2.5)	0.824	0.05	0.347	0.067
		(0, 3)	0.945	0.051	0.512	0.068
		(0, 3.5)	0.992	0.058	0.664	0.065
95% Normal 5% Cauchy	(0, 1)	(0, 1)	0.045	0.05	0.053	0.063
		(0.5, 1)	0.421	0.335	0.297	0.405
		(1, 1)	0.924	0.765	0.819	0.922
		(1.5, 1)	0.999	0.884	0.993	0.999
		(2, 1)	1	0.929	1	1
Cauchy	(0, 1)	(0, 1)	0.043	0.021	0.033	0.043
		(0.5, 1)	0.138	0.037	0.185	0.185
		(1, 1)	0.403	0.082	0.557	0.518
		(1.5, 1)	0.647	0.128	0.859	0.78
		(2, 1)	0.874	0.224	0.976	0.945
Cauchy	(0, 1)	(0, 1)	0.056	0.025	0.04	0.068
		(0, 2)	0.192	0.02	0.098	0.05

		(0, 3)	0.407	0.015	0.206	0.051
		(0, 4)	0.618	0.015	0.362	0.065
		(0, 5)	0.777	0.018	0.489	0.07
Beta	(2, 2)	(2, 2)	0.053	0.054	0.036	0.054
		(1.8, 2)	0.09	0.089	0.048	0.082
		(1.6, 2)	0.141	0.146	0.095	0.142
		(1.4, 2)	0.297	0.288	0.191	0.286
		(1.2, 2)	0.527	0.532	0.379	0.531
Beta	(2, 2)	(2, 2)	0.053	0.054	0.032	0.051
		(2, 1.8)	0.068	0.07	0.043	0.066
		(2, 1.6)	0.142	0.139	0.099	0.137
		(2, 1.4)	0.314	0.302	0.225	0.302
		(2, 1.2)	0.542	0.543	0.401	0.536
Gamma	(2, 2)	(2, 2)	0.052	0.051	0.033	0.046
		(1.8, 2)	0.106	0.089	0.063	0.108
		(1.6, 2)	0.228	0.215	0.162	0.247
		(1.4, 2)	0.459	0.42	0.331	0.49
		(1.2, 2)	0.754	0.705	0.631	0.779
Gamma	(2, 2)	(2, 2)	0.053	0.058	0.043	0.056
		(2, 1.8)	0.068	0.07	0.051	0.064
		(2, 1.6)	0.219	0.205	0.124	0.191
		(2, 1.4)	0.472	0.489	0.309	0.443
		(2, 1.2)	0.77	0.79	0.578	0.727
Fisher	(1, 4)	(1, 4)	0.05	0.026	0.029	0.042
		(2, 4)	0.085	0.037	0.141	0.178
		(3, 4)	0.157	0.052	0.31	0.322
		(4, 4)	0.173	0.051	0.399	0.401
		(5, 4)	0.212	0.054	0.49	0.444
		(6, 4)	0.252	0.069	0.577	0.485
Fisher	(1, 6)	(1, 6)	0.049	0.035	0.04	0.051
		(1, 5)	0.048	0.035	0.037	0.053
		(1, 4)	0.057	0.036	0.043	0.049
		(1, 3)	0.1	0.05	0.037	0.056
		(1, 2)	0.234	0.086	0.066	0.095
		(1, 1)	0.741	0.094	0.213	0.258
Fisher	(6, 6)	(6, 6)	0.057	0.043	0.029	0.049
		(5, 5)	0.066	0.042	0.032	0.055
		(4, 4)	0.092	0.054	0.045	0.051
		(3, 3)	0.177	0.089	0.073	0.057
		(2, 2)	0.464	0.133	0.142	0.047

		(1, 1)	0.902	0.091	0.466	0.062
Student	(1, 0)	(1, 0)	0.057	0.02	0.043	0.043
		(1, 0.2)	0.093	0.033	0.072	0.112
		(1, 0.4)	0.185	0.07	0.188	0.251
		(1, 0.6)	0.385	0.143	0.395	0.475
		(1, 0.8)	0.602	0.231	0.64	0.73
		(1, 1)	0.733	0.298	0.806	0.887
Student	(1, 0)	(1, 0)	0.046	0.011	0.03	0.037
		(2, 0)	0.124	0.023	0.039	0.04
		(3, 0)	0.238	0.023	0.049	0.06
		(4, 0)	0.284	0.019	0.044	0.042
		(5, 0)	0.323	0.014	0.047	0.044
		(6, 0)	0.333	0.016	0.057	0.051
Uniform	(0, 1)	(0, 0)	0.044	0.038	0.032	0.037
		(0.1, 0.1)	0.209	0.253	0.118	0.25
		(0.2, 0.2)	0.702	0.772	0.423	0.724
		(0.3, 0.3)	0.964	0.981	0.832	0.955
		(0.4, 0.4)	0.999	1	0.984	0.997
Weibull	(1, 3)	(1, 3)	0.053	0.043	0.027	0.045
		(1, 2.5)	0.103	0.093	0.054	0.089
		(1, 2)	0.335	0.321	0.177	0.274
		(1, 1.5)	0.703	0.707	0.464	0.608
		(1, 1)	0.987	0.984	0.881	0.95

Таблица А.3: Мощность тестов для выборок размера  $n = 100$ 

Distr	F1	F2	$K_6$	t.test	ks.test	wilcox.test
Normal	(0, 1)	(0, 1)	0.047	0.053	0.036	0.047
		(0.5, 1)	0.92	0.94	0.821	0.929
		(1, 1)	1	1	0.999	1
		(1.5, 1)	1	1	1	1
		(2, 1)	1	1	1	1
Normal	(0, 1)	(0, 1)	0.053	0.051	0.035	0.058
		(0, 1.3)	0.21	0.048	0.089	0.043
		(0, 1.6)	0.74	0.045	0.345	0.049
		(0, 1.9)	0.985	0.051	0.687	0.054
		(0, 2.2)	0.999	0.053	0.913	0.074
		(0, 2.5)	1	0.049	0.988	0.064
95% Normal	(0, 1)	(0, 1)	0.056	0.041	0.048	0.058
5% Cauchy		(0.5, 1)	0.889	0.63	0.802	0.897

		(1, 1)	1	0.853	1	1
		(1.5, 1)	0.999	0.902	1	1
		(2, 1)	1	0.936	1	1
Cauchy	(0, 1)	(0, 1)	0.052	0.03	0.035	0.054
		(0.5, 1)	0.264	0.033	0.53	0.485
		(1, 1)	0.789	0.072	0.982	0.958
		(1.5, 1)	0.988	0.13	1	0.997
		(2, 1)	0.995	0.207	1	1
Cauchy	(0, 1)	(0, 1)	0.044	0.025	0.045	0.049
		(0, 2)	0.495	0.015	0.348	0.043
		(0, 3)	0.898	0.026	0.855	0.058
		(0, 4)	0.974	0.016	0.983	0.069
		(0, 5)	0.987	0.016	0.998	0.05
Beta	(2, 2)	(2, 2)	0.055	0.051	0.034	0.048
		(1.8, 2)	0.126	0.135	0.079	0.125
		(1.6, 2)	0.401	0.409	0.273	0.4
		(1.4, 2)	0.746	0.772	0.621	0.772
		(1.2, 2)	0.969	0.967	0.922	0.968
Beta	(2, 2)	(2, 2)	0.056	0.052	0.042	0.054
		(2, 1.8)	0.12	0.132	0.081	0.122
		(2, 1.6)	0.371	0.391	0.267	0.373
		(2, 1.4)	0.745	0.756	0.638	0.754
		(2, 1.2)	0.967	0.962	0.918	0.964
Gamma	(2, 2)	(2, 2)	0.043	0.048	0.031	0.052
		(1.8, 2)	0.203	0.187	0.161	0.209
		(1.6, 2)	0.587	0.557	0.516	0.641
		(1.4, 2)	0.937	0.907	0.866	0.952
		(1.2, 2)	0.998	0.989	0.995	0.999
Gamma	(2, 2)	(2, 2)	0.048	0.044	0.024	0.045
		(2, 1.8)	0.178	0.194	0.107	0.174
		(2, 1.6)	0.587	0.635	0.389	0.556
		(2, 1.4)	0.921	0.94	0.816	0.907
		(2, 1.2)	0.996	0.998	0.984	0.994
Fisher	(1, 4)	(1, 4)	0.054	0.046	0.036	0.046
		(2, 4)	0.156	0.039	0.493	0.446
		(3, 4)	0.392	0.053	0.864	0.719
		(4, 4)	0.581	0.057	0.97	0.845
		(5, 4)	0.735	0.06	0.989	0.896
		(6, 4)	0.811	0.062	0.994	0.93
Fisher	(1, 6)	(1, 6)	0.053	0.048	0.032	0.045

		(1, 5)	0.062	0.05	0.037	0.05
		(1, 4)	0.086	0.079	0.052	0.068
		(1, 3)	0.223	0.177	0.059	0.07
		(1, 2)	0.68	0.384	0.167	0.228
		(1, 1)	1	0.21	0.84	0.74
Fisher	(6, 6)	(6, 6)	0.046	0.04	0.034	0.038
		(5, 5)	0.075	0.067	0.047	0.059
		(4, 4)	0.201	0.142	0.078	0.053
		(3, 3)	0.613	0.359	0.198	0.048
		(2, 2)	0.971	0.532	0.669	0.039
		(1, 1)	1	0.229	0.999	0.064
Student	(1, 0)	(1, 0)	0.048	0.024	0.049	0.055
		(1, 0.2)	0.168	0.05	0.192	0.252
		(1, 0.4)	0.463	0.123	0.576	0.658
		(1, 0.6)	0.813	0.232	0.918	0.957
		(1, 0.8)	0.958	0.362	0.995	0.998
		(1, 1)	0.988	0.439	1	1
Student	(1, 0)	(1, 0)	0.04	0.026	0.038	0.045
		(2, 0)	0.494	0.02	0.059	0.051
		(3, 0)	0.816	0.022	0.096	0.053
		(4, 0)	0.894	0.015	0.106	0.05
		(5, 0)	0.946	0.019	0.149	0.044
		(6, 0)	0.967	0.028	0.159	0.058
Uniform	(0, 1)	(0, 0)	0.041	0.043	0.024	0.043
		(0.1, 0.1)	0.622	0.693	0.363	0.665
		(0.2, 0.2)	0.996	0.998	0.977	0.993
		(0.3, 0.3)	1	1	1	1
		(0.4, 0.4)	1	1	1	1
Weibull	(1, 3)	(1, 3)	0.055	0.05	0.037	0.056
		(1, 2.5)	0.214	0.237	0.11	0.178
		(1, 2)	0.78	0.801	0.559	0.687
		(1, 1.5)	0.997	0.999	0.959	0.986
		(1, 1)	1	1	1	1

## Приложение Б

## Графическое представление мощности тестов

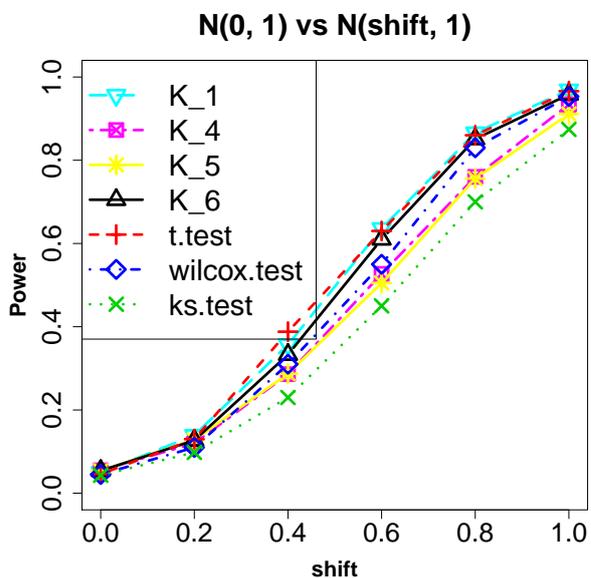


Рис. Б.0.1. Нормальное распределение

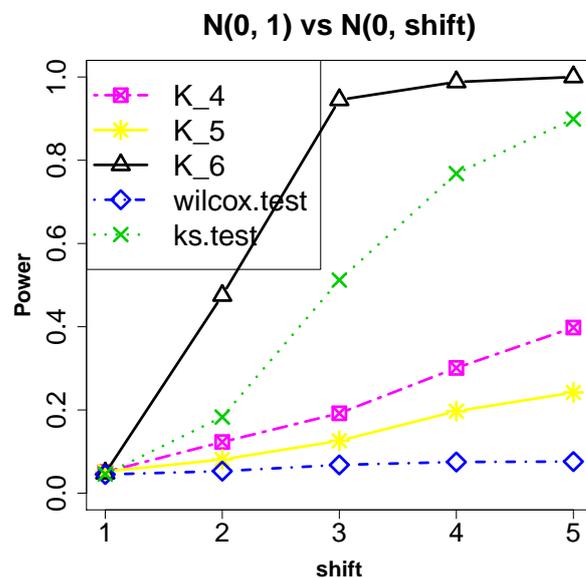


Рис. Б.0.2. Нормальное распределение

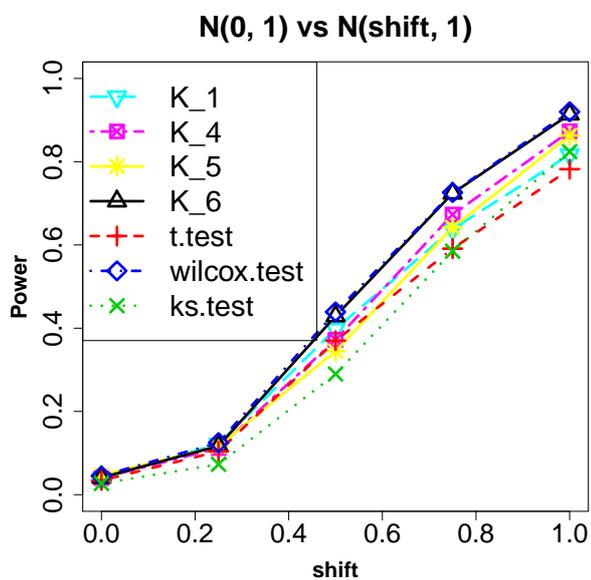


Рис. Б.0.3. Загрязненное нормальное распределение

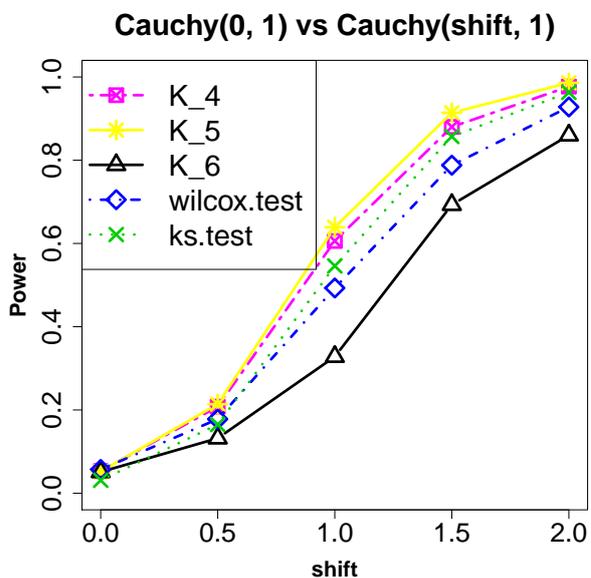


Рис. Б.0.4. Распределение Коши

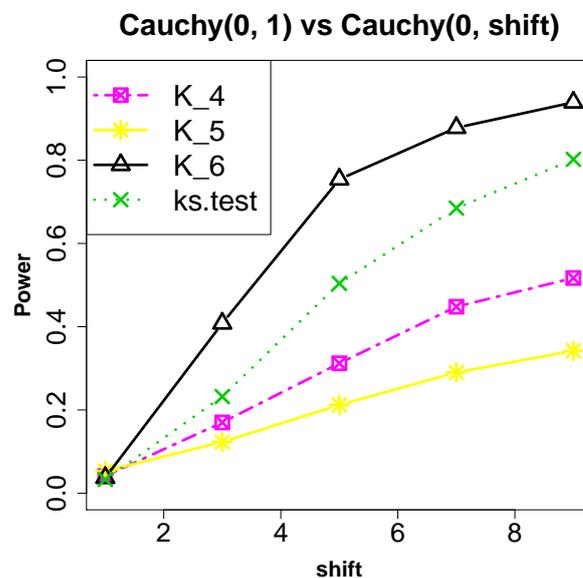


Рис. Б.0.5. Распределение Коши

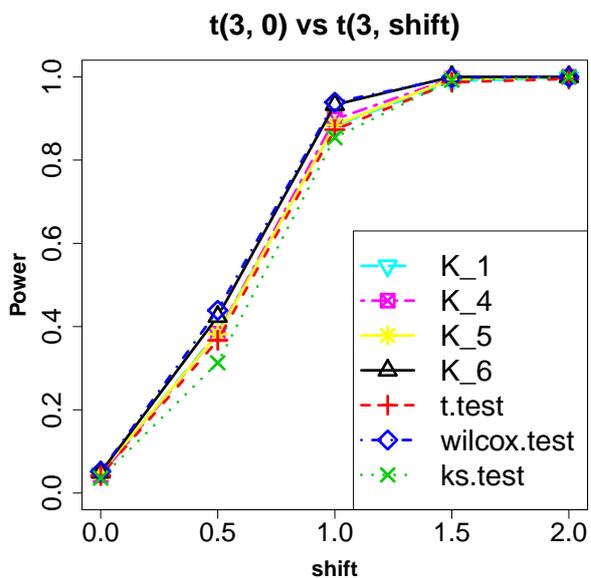


Рис. Б.0.6. Распределение Стьюдента

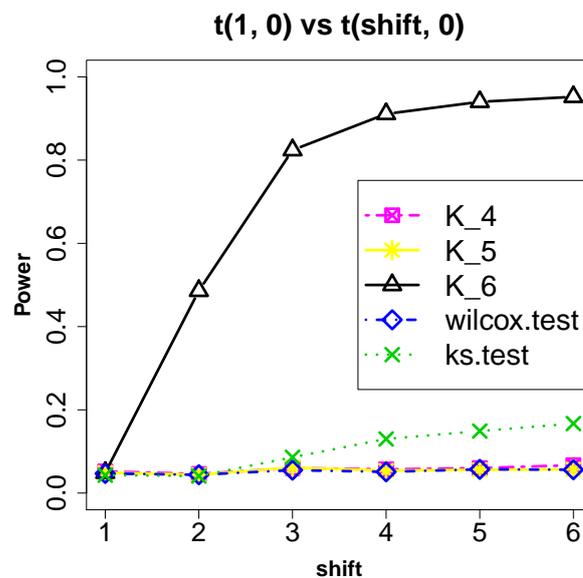


Рис. Б.0.7. Распределение Стьюдента

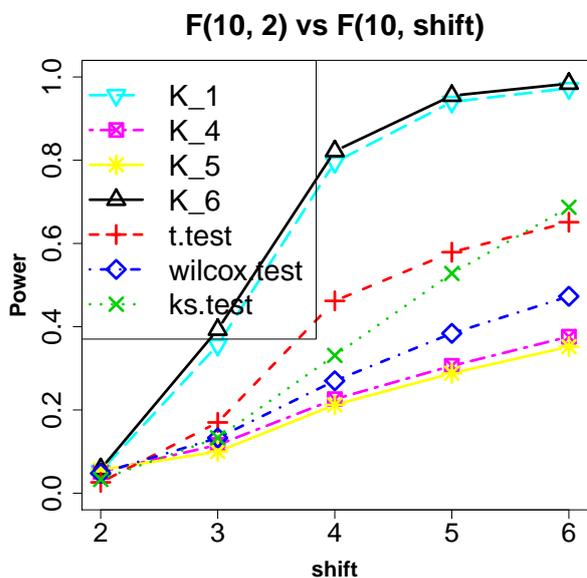


Рис. Б.0.8. Распределение Фишера

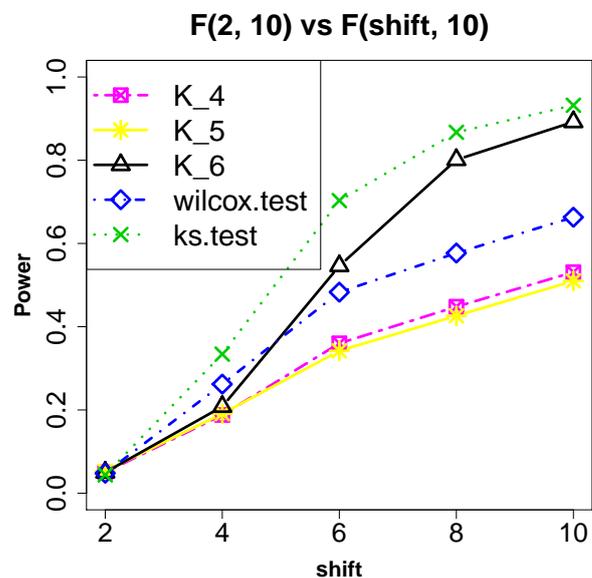


Рис. Б.0.9. Распределение Фишера

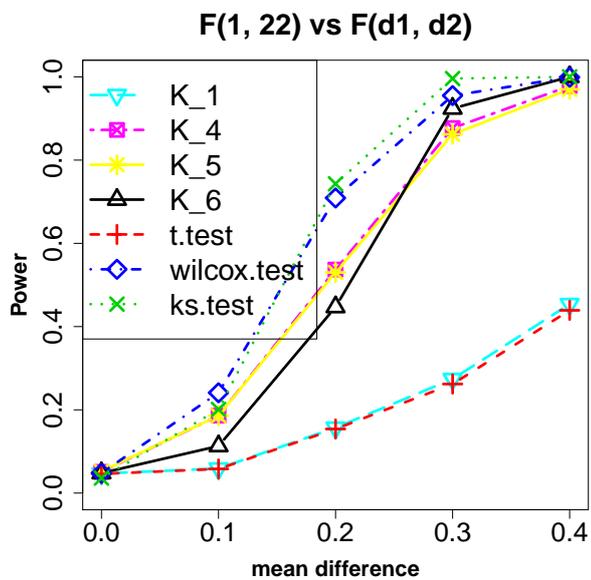


Рис. Б.0.10. Распределение Фишера

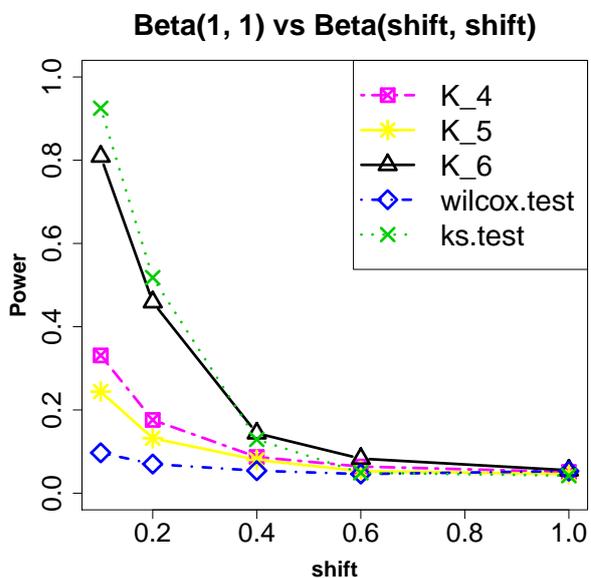


Рис. Б.0.11. Бета-распределение

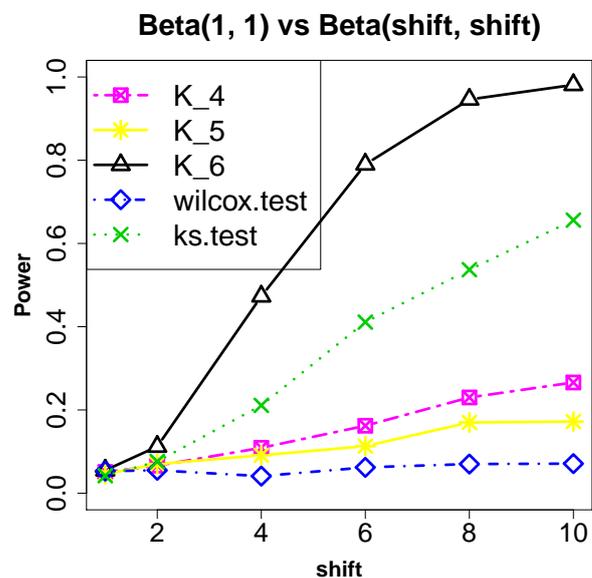


Рис. Б.0.12. Бета-распределение

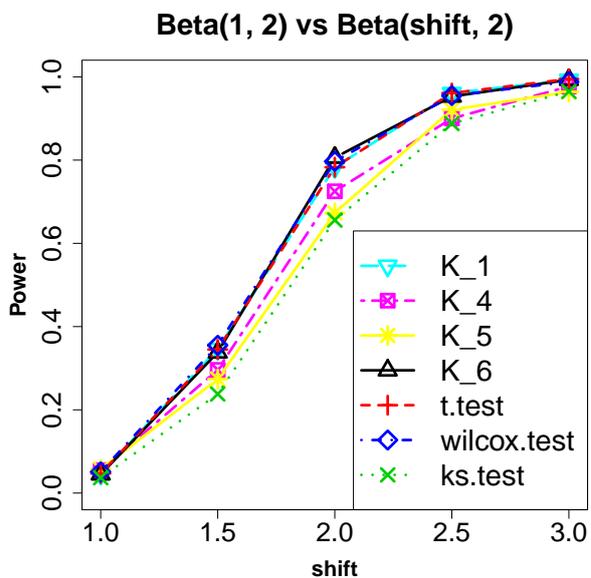


Рис. Б.0.13. Бета-распределение

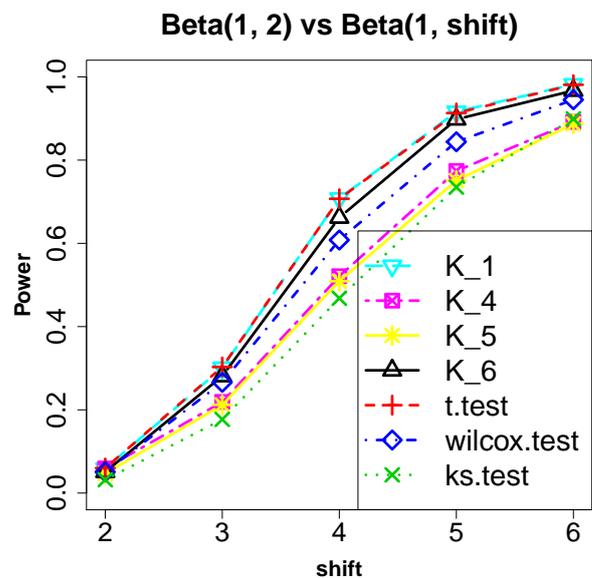


Рис. Б.0.14. Бета-распределение

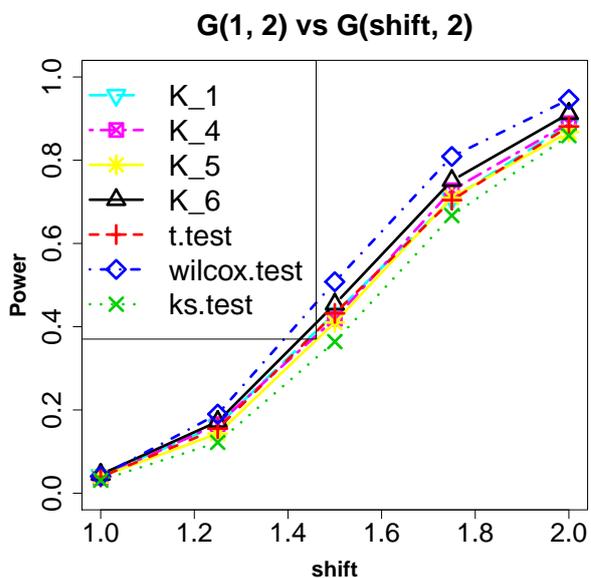


Рис. Б.0.15. Гамма-распределение

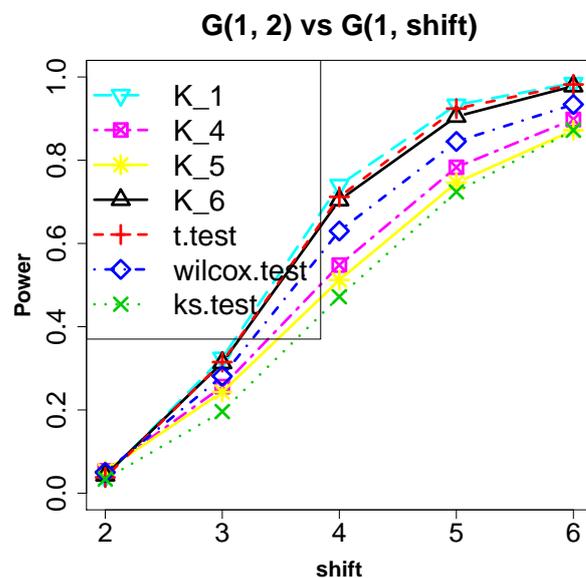


Рис. Б.0.16. Гамма-распределение

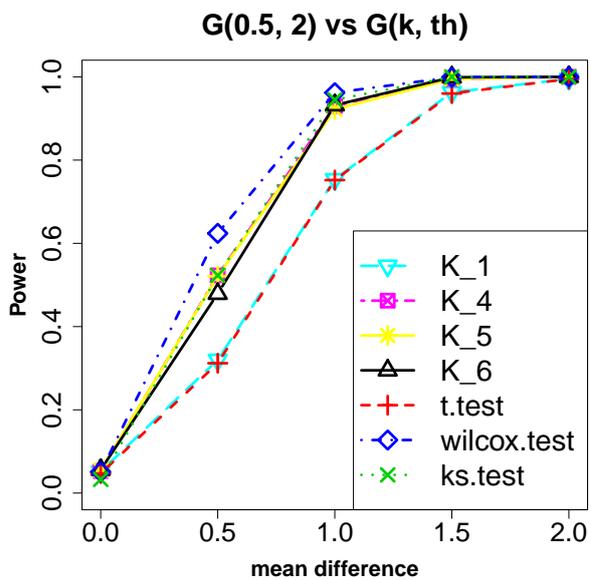


Рис. Б.0.17. Гамма-распределение

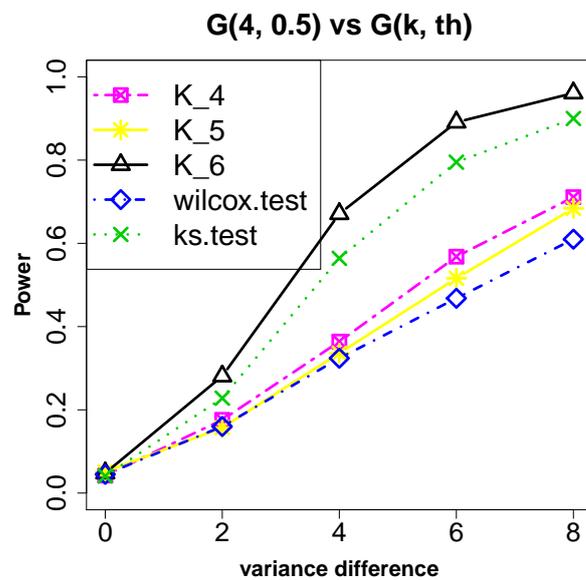


Рис. Б.0.18. Гамма-распределение

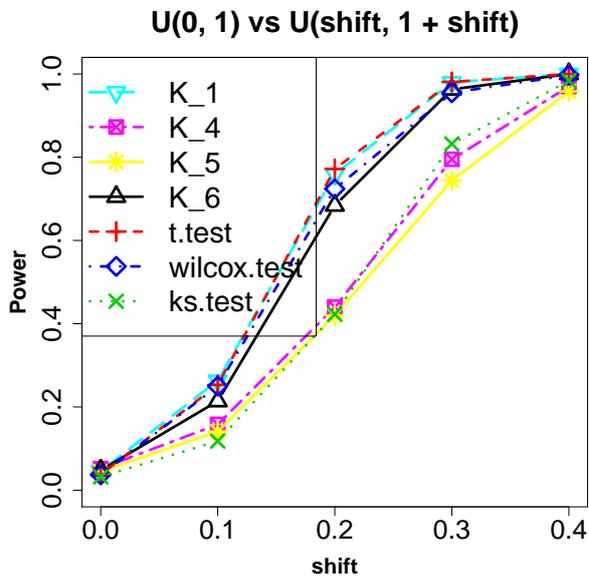


Рис. Б.0.19. Равномерное распределение

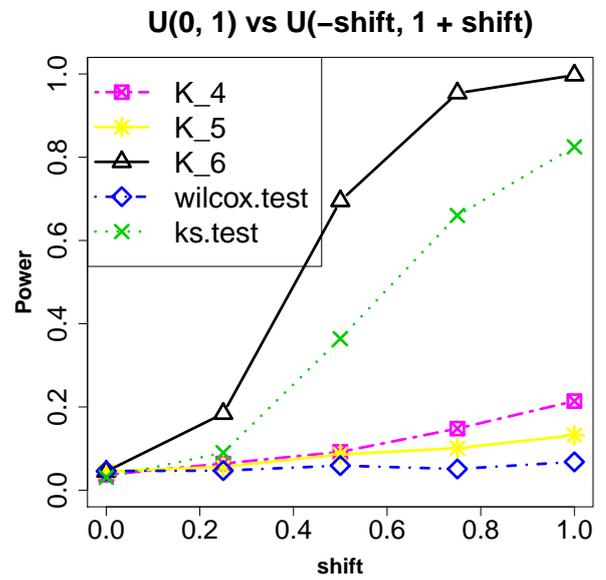


Рис. Б.0.20. Равномерное распределение

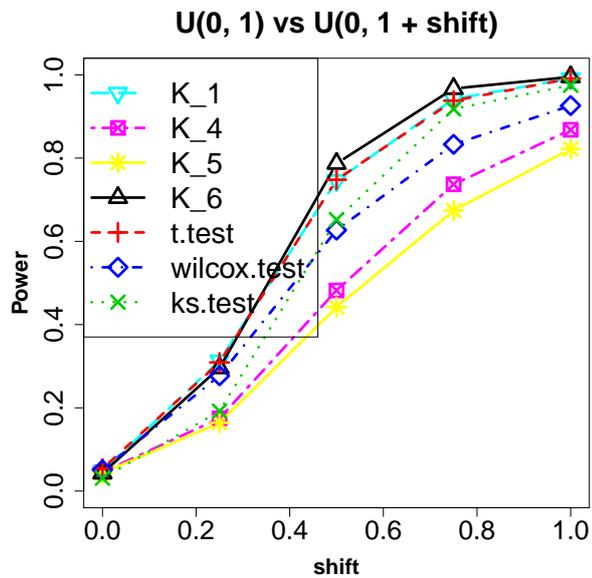


Рис. Б.0.21. Равномерное распределение

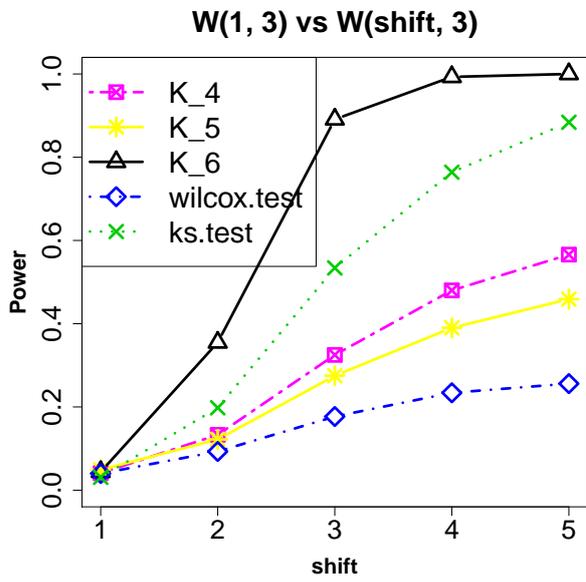


Рис. Б.0.22. Распределение Вейбулла

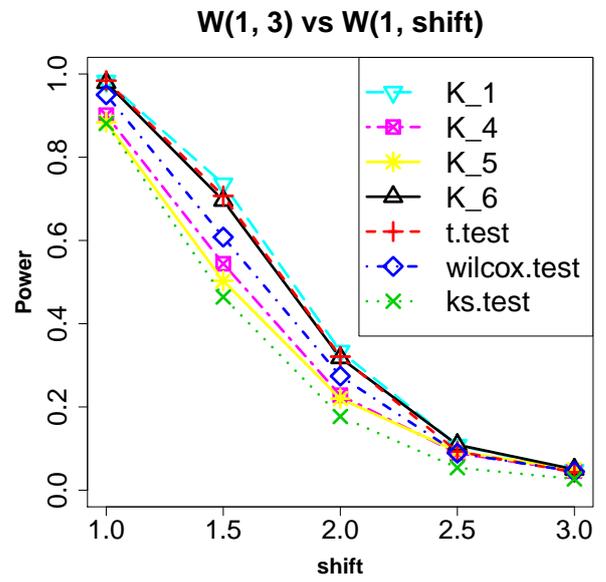


Рис. Б.0.23. Распределение Вейбулла

## Приложение В

### Реализация алгоритма

Задача была рассмотрена для различных распределений.

Рассмотрим на примере нормального распределения с изменяющимся средним реализацию алгоритма на языке R.

Объявим функцию, моделирующую один эксперимент, принимающую три параметра - размер выборки  $n$ , среднее *shift* и уровень значимости  $a$ :

```
1 power.test <- function(n, shift, a = 0.05) {
```

Промоделируем 2 выборки из нормального распределения объемом  $n$  с общей дисперсией  $\sigma^2 = 1$  и посчитаем их медианы для сокращения вычислений:

```
2   x <- rnorm(n, mean = 0, sd = 1)
3   y <- rnorm(n, mean = shift, sd = 1)
4   x.m <- median(x)
5   y.m <- median(y)
```

Вычислим начальную статистику перестановочных критериев:

```
6   stat.K_1 <- (mean(x) - mean(y)) ^ 2
7   stat.K_4 <- (x.m - y.m) ^ 2
8   stat.K_5 <- (sum(abs(x - x.m) + abs(y - y.m))) ^ 2
9   stat.K_6 <- sum(sapply(1:n, function(i){ sum(abs(x - y[i])) })))
```

Далее с помощью стандартной функции `sample()` генерируем перестановки.

`sample(c(x, y))` генерирует случайную перестановку совместной выборки. Выделив из общей совокупности  $x$  и  $y$ , вычисляем статистики перестановочных критериев:

```
10  perm.test = function(x, y) {
11    z <- sample(c(x, y))
12
13    x.perm <- z[1:n]
14    y.perm <- z[(n + 1):(2 * n)]
15
16    x.m <- median(x.perm)
17    y.m <- median(y.perm)
18
19    perm.stat.K_1 <- (mean(x.perm) - mean(y.perm)) ^ 2
20    perm.stat.K_4 <- (x.m - y.m) ^ 2
21    perm.stat.K_5 <- (sum(abs(x.perm - x.m) + abs(y.perm - y.m))) ^ 2
22    perm.stat.K_6 <- sum(sapply(1:n,
23      function(i){ sum(abs(x.perm - y.perm[i])) })))
24
25    list(perm.stat.K_1, perm.stat.K_4, perm.stat.K_5, perm.stat.K_6)
```

26 }  
 }  
 }

Прогоним функцию `perm.test()` 1600 раз:

```
27 many.perm.stat = replicate(1600, perm.test(x, y))
```

Получим  $p$ -значение для тестов,  $p = \frac{r_1}{r_2}$ , где  $r_2 = 1600$ :

```
28 p.K_1 <- mean(many.perm.stat[1,] > stat.K_1)
```

```
29 p.K_4 <- mean(many.perm.stat[2,] > stat.K_4)
```

```
30 p.K_5 <- mean(many.perm.stat[3,] > stat.K_5)
```

```
31 p.K_6 <- mean(many.perm.stat[4,] > stat.K_6)
```

Проверяем, была ли отвергнута нулевая гипотеза для каждого критерия, возвращаем результат и завершаем функцию:

```
32 power.K_1 <- if (p.K_1 < a) 1 else 0
```

```
33 power.K_4 <- if (p.K_4 < a) 1 else 0
```

```
34 power.K_5 <- if (p.K_5 > 1-a) 1 else 0
```

```
35 power.K_6 <- if (p.K_6 < a) 1 else 0
```

```
36
```

```
37 return(c(power.K_1, power.K_4, power.K_5, power.K_6))
```

```
38 }
```

Положим  $m = 1000$  и будем моделировать выборки  $x$  и  $y$   $m$  раз. Тогда мощность тестов равна:

```
39 m <- 1000
```

```
40 tmp <- replicate(m, power.test())
```

```
41 print(mean(tmp[1,]), mean(tmp[2,]), mean(tmp[3,]), mean(tmp[4,]))
```