

**«Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)**

Направление	01.04.02 – Прикладная математика и информатика
Программа	Математическое и программное обеспечение вычислительных машин
Факультет	КТИ
Кафедра	МО ЭВМ

К защите допустить

Зав. кафедрой

Кринкин К.В.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
МАГИСТРА**

**Тема: РАЗРАБОТКА ПРОГРАММНОГО ПРОДУКТА ДЛЯ ПРОВЕРКИ
ГИПОТЕЗЫ ЭКСПОНЕНЦИАЛЬНОСТИ**

Студент		<hr/>	Носов В.Е.
		<i>подпись</i>	
Руководитель	Д.Т.Н., проф. <i>(Уч. степень, уч. звание)</i>	<hr/>	Григорьев Ю.Д.
		<i>подпись</i>	
Консультанты	К.Т.Н. <i>(Уч. степень, уч. звание)</i>	<hr/>	Яновский В.В.
		<i>подпись</i>	
	К.Т.Н., доцент <i>(Уч. степень, уч. звание)</i>	<hr/>	Иванов А.Н.
		<i>подпись</i>	

Санкт-Петербург
2016

ЗАДАНИЕ НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ

Утверждаю
Зав. кафедрой МО ЭВМ
_____ Кринкин К.В.
« » _____ 20__ г.

Студент	Носов В. Е.	Группа	0381
---------	-------------	--------	------

Тема работы: Разработка программного продукта для проверки гипотезы экспоненциальности

Место выполнения ВКР: СпбГЭТУ «ЛЭТИ»

Исходные данные (технические требования):

Комплекты средств разработки языков Java и R, сервер Tomcat, СУБД MySQL

Содержание ВКР:

- Рассмотрение задачи проверки гипотезы экспоненциальности;
- Обзор программных составляющих разработанного продукта.

Перечень отчетных материалов: пояснительная записка, иллюстративный материал

Дополнительные разделы: Специальные вопросы обеспечения безопасности

Дата выдачи задания
« » 20 г.

Дата представления ВКР к защите
« » 20 г.

Студент Носов В.Е.

Руководитель _____ д.т.н., проф. Григорьев Ю.Д.
(Уч. степень, уч. звание)

Консультант к.т.н., доцент Иванов А.Н.
(Уч. степень, уч. звание)

КАЛЕНДАРНЫЙ ПЛАН ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Утверждаю
Зав. кафедрой МО ЭВМ
_____ Кринкин К. В.
« ____ » _____ 20 ____ г.

Студент Носов В.Е.

Группа 0381

Тема работы: Разработка программного продукта для проверки гипотезы экспоненциальности

№ п/п	Наименование работ	Срок выполнения
1	Обзор литературы по теме работы	20.02 – 25.09
2	Изучение необходимых теоретических основ и понятий	20.09 – 15.12
3	Разработка программного продукта	15.12 – 10.04
5	Оформление пояснительной записки	29.02 – 01.06
6	Оформление иллюстративного материала	29.02 – 00.00

Студент _____

Носов В.Е.

Руководитель _____

Григорьев Ю.Д.

РЕФЕРАТ

Пояснительная записка 80 стр., 19 рис., 18 табл., 30 ист., 0 прил.

ЭКСПОНЕНЦИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ, ПРОВЕРКА ГИПОТЕЗ, ПРОГРАММНЫЙ ПРОДУКТ, СТАТИСТИКИ, СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, МОЩНОСТЬ, АСИМПТОТИЧЕСКИ ОПТИМАЛЬНАЯ ЭФФЕКТИВНОСТЬ.

Объектом исследования является проверка гипотезы экспоненциальности на основе выборочных данных с помощью различных критериев.

Целью данной работы является разработка программного продукта, предназначенного для проверки гипотезы экспоненциальности.

В ходе данной выпускной работы рассматривается множество критериев для проверки гипотезы экспоненциальности, выполняется моделирование их критических значений и мощностей. Также описываются архитектура и различные характеристики разработанного программного продукта, рассматриваются примеры его применения на реальных задачах.

ABSTRACT

Exponential distribution is one of the most popular distributions in the field of Mathematical Statistics. The assumption of exponentiality is used in many life testing studies. That's why a lot of tests for exponentiality have been already proposed. The goal of this work is development of an software product for testing for exponentiality by some tests. One of the problems, that were solved in this work, is modulating of critical values for statistics, which have unknown distributions of values. Besides, the product has an ability to modulate power values for statistics against various alternative distributions. The mission of the product is giving an information, which should help to accept or reject the hypothesis of exponentiality.

Оглавление

Введение.....	7
Глава 1. Проверка статистических гипотез	9
1.1. Постановка задачи.....	9
1.2. Мощность критериев. Простые и сложные критерии	12
1.3. Проверка гипотезы экспоненциальности	14
1.4. Асимптотически оптимальная эффективность	23
1.5. Статистическое моделирование	26
Глава 2. Программный продукт.....	35
2.1. Структура продукта	35
2.2. Имитационный блок	44
2.3. Статистики	57
2.3. Примеры использования программного продукта	65
Глава 3. Специальные вопросы обеспечения безопасности.....	70
Заключение	76
Список использованных источников	78
Приложение А. Таблицы квантилей статистик.....	81

ВВЕДЕНИЕ

Экспоненциальное распределение, наряду с нормальным, является одним из самых востребованных распределений в теории надежности. Также имеется ряд инженерных задач, использующих данное распределение. По этой причине возникает необходимость в проверке гипотезы экспоненциальности. На данный момент существует большое количество различных критериев для проверки выборочных данных на экспоненциальность, что подтверждает популярность данного распределения.

При проверке различных гипотез возникает необходимость знать критические значения (квантили) используемых статистик. Для ряда критериев известны их функции распределения, что позволяет аналитически вычислять квантили. Для остальных данные значения моделируются. При этом возникает вопрос, какой из этих статистик отдать предпочтение, если результаты проверки гипотезы с их помощью на одних и тех же выборочных данных противоречат друг другу? Зачастую альтернативное распределение при проверке гипотезы не задано. В таком случае проверка выполняется на согласованность. В противном случае, если альтернативное распределение задано, возникает возможность сравнить характеристики используемых критериев против данной альтернативы. Имеются различные подходы к принятию решений в подобных ситуациях, основанные на понятиях мощности и асимптотически оптимальной эффективности.

Целью данной работы является разработка программного продукта, предназначенного для проверки гипотезы экспоненциальности.

Для достижения поставленной цели были сформулированы следующие задачи:

- наработка библиографической базы, состоящей из различных источников информации о статистиках проверки выборочных данных на экспоненциальность;

- создание таблиц квантилей статистик, для которых такие таблицы не приведены в литературе, а также дополнение уже имеющихся таблиц квантилей;
- разработка архитектуры приложения, отвечающей современным требованиям;
- проверка разработанного программного продукта на реальных задачах, требующих от распределения случайной величины экспоненциальности.

Глава 1. ПРОВЕРКА ГИПОТЕЗЫ ЭКСПОНЕНЦИАЛЬНОСТИ

Проверка гипотезы экспоненциальности является частным случаем проверки статистических гипотез. Имеется множество литературных источников, описывающих теоретические основы данного раздела математической статистики, например [4]. В данной главе приводится краткое описание основных понятий рассматриваемой прикладной области, а также ряд статистик для проверки гипотезы экспоненциальности, использованных в разработанном программном продукте.

1.1. Постановка задачи

Пусть $X = (X_1, \dots, X_n)$ – повторная выборка, т. е. набор независимых одинаково распределенных случайных величин (н. о. р. с. в.), \mathcal{F} – совокупность априори допустимых распределений X , F – истинное распределение с. в. ($F \in \mathcal{F}$). *Статистической гипотезой* называют любое утверждение о виде распределения F . Задача состоит в том, чтобы по результатам наблюдений (выборочным данным) принять или отклонить эту гипотезу. В общем случае для проверки гипотезы необходимо выделить некоторое множество допустимых значений \mathcal{F}_0 ($\mathcal{F}_0 \subset \mathcal{F}$) и проверить, справедливо ли утверждение $H_0: F \in \mathcal{F}_0$. H_0 принято называть *основной (нулевой) гипотезой*. В случае если гипотеза H_0 справедлива, то она принимается. Если H_0 ложна, гипотеза отвергается, и принимается другая гипотеза $H_1: F \in \mathcal{F}_1 = \mathcal{F} \setminus \mathcal{F}_0$, называемая *альтернативной (либо конкурирующей)*.

Правило, согласно которому гипотеза H_0 принимается, или отвергается, называется *статистическим критерием* для проверки гипотезы H_0 . Вид критерия существенным образом зависит от поставленной задачи. Довольно часто выделяют две группы задач проверки статистических гипотез. Первую группу составляют задачи, связанные с проверкой вида семейства распределений, во вторую входят задачи, связанные с проверкой конкретных параметров, определяющих распределения. Рассматриваемая в данной работе за-

дача о проверке гипотезы экспоненциальности является одним из примеров задач из первой группы. В качестве примеров из второй группы можно привести критерии для проверки гипотезы о значении параметра экспоненциального распределения.

В случае, когда класс \mathcal{F} содержит все распределения на заданном выборочном пространстве, нет необходимости в конкретизации альтернативной гипотезы $H_1 = \overline{H_0}$. В данном случае требуется проверить, согласуются ли выборочные данные с основной гипотезой H_0 . Если данные не согласуются с H_0 , то считается, что они её опровергают. Соответствующие критерии называются *критериями согласия*. В других случаях класс всех допустимых распределений задается параметрическим семейством $\mathcal{F} = \{F(x, \theta), \theta \in \Theta\}$. Критерии, соответствующие данному классу распределений, называют *параметрическими*. Гипотезы, проверяемые в данных критериях, имеют вид

$$H_0: \theta \in \Theta_0 \subset \Theta; \quad H_1: \theta \in \Theta_1 = \Theta \setminus \Theta_0 \quad (1.1)$$

Каждому статистическому критерию соответствует некоторое разбиение выборочного пространства \mathfrak{X} на два множества \mathfrak{X}_0 и \mathfrak{X}_1 ($\mathfrak{X}_0 \cup \mathfrak{X}_1 = \mathfrak{X}$, $\mathfrak{X}_0 \cap \mathfrak{X}_1 = \emptyset$), где в множество \mathfrak{X}_0 входят все возможные выборки X , для которых гипотеза H_0 принимается, а в \mathfrak{X}_1 – все выборки X , для которых H_0 отвергается (и принимается H_1). Множество \mathfrak{X}_0 принято называть *областью принятия гипотезы H_0* , а \mathfrak{X}_1 – *областью её отклонения* (или *критической областью*).

Правило проверки гипотезы H_0 : Пусть x – наблюдаемая реализация выборки X ($x \in \mathfrak{X}$). Если $x \in \mathfrak{X}_1$, то считается, что произошло маловероятное событие при справедливости H_0 . В данном случае говорят о несогласованности гипотезы H_0 с выборочными данными (гипотеза H_0 отклоняется). Если $x \in \mathfrak{X}_0$, считается, что данные согласуются с H_0 , и гипотеза принимается.

Одной из основных проблем при проверке гипотез является выбор областей \mathfrak{X}_0 и \mathfrak{X}_1 . Согласно правилу проверки гипотез критическая область \mathfrak{X}_1

должна быть выбрана таким образом, чтобы уменьшить условную вероятность $P\{x \in \mathfrak{X}_1 | H_0\}$. Обычно её ограничивают некоторым малым числом α , называемым *уровнем значимости*:

$$P\{x \in \mathfrak{X}_1 | H_0\} \leq \alpha. \quad (1.2)$$

Величину α выбирают в зависимости от конкретной задачи, но обычно это значения 0.01, 0.05, 0.1. Уровень значимости можно считать вероятностью ложного отвержения гипотезы H_0 , когда она верна (или вероятностью *ошибки первого рода*).

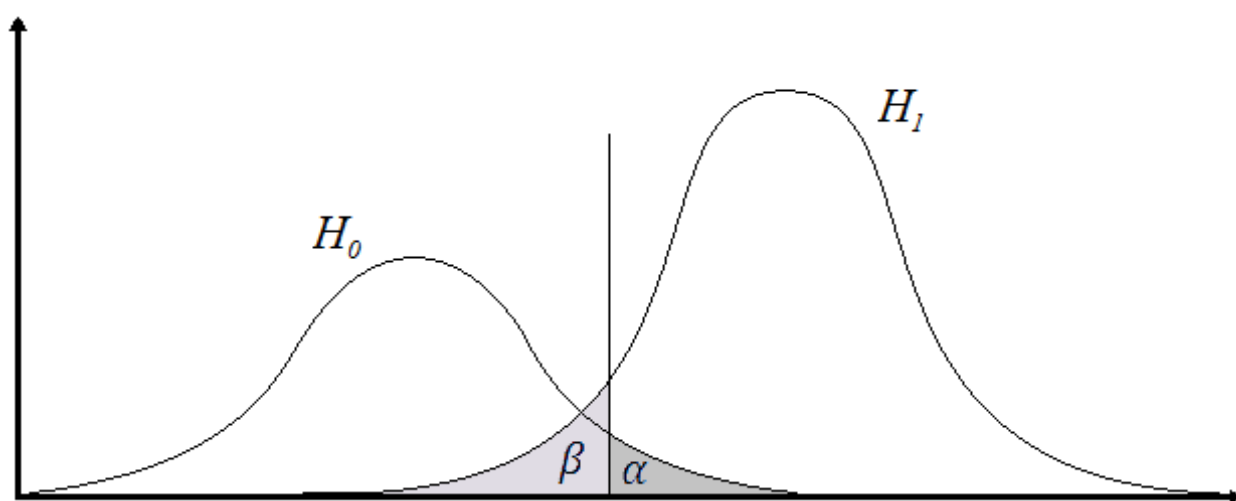


Рисунок 1.1. Ошибки первого и второго рода.

Ошибки первого и второго рода: При проверке некоторой гипотезы может быть принято правильное решение, либо может быть допущена одна из двух ошибок: *ошибка первого рода*, когда отвергается гипотеза H_0 , хотя она является верной, и *ошибка второго рода*, когда H_0 является ложной, но при этом принимается.

На рисунке 1.1 изображены плотности распределений вероятностей значений статистического критерия в случаях, когда верна нулевая гипотеза (H_0), либо верна альтернативная гипотеза (H_1). Вертикальная линия отображает заданный уровень значимости α . Если полученное значение статистики для выборки попадает левее данной линии, то гипотеза H_0 принимается. В ином случае принимается гипотеза H_1 . Ошибки первого и второго рода отме-

чены на графике символами α и β соответственно. При смещении уровня значимости влево (увеличивая его) можно уменьшить объем ошибок второго рода β , но при этом увеличится объем ошибок первого рода α . Аналогично, смещая уровень значимости вправо, таким образом, уменьшая его, уменьшится значение α , но при этом увеличится значение β . По этой причине необходимо искать некую «золотую середину», опираясь на условия конкретной задачи.

Ошибки второго рода являются более нежелательными и опасными, могут приводить к более большим потерям, чем ошибки первого рода. Поэтому стоит стремиться к уменьшению вероятности ошибки второго рода, допуская при этом сравнительно большие значения уровня значимости α .

1.2. Мощность критериев. Простые и сложные критерии

Важнейшей числовой характеристикой критерия является его *функция мощности*

$$W(F) = P\{x \in \mathfrak{X}_1 | F\}, \quad (1.3)$$

где $F \in \mathcal{F}$, т. е. вероятность попадания значения выборки x в критическую область \mathfrak{X}_1 , когда F – её истинное распределение. При $F \in \mathcal{F}_1$ значение $W(F)$ принято называть *мощностью критерия при альтернативе F* .

Через функцию мощности можно выразить ошибки первого и второго рода:

$$P_{\text{ошибки 1 рода}} = P\{H_1 | H_0\} = W(F) \text{ при } F \in \mathcal{F}_0 \quad (1.4)$$

$$P_{\text{ошибки 2 рода}} = P\{H_0 | H_1\} = 1 - W(F) \text{ при } F \in \mathcal{F}_1 \quad (1.5)$$

Из (1.5) следует, что мощность критерия при альтернативе $F \in \mathcal{F}_1$ $W(F) = 1 - P\{H_0 | H_1\}$.

Также можно переписать формулу (1.2) через функцию мощности:

$$W(F) \leq \alpha \quad \forall F \in \mathcal{F}_0 \quad (1.6)$$

Мощность, как характеристика критерия, очень важна. При наличии нескольких критериев для проверки основной гипотезы против конкретной конкурирующей гипотезы выбор нужно сделать в пользу наиболее мощного критерия против данной альтернативы. Критерий T_1 считается *мощнее* критерия T_2 против конкретной альтернативной гипотезы $H_1: F \in \mathcal{F}_1$, если

$$W_{T_1}(F) > W_{T_2}(F), F \in \mathcal{F}_1. \quad (1.7)$$

Параметрические гипотезы: Как видно из (1.1), параметры распределений из семейства \mathcal{F} для гипотез H_0 и H_1 образуют два множества Θ_0 и $\Theta_1 = \Theta \setminus \Theta_0$. Примерами подобных параметрических множеств являются:

1. $\Theta' = \{\theta_0\}$ – множество, состоящее лишь из одного значения параметра;
2. $\Theta' = \{\theta_0, \dots, \theta_n\}$ – множество, состоящее из фиксированного набора некоторых значений параметра;
3. $\Theta' = \{\forall \theta \in \Theta: G(\theta) = g_0\}$ – множество, состоящее из всех значений параметра, для которых выполнено условие $G(\theta) = g_0$, где $G(\theta)$ – некоторая функция, g_0 – фиксированное значение, $\Theta' \subset \Theta$.

Если в множество Θ_0 (Θ_1) входит только одно значение параметра, то гипотезу H_0 (H_1) называют *простой*. В противном случае гипотезу называют *сложной*. Случаи, когда основная и альтернативная гипотезы одновременно являются простыми, встречаются довольно редко. Намного чаще встречаются ситуации, когда обе гипотезы (или хотя бы одна из них) сложные.

В случае, когда обе гипотезы являются простыми ($\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$), ошибки первого и второго рода равны $W(\theta_0)$ и $1 - W(\theta_1)$ соответственно ($W(\theta)$ – функция мощности). Для того чтобы найти оптимальный критерий, необходимо из всех критериев с вероятностью ошибки первого рода $W(\theta_0) \leq \alpha$ выбрать критерий, имеющий наименьшую вероятность ошибки второго рода $1 - W(\theta_1)$ или наибольшую мощность $W(\theta_1)$.

Из двух критериев, имеющих одинаковые вероятности ошибок первого рода, критерий K_1 *равномерно более мощный* [4], чем K_2 , если

$$W_1(\theta) \geq W_2(\theta), \text{ где } \theta \in \Theta_1. \quad (1.8)$$

Если критерий K_1 равномерно более мощный, чем любой другой критерий с такой же вероятностью ошибки первого рода, то он называется *равномерно наиболее мощным* [4].

1.3. Проверка гипотезы экспоненциальности

Экспоненциальное распределение играет важную роль в таких областях, как теория надежности, анализ данных типа времени жизни и др. С середины XX века было предложено огромное количество критериев проверки гипотезы экспоненциальности. Одними из первых были такие статистики, как критерий Гринвуда[21], критерий Морана[27].

Классические критерии проверки гипотез, обычно излагаемые в учебниках (такие, как критерий Неймана-Пирсона, Вальда[9, с. 369]), редко имеют удобный для вычислений вид и не всегда являются оптимальными против конкретных альтернатив. В предлагаемом в данной работе продукте представлены критерии, квантили и мощности которых против конкретных альтернатив могут быть вычислены.

Одним из главных свойств экспоненциального распределения с точки зрения теории надежности является тот факт, что его плотность *интенсивности отказов* постоянна.

В [1] приводится определение интенсивности отказов, как вероятности отказа элемента в течение конечного интервала времени y при условии, что возраст элемента равен x . Функция интенсивности отказов имеет вид

$$\frac{F(x+y)-F(x)}{1-F(x)}, \quad (1.9)$$

где $F(x)$ – функция распределения с. в. x . Плотность интенсивности отказов равна

$$\eta(x) = \frac{f(x)}{1-F(x)}, \quad (1.10)$$

где $f(x)$ – плотность распределения с. в. x .

Плотность интенсивности отказов постоянна тогда и только тогда, когда с. в. имеет экспоненциальное распределение. По этой причине при проверке гипотезы экспоненциальности можно также проводить анализ функции интенсивности отказов на основе выборочных данных. Например, непараметрический критерий Манна, разработанный в [26], предназначен поиска какой-либо тенденции изменения интервалов между отказами.

Пусть $X = (X_1, \dots, X_n)$ – повторная выборка, F – функция распределения случайных величин X . Необходимо на основании X выполнить проверку сложной гипотезы $H_0 : F \in \mathcal{F}_0 \equiv \text{Exp}(\lambda)$, где

$$\text{Exp}(\lambda) = \{F : F(x) = 1 - e^{-\lambda x}, \lambda > 0, x \geq 0\}. \quad (1.11)$$

– семейство экспоненциальных распределений с неизвестным параметром масштаба λ . Альтернативная гипотеза H_1 состоит в том, что F не принадлежит классу $\text{Exp}(\lambda)$:

$$H_1 : F \in \mathcal{F}_1 \equiv \mathcal{F} \setminus \text{Exp}(\lambda). \quad (1.12)$$

Для выполнения данной проверки нужно воспользоваться одним из критериев, которые будут описаны ниже. При выборе критерия необходимо учесть, что должны быть известны его *критические значения (квантили)* при текущем объеме выборки n и заданном уровне значимости α . В общем случае квантиль – значение, которое заданная с. в. не превышает с фиксированной вероятностью.

На рис. 1.2 изображена плотность распределения вероятности F некоторой случайной величины x , а также её α -квантиль, изображенная в виде вертикальной разделяющей полосы. Площадь серой области справа от полосы равна α от общей площади под графиком плотности F . Соответственно, вероятность попадания случайной величины x в эту область также равна α .

Существует ряд статистик, для которых известны их распределения, поэтому можно аналитически посчитать их квантили, или воспользоваться готовыми квантилями этих распределений. Для остальных данные значения моделируются. Таблицы квантилей для некоторых критериев при наиболее популярных объемах выборки и уровнях значимости можно найти в статьях и справочниках. Одним из назначений разработанного продукта является пополнение этих таблиц, а также моделирование квантилей статистик, для которых подобных таблиц не имеется. Методы моделирования рассматривается в следующем параграфе.

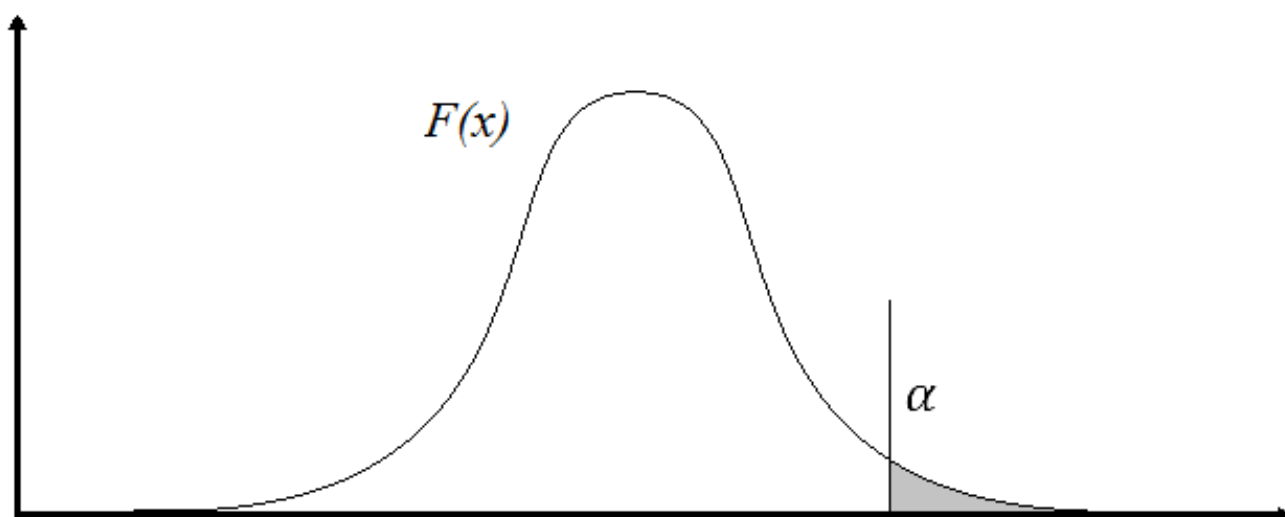


Рисунок 1.2. Квантиль некоторого распределения F .

Процесс проверки гипотезы экспоненциальности на основе повторной выборки X объемом n с помощью некоторого критерия T при заданном уровне значимости α выглядит следующим образом:

1. вычисляется значение статистики для заданной выборки – $T(X)$;
2. далее это число сравнивается с квантилями статистики (с двумя квантилями $T_{n,\alpha} \leq T(X) \leq T_{n,1-\alpha}$, если статистика *двухсторонняя*, либо с одним квантилем $T(X) \leq T_{n,1-\alpha}$, если статистика *односторонняя*);
3. если условие выполняется, то считается, что выборка X не противоречит гипотезе об экспоненциальности, и гипотеза H_0 принимается. В против-

ном случае, если условие не выполняется, считается, что данные не согласуются с H_0 , и гипотеза отвергается.

Пример 1.1. В [5, с. 257] рассматривается проверка гипотезы экспоненциальности для моментов появления отказов автомобиля во время стендовых виброиспытаний. Продолжительность испытаний равна 245 часов. В таблице 1.1 представлены интервалы между отказами.

Таблица 1.1. Интервалы между отказами автомобиля во время стендовых виброиспытаний.

<i>n</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
	21.2	26.7	11.3	2.8	12.6	0.1	2.1	7.5	6.7	2.3
<i>n</i>	<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>	<i>17</i>	<i>18</i>	<i>19</i>	<i>20</i>
	15.3	4.3	14.1	16.9	7.7	5.8	7.3	32.1	17.6	4.5

Для проверки гипотезы экспоненциальности в [5, с. 257] используется критерий Бартлетта [15]. Данная статистика имеет вид

$$B_n = \frac{2n \left[\ln(\bar{X}) + \frac{1}{n} (\sum_{i=1}^n \ln X_i) \right]}{1 + \frac{(n+1)}{6n}},$$

где X_i – случайная величина, n – объем выборки, \bar{X} – выборочное среднее случайных величин.

В примере используется уровень значимости $\alpha = 0.05$. Значение статистики для повторной выборки из таблицы 1.1 равно:

$$B_{20} = \frac{2 \cdot 20 \cdot \left[\ln\left(\frac{218.9}{20}\right) - \frac{1}{20} (\sum_{i=1}^{20} \ln X_i) \right]}{1 + \frac{21}{120}} = 15.42.$$

Критерий Бартлетта является двусторонним. Критические значения для него при $n = 20$ и $\alpha = 0.05$ имеют вид

$$B_{20; 0.05} = 10.12, \quad B_{20; 0.95} = 30.14.$$

Полученное значение статистики для выборки попадает в интервал $[10.12, 30.14]$ между квантилями. Следовательно, выборочные данные из табл. 1.1 не противоречат гипотезе экспоненциальности. ■

В большинстве случаев при проверке гипотезы H_0 альтернатива неизвестна. Тогда выполняется проверка на согласованность выборочных данных гипотезе экспоненциальности. Но иногда класс альтернатив сужается,

например, до конкретного распределения с известными параметрами, либо до класса альтернатив с возрастающей или убывающей функциями интенсивности отказов $h(x) = F'(x)/(1 - F(x))$ (соответственно ВФИ или УФИ).

Если класс альтернатив H_1 сузить до семейства распределений Вейбулла

$$We(\theta) = \{F: F(x, \theta) = 1 - e^{-\lambda x^k}, \theta = (k, \lambda), x \geq 0, k > 0, \lambda > 0\}, \quad (1.13)$$

то гипотезу H_0 можно записать, как

$$H_0: F \in \mathcal{F}_0 \equiv \{We(\theta), x \geq 0, k = 1, \lambda > 0\}, \quad (1.14)$$

т. е. H_0 при данной альтернативе равносильна утверждению, что $F \in We(\theta)$ при $k = 1$.

Далее будут рассмотрены критерии проверки гипотезы экспоненциальности, свободные от масштаба. Для этого удобно перейти к нормированным случайным величинам $U_i = X_i/\bar{X}$, $i = 1, \dots, n$, где \bar{X} – выборочное среднее случайных величин из выборки X :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}. \quad (1.15)$$

Количество критериев, свободных от масштаба, сложно точно подсчитать, но однозначно их существует несколько десятков. По этой причине выделяются некоторые *классы статистик* для проверки гипотезы H_0 , что даёт возможность с единых позиций рассматривать свойства различных критериев. Один из примеров подобных классификаций рассматривается в разделе 1.4.

Некоторые статистики используют случайные величины повторной выборки, упорядоченные по возрастанию. Они называются *порядковыми статистиками* выборки X . Далее они будут обозначаться, как $X_{(i)}$, $i = 1, \dots, n$. Также положим $U_{(i)} = X_{(i)}/\bar{X}$.

Пример 1.2. Пусть имеется повторная выборка $X = (5, 2, 7, 1)$ и $X_{(i)}$ – i -я порядковая статистика выборки X . Тогда $X_{(1)} = 1$, $X_{(2)} = 2$, $X_{(3)} = 5$, $X_{(4)} = 7$, $\bar{X} = 3.75$, $U_{(1)} \approx 0.267$, $U_{(2)} \approx 0.533$, $U_{(3)} \approx 1.333$, $U_{(4)} \approx 1.867$.

Далее рассматривается весь набор статистик, включенных в разрабатываемый продукт:

1. *Критерий Морана.* Статистика является одним из первых предложенных критериев для проверки гипотезы экспоненциальности. Является оптимальной свободной от масштаба статистикой против гамма-альтернатив [10]. Данная статистика имеет вид

$$M_n = n^{-1} \sum \log U_i + C, \quad (1.16)$$

где $C = -\Gamma'(1) = 0.5772$ – постоянная Эйлера.

2. *Критерий Гринвуда.* Статистики имеет асимптотически нормальное распределение (после умножения на \sqrt{n}).

Данная статистика имеет вид

$$Gr_n = 2 - n^{-1} \sum U_i^2. \quad (1.17)$$

3. *Критерий Кокса-Оукса* [7]. Данная статистика имеет вид

$$CO_n = 1 + n^{-1} \sum (1 - U_i) \log U_i. \quad (1.18)$$

Данную функцию в литературе называют *функцией вклада* [5, с. 47]. Она также используется в критерии Рао.

4. *Критерий Джини* [20]. Распределение статистики асимптотически нормальное. По причине того, что критерий Джини симметричен относительно точки $x = 0.5$, квантили u_α нижнего хвоста распределения $F(x)$ вычисляются согласно равенству $u_\alpha = 1 - u_{1-\alpha}$.

Статистика имеет вид

$$G_n = \frac{\sum_{i=1}^n \sum_{j=1}^n |U_j - U_i|}{2n(n-1)}, \quad (1.19)$$

Также формулу (1.19) можно привести к более удобному для моделирования виду:

$$\mathbf{G}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} i(n-i)(U_{(i+1)} - U_{(i)}). \quad (1.20)$$

В случае, когда $H_1 = We(k)$, отклонение значения статистики меньше нижней квантили u_a свидетельствует о том, что $k > 1$, а значит функция интенсивности распределения возрастающая. Аналогично, отклонение больше верхней квантили u_{1-a} свидетельствует, что $k < 1$ и распределение имеет УФИ.

5. *Критерий Джексона* [22]. Статистика имеет вид

$$\mathbf{J}_n = n^{-1} \sum_{i=1}^n q_i U_{(i)}, \quad (1.21)$$

где $q_i = \sum_{k=1}^i \frac{1}{n-k+1}$.

6. *Критерий Фортиана-Гране* [17]. Данная статистика имеет вид

$$\mathbf{FG}_n = n^{-1} \sum \omega_i U_{(i)}, \quad (1.22)$$

где $\omega_i = (n-i) \log(n-i) - (n-i+1) \log(n-i+1) + \log n$.

7. *Критерий Фроцини* [18]. Данная статистика имеет вид

$$\mathbf{Fr}_n = n^{-\frac{1}{2}} \sum_{i=1}^n \left(|1 - e^{-U_{(i)}}| - n^{-1} \left(i - \frac{1}{2} \right) \right). \quad (1.23)$$

8. *Критерий Шермана* [28]. Данная статистика имеет вид

$$\mathbf{Sh}_n = (2n)^{-1} \sum |U_i - 1|. \quad (1.24)$$

9. *Критерий Шапиро-Уилка* [29]. Данная статистика имеет вид

$$\mathbf{SW}_n = n^{-2} \sum (U_i - 1)^2. \quad (1.25)$$

10. *Критерий Бартлетта*. Данный критерий не относится к числу наиболее часто применяемых на практике. Однако в [5, с. 257] указывается на тот факт, что он является наиболее мощным по сравнению с другими су-

ществующими критериями для обнаружения возрастающей или убывающей интенсивности отказов.

Статистика имеет вид

$$B_n = \frac{2}{1+(n+1)/6n} \sum \log U_i. \quad (1.26)$$

11. *Критерий Энпса-Пулли* [16]. Данная статистика имеет вид

$$EP_n = \sqrt{48n} \left(\frac{1}{n} \sum e^{-U_i} - \frac{1}{2} \right). \quad (1.27)$$

12. *Критерий Стивенса* [30]. Является модификацией статистики Шапиро-Уилка. Данная статистика имеет вид

$$St_n = \frac{1}{\sum U_i^2 + n^{-1} \sum (U_i^2 - 1)}. \quad (1.28)$$

13. *Критерий Гнеденко* [2, с. 249]. Данная статистика имеет вид

$$Gn_n = \frac{n-r}{r} \cdot \frac{V_r}{1-V_r}, \text{ где } V_r = n^{-1} \sum_{i=1}^r U_i. \quad (1.29)$$

14. *Критерий Кимбера-Мичела* [23]. Данная статистика имеет вид

$$KM_n = \frac{2}{\pi} \max_i |s_i - r_i|, \quad (1.30)$$

где $s_i = \arcsin \sqrt{1 - e^{-U_i}}$, $r_i = \arcsin \sqrt{\frac{i-0.5}{n}}$.

15. *Критерий Рао* [9]. Наряду с критериями Неймана-Пирсона и Вальда, является одним из классических критериев. В отличие от двух других, полностью определяется случайными величинами $\{U_i\}_{i=1}^n$.

Статистика Рао в случае, когда выполняется проверка на экспоненциальность, имеет вид

$$R_n = \frac{nV_0^2}{1+n^{-1} \sum U_i (\log U_i)^2 - (n^{-1} \sum U_i \log U_i)^2}, \quad (1.31)$$

где $V_0 = 1 + n^{-1} \sum (1 - U_i) \log U_i$ — функция вклада, также используемая, как статистика Кокса-Оукса.

Таблица 1.2. Результаты проверки гипотезы экспоненциальности по выборочным данным из примера 1.1 с помощью ряда критериев, представленных в данной работе. $\alpha = 0.05$.

Статистика	Значение статистики	Квантили статистики		Результат
Морана	0.1245	-0.2863	0.2863	принимается
Гринвуда	1.5907	1.4558	2.5933	принимается
Фроцини	0.2159	0.3813 (односторонняя)		принимается
Шермана	0.3166	0.2281	0.4491	принимается
Шапиро-Уилка	0.0295	0.0228	0.0794	принимается
Кимбера-Мичела	0.0978	0.1634 (односторонняя)		принимается
Джини	0.4419	0.3879	0.6121	принимается
Рао	0.9124	3.7863 (односторонняя)		принимается
Кокса-Оукса	0.2546	-0.4444	0.4569	принимается
Фортиана-Гране	0.9640	0.6188	1.1873	принимается

Пример 1.3. Рассмотрим задачу из примера 1.1, используя другие критерии для проверки гипотезы H_0 . В таблице 1.2 представлены результаты проверки выборочных данных из примера 1.1 на экспоненциальность с помощью десяти критериев. Опираясь на тот факт, что для каждой из статистик из табл. 1.2 гипотеза H_0 принимается, можно считать, что выборочные данные из примера 1.1 на противоречат гипотезе экспоненциальности. ■

Таблица 1.3. Выборочные данные из примера 1.4.

n	1	2	3	4	5	6	7	8	9	10
	0.373	0.704	0.314	1.155	0.604	1.605	0.984	2.04	0.790	1.427

Пример 1.4. Пусть дана повторная выборка, представленная в табл. 1.3. Результаты проверок гипотезы H_0 данной выборке с помощью нескольких критериев представлены в табл. 1.4. Анализ данной таблицы показывает, что для разных статистик результаты проверки гипотезы по одной и той же выборке противоречат друг другу. В таких случаях можно проводить более глубокое изучение используемых статистик. Например, когда задано альтернативное распределение, имеет смысл выбирать наиболее мощный критерий, или критерий с максимальной *асимптотически оптимальной эффективностью*[10] для данной альтернативы. Также к менее противоречивым результатам может привести увеличение объема выборки при повторении эксперимента. ■

Таблица 1.4. Результаты проверки гипотезы экспоненциальности по выборочным данным из примера 1.4 с помощью ряда критериев, представленных в данной работе. $\alpha = 0.05$.

Статистика	Значение статистики	Квантили статистики		Результат
Морана	0.4211	-0.3802	0.3802	отклоняется
Гринвуда	1.2807	1.3112	2.6672	отклоняется
Фроцини	0.3443	0.3730 (односторонняя)		принимается
Шермана	0.2228	0.2281	0.4491	отклоняется
Шапиро-Уилка	0.0280	0.0309	0.1643	отклоняется
Кимбера-Мичела	0.2043	0.2084 (односторонняя)		принимается
Джини	0.3313	0.3385	0.6574	отклоняется
Рао	3.9663	3.7292 (односторонняя)		отклоняется
Кокса-Оукса	0.7033	-0.5775	0.6384	отклоняется
Фортиана-Гране	0.4838	0.4754	1.1639	принимается

1.4. Асимптотически оптимальная эффективность

Помимо мощности, для оценки качества критерия против определенной альтернативы можно использовать асимптотически оптимальную эффективность (АОЭ) по Бахадуру. Чем выше АОЭ, тем лучше критерий улавливает отклонение H_0 от H_1 при их сближении.

Суть вычисления АОЭ по Бахадуру для некоторой последовательности $\{T_n\}$ состоит в следующем. Пусть $T_n \rightarrow b(\theta)$ по P_0 -вероятности, $\theta \in \Theta_1$, где $-\infty < b(\theta) < \infty$, и $k(a)$ – функция уклонений. Тогда величина

$$c_T(\theta) = -2k(b(\theta)) \quad (1.32)$$

называется точным наклоном по Бахадуру. При всех $\theta \in \Theta_1$

$$c_T(\theta) \leq 2Q(\theta), \quad (1.33)$$

где

$$Q(\theta) = \inf_{\theta_0 \in \Theta_1} \{K(\theta, \theta_0) : \theta_0 \in \Theta_0\}. \quad (1.34)$$

Здесь $K(\theta, \theta_0)$ – информация Кульбака–Лейблера [8, с. 19]. Если в (1.33) достигается знак равенства, то последовательность $\{T_n\}$ называется *асимптотически оптимальной* по Бахадуру.

Асимптотической эффективностью по Бахадуру последовательности $\{T_n\}$ является отношение точного наклона к его потенциальной границе:

$$e_T(\theta) = \frac{c_T(\theta)}{2Q(\theta)}. \quad (1.35)$$

В [10] предложена классификация критериев проверки гипотезы экспоненциальности, свободных от масштаба, состоящая из трех групп.

В первой группе представлены статистики критериев вида $T_n = n^{-1} \sum_{i=1}^n r(U_i)$, где r – некоторая функция, заданная на положительной полуоси. Примеры статистик из первой группы представлены в табл. 1.5.

Таблица 1.5. Примеры статистик из первой группы классификации, предложенной в [10].

Здесь $U_i = X_i/\bar{X}$, $C = -\Gamma'(1) = 0.5772$ – постоянная Эйлера.

Статистика Гринвуда	Статистика Морана	Статистика Кокса–Оукса
$2 - n^{-1} \sum U_i^2$	$n^{-1} \sum \log U_i + C$	$1 + n^{-1} \sum (1 - U_i) \log U_i$

Вторая группа рассматриваемой классификации содержит *нормированные L-статистики*, т. е. линейные комбинации порядковых статистик $U_{(i)}$ вида $T_n = n^{-1} \sum_{i=1}^n w_i U_{(i)}$, где w_i – некоторые коэффициенты. Примеры статистик из второй группы представлены в табл. 1.6.

Таблица 1.6. Примеры статистик из второй группы классификации, предложенной в [10].

Здесь $U_{(i)} = X_{(i)}/\bar{X}$, $p_i = \frac{2(i-1)}{n-1} - \frac{3}{2}$, $q_1 = \mathbb{M}X_{(i)} = \frac{1}{n-i+1}$, $\omega_i = (n-i) \log(n-i) - (n-i+1) \log(n-i+1) + \log n$. В качестве статистики Джини рассматривается ее центрированный ва-

$$\text{риант } T_n = G_n - \frac{1}{2}.$$

Статистика Джини	Статистика Джексона	Статистика Фортиана–Гране
$n^{-1} \sum_{i=1}^n p_i U_{(i)}$	$n^{-1} \sum_{i=1}^n q_i U_{(i)}$	$n^{-1} \sum \omega_i U_{(i)}$

В третью группу входят критерии вида $T_n = \sup_{x \geq 0} n^{-1} \sum_{i=1}^n r(U_i, x)$, где $r(s, x)$ – функция, определенная на первом квадранте. Сюда относятся

критерии Лилефорса[25], Барингхауза-Хенце[13], а также их взвешенные варианты. В данной работе они не рассматриваются.

Преимуществом данного подхода является возможность вычисления АОЭ не для каждой отдельной статистики, а для всей группы. В рамках данной работы не стоит задачи самостоятельно вычислять значения e_T для используемых статистик. При этом в программный продукт имеет смысл добавить уже посчитанные в [10] значения АОЭ для некоторых статистик. Данные значения представлены в таблице 1.7. В ней также присутствуют значения АОЭ для статистики Кокса-Оукса, вычисленные в [3].

Таблица 1.7. Значения АОЭ для ряда статистик против некоторых альтернатив.

H_1	Статистики для проверки гипотезы H_0					
	Гринвуд	Джексон	Джини	Кокс-Оукс	Моран	Фортиана-Гране
We	0.6079	0.6079	0.8762	1.0	0.9426	0.6079
Ga	0.3876	0.3876	0.6941	0.9426	1.0	0.3876

Анализ таблицы 1.7 показывает, что для We -альтернативы наибольшей АОЭ среди предложенных статистик обладает критерий Кокса-Оукса, для Ga -альтернативы – критерий Морана. Кроме того статистики Гринвуда, Джексона и Фортиана-Гране имеют одинаковые значения АОЭ по Бахадуру для данных альтернатив. Это означает, что необходимо более детальное исследование их характеристик при получении противоречивых результатов этими статистиками.

Пример 1.5. В примере 1.4 для различных статистик результаты проверок гипотезы H_0 противоречат друг другу. Для того, чтобы предпочесть один из имеющихся в табл. 1.4 статистик для принятия окончательного решения, рассмотрим характеристики этих статистик с точки зрения АОЭ по Бахадуру. Пусть в качестве альтернативы H_1 задано семейство распределений We . Из предложенных статистик значение АОЭ для данной альтернативы имеется для критериев Гринвуда, Морана, Джини, Кокс-Оукса и Фортиана-Гране. Для остальных критериев информации о их АОЭ по Бахадуру не имеется. Максимальную АОЭ среди данных статистик имеет критерий Кокса-Оукса, следовательно, если учитывать данную характеристику статистик, при условии альтернативы $H_1 = We(k)$ стоит

предпочесть результат, полученный данной статистикой. Для критерия Кокса-Оукса в примере 1.4 гипотеза экспоненциальности отклоняется. ■

1.5. Статистическое моделирование

Во многих источниках, посвященных гипотезе экспоненциальности (например, в [6]), приводятся таблицы квантилей статистик для ее проверки. Для таких таблиц квантили считаются статистическим моделированием методом Монте-Карло. Этот способ имеет точность вычислений всего несколько знаков после запятой, которая намного меньше той, что достигается при аналитическом подходе, но для практических надобностей и для высказывания тех или иных теоретических предположений, она оказывается достаточной. Подобные таблицы на данный момент имеются далеко не для всех статистик, поэтому одним из назначений разрабатываемого продукта является моделирование неизвестных квантилей для предложенных в нем критериев.

Метод Монте-Карло. Данный метод статистического моделирования используется во многих областях науки. Его популярность объясняется тем, что он очень прост в реализации.

Данный метод, в случае моделирования квантилей для статистик проверки гипотезы экспоненциальности, состоит из нескольких шагов:

1. моделируются повторные выборки заданного объема n , извлекаемые из экспоненциального распределения,
2. для смоделированных выборок считается значение статистики, которое записывается в некоторый массив,
3. 1-й и 2-й шаги повторяются заданное количество итераций (испытаний),
4. полученные значения статистик из данного массива упорядочиваются по возрастанию. Квантилями считаются значения в массиве с индексами an и $(1 - a)n$ (a и $1 - a$ квантили соответственно), где a – выбранный уровень значимости.

Далее производится сравнение смоделированных методом Монте-Карло квантилей статистики Джини с их аналитически вычисленными аналогами. Для статистики Джини известно точное выражение её функции распределения $F(x)$. Поэтому можно аналитически вычислить точные значения ее квантилей. Таблица с вычисленными значениями статистики приводится в [20].

Таблица 1.8. Сравнение смоделированных значений квантилей статистики Джини с их точными аналогами. n – объем выборок, M – количество испытаний.

n	M	Значения квантилей		
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
10	100	0.6111	0.6762	0.7292
	1000	0.6529	0.6952	0.7439
	10000	0.6596	0.6884	0.7202
	50000	0.6590	0.6875	0.7198
	100000	0.6589	0.6875	0.7198
	точное	0.65855	0.68768	0.72070
12	100	0.6394	0.6522	0.6549
	1000	0.6346	0.6606	0.7006
	10000	0.6444	0.6662	0.6994
	50000	0.6434	0.6733	0.7006
	100000	0.6434	0.6702	0.7014
	точное	0.64337	0.66992	0.70020
14	100	0.6143	0.6786	0.6635
	1000	0.6296	0.6557	0.6782
	10000	0.6282	0.6561	0.6801
	50000	0.6322	0.6576	0.6846
	100000	0.6315	0.6564	0.6833
	точное	0.63185	0.65641	0.68448
16	100	0.6528	0.6272	0.6874
	1000	0.6165	0.6445	0.6692
	10000	0.6219	0.6438	0.6718
	50000	0.6224	0.6451	0.6727
	100000	0.6223	0.6457	0.6712
	точное	0.62273	0.64567	0.67197

В таблице 1.8 представлены смоделированные квантили статистики Джини. Сравним результаты моделирования для разного количества испыта-

ний N . Так как G_n симметрична относительно точки $x = 0.5$, то квантили u_α вычисляются согласно равенству $u_\alpha = 1 - u_{1-\alpha}$.

Анализ табл.1.8 показывает, что с помощью моделирования квантили вычисляются с точностью, примерно равной двум знакам после запятой. При увеличении количества испытаний M точность моделирования растет. При этом для значений $M = 50000$ и $M = 100000$ точность моделирования примерно одинаковая. Это подтверждает закон больших чисел [11], который говорит о том, что, увеличивая N до бесконечности, невозможно получить с помощью моделирования значения квантилей с точностью до любого числа знаков. Из табл. 1.8 видно, что моделирование дает довольно хорошую точность уже при $M = 10000$. Данной точности зачастую достаточно для принятия решений при проверке гипотезы H_0 , поэтому использовать более большие значения количества испытаний при моделировании не всегда целесообразно. В противном случае, при использовании слишком больших значений M , время моделирования квантилей может увеличиться на порядок. Например, для $M = 50000$ моделирование будет занимать в пять раз больше времени, чем для $M = 10000$. При этом точности для этих значений отличаются незначительно.

Таблицы различных статистик приведены в прил. А.

Таблица 1.9. Таблица смоделированных квантилей для статистики Шапиро-Уилка. $N = 10000$.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	0.0378	0.1370	0.0311	0.1645	0.0211	0.2366
11	0.0362	0.1246	0.0302	0.1498	0.0207	0.2052
12	0.0351	0.1125	0.0292	0.1367	0.0206	0.1963
13	0.0330	0.0051	0.0283	0.1253	0.0203	0.1753
14	0.0324	0.0989	0.0275	0.1166	0.0202	0.1593
15	0.0311	0.0915	0.0265	0.1082	0.0195	0.1464
16	0.0298	0.0844	0.0256	0.1012	0.0192	0.1407
17	0.0285	0.0807	0.0248	0.0944	0.0187	0.1309
18	0.0277	0.0756	0.0241	0.0888	0.0185	0.1225
19	0.0273	0.0718	0.0235	0.0839	0.0183	0.1145

В параграфе 1.2 рассматривается понятие *мощности критерия*. Пусть \mathfrak{X} – выборочное множество всех возможных значений статистики T_n , множества \mathfrak{X}_0 и \mathfrak{X}_1 – разбиение множества \mathfrak{X} ($\mathfrak{X}_0 \cup \mathfrak{X}_1 = \mathfrak{X}$, $\mathfrak{X}_0 \cap \mathfrak{X}_1 = \emptyset$), соответствующие H_0 и H_1 . При обилии имеющихся критериев для проверки гипотезы H_0 , чтобы отдать какому-либо из них предпочтение, требуется вычислять мощность критериев $W(T_n) = P\{T_n(X) \in \mathfrak{X}_1 | H_1\}$ и затем делать выбор, т. е. если задан набор из k критериев, то выбирают i -й критерий из условий

$$P\{T_{ni} > u_\alpha | H_0\} = \alpha, \quad P\{T_{ni} > u_\alpha | H_1\} = \max_{1 \leq j \leq k} P\{T_{nj} > u_\alpha | H_1\}. \quad (0.1)$$

По причине того, что мощности критериев вычислить аналитически невозможно, данные значения необходимо моделировать. Для моделирования мощности против заданной альтернативы обычно используется метод Монте-Карло. В данном случае он заключается в следующем:

1. выбирается конкретная альтернатива, против которой для некоторой статистики T_n будет моделироваться мощность. Например, распределение Вейбулла с параметром $k = 1.5$,
2. заводится некоторый счетчик r , изначально равный 0,
3. моделируется повторная выборка X , подчиняющаяся выбранной альтернативе,
4. выборка X проверяется критерием T_n на гипотезу экспоненциальности. Если статистика дает отрицательный результат, и гипотеза отвергается, то счетчик r увеличивается на 1,
5. 3-й и 4-й шаги выполняются заданное количество испытаний M ,
6. счетчик r делится на количество испытаний M . Полученное частное является результатом моделирования мощности статистики T_n против выбранной альтернативы.

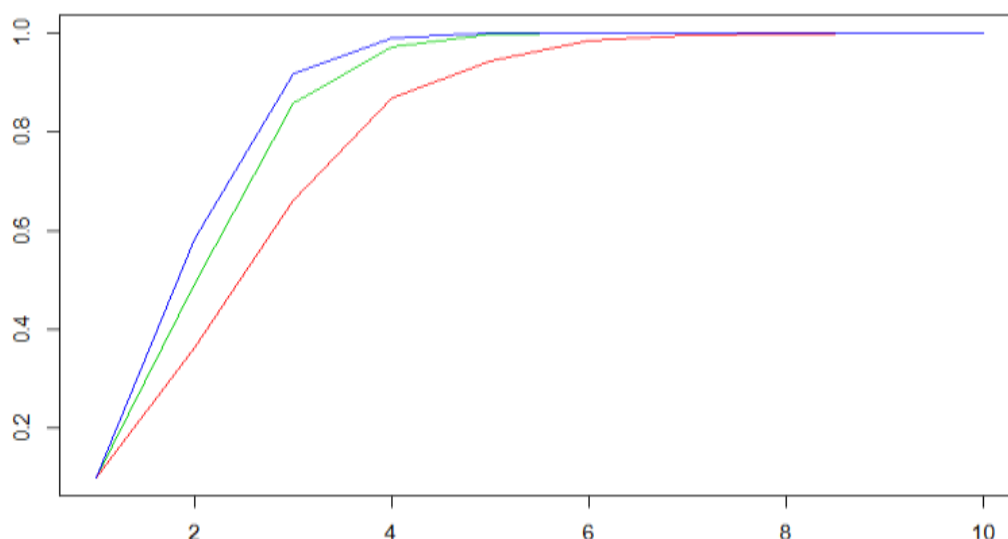


Рисунок 1.3. Графики мощностей статистики Джини против Ga -распределения: красным цветом отображены мощности для $n = 10$, зеленым – для $n = 15$, синим – для $n = 20$. На графике ось ординат – значение мощности, ось абсцисс – значение параметра α Ga -распределения.

Пример 1.6. В табл. 1.10 представлены результаты вычислений мощности критерия Джини для различных альтернатив $H_1 = We(k)$ и объемов выборки n при $\alpha = 0.05$. Анализ таблицы показывает, что при приближении параметра формы k распределения We к значению 1 мощность статистики падает. Также можно сделать вывод, что с ростом n мощность критерия G_n растет, что свидетельствует о его состоятельности.■

Таблица 1.10. Результаты моделирования мощностей против Вейбулла распределения с разным значением параметра k и при разном объеме выборки n для критерия Джини. $\alpha = 0.05$.

k	Значение параметра формы распределения Вейбулла							
	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8
$n = 5$	0.892	0.612	0.36	0.17	0.112	0.145	0.233	0.286
$n = 10$	0.993	0.884	0.581	0.24	0.163	0.282	0.46	0.589
$n = 15$	1	0.962	0.71	0.3	0.167	0.392	0.624	0.822
$n = 20$	1	0.995	0.816	0.393	0.233	0.494	0.762	0.915
$n = 25$	1	0.996	0.868	0.398	0.239	0.589	0.845	0.995
$n = 30$	1	1	0.924	0.44	0.302	0.654	0.921	0.99
$n = 35$	1	1	0.935	0.518	0.329	0.718	0.943	0.993

Таблица 1.11. Результаты моделирования мощностей против

Вейбулла распределения с разным значением параметра k для статистик M_n – Морана, Gr_n – Гринвуда, Fr_n – Фроцини, Sh_n – Шермана, SW_n – Шапиро-Уилка, KM_n – Кимбера-Мичела, G_n – Джини, R_n – Рао. $n = 10$, $\alpha = 0.05$.

H_1	Статистики									
	CO_{10}	FG_{10}	M_{10}	Gr_{10}	Fr_{10}	Sh_{10}	SW_{10}	KM_{10}	G_{10}	R_{10}
$We(0.1)$	1	0.997	1	0.998	1	1	0.997	0.999	1	1
$We(0.25)$	0.999	0.952	0.997	0.952	0.976	0.986	0.954	0.991	0.993	0.998
$We(0.5)$	0.838	0.647	0.848	0.652	0.601	0.731	0.638	0.615	0.729	0.785
$We(1.5)$	0.43	0.305	0.322	0.36	0.244	0.33	0.351	0.226	0.344	0.286
$We(2.0)$	0.826	0.699	0.739	0.777	0.639	0.771	0.757	0.54	0.769	0.68
$We(3.0)$	0.995	0.99	0.994	0.989	0.971	0.992	0.994	0.916	0.993	0.98

Пример 1.7. Как было замечено ранее, в примере 1.4 результаты проверки с помощью разных статистик заданной выборки X на экспоненциальность противоречат друг другу. В примере 1.5 рассматриваются характеристики данных статистик с точки зрения АОЭ по Бахадуру в качестве возможного разрешения неопределенности и выбора более предпочтительного результата. В текущем примере критерии из примера 1.4 рассматриваются с точки зрения их мощности. Рассмотрим шесть случаев, когда в качестве альтернативы задано распределение Вейбулла We с параметром k , равным 0.1, 0.25, 0.5, 1.5, 2.0 и 3.0 соответственно. Для принятия конечного решения в каждом из случаев необходимо выбрать среди критериев наиболее мощный против заданной альтернативы. Результаты моделирования мощностей приведены в табл. 1.11.

Анализ таблицы 1.11 показывает, что среди представленных в ней критериев нет равномерно наиболее мощного. Например, для случаев $We(0.25)$, $We(1.5)$, $We(2)$ и $We(3)$ наиболее мощным оказался критерий Кокса-Оукса. Следовательно, для данных альтернатив следует предпочесть результат, полученный при проверке гипотезы H_0 по выборке X с помощью этого критерия и отклонить H_0 . С другой стороны, если параметр Вейбулла распределения равен $k = 0.5$, наиболее мощным оказался критерий Морана. В текущем примере результатом проверки выборки на экспоненциальность с помощью критерия Морана также является отклонение гипотезы H_0 .

Для случая $We(0.1)$ большинство критериев из таблицы достигает максимального значения мощности, что означает, что вероятность допустить ошибку второго рода при проверке гипотезы H_0 с помощью этих статистик против данной альтернативы минимальна. ■

Имеется множество решений при возникновении ситуации, когда ряд критериев, имеющие одинаковую мощность, которая является максимальной для всего набора рассматриваемых статистик, при проверке гипотезы H_0 по одной и той же выборке ведут себя по-разному. Далее приводятся несколько подобных решений данной проблемы:

- Сделать выбор в пользу результата, полученного наибольшим количеством статистик с максимальной среди них мощностью. В примере 1.7 среди статистик с мощностью $W = 1$ против альтернативы $We(0.1)$ для критериев Морана, Кокса-Оукса, Джини, Рао и Шермана гипотеза H_0 отклоняется, и лишь для критерия Фроцини гипотеза принимается. Следовательно, стоит отклонить гипотезу экспоненциальности в пользу распределения Вейбулла с параметром $k = 0.1$;
- Сделать выбор в пользу критерия с максимальной АОЭ по Бахадuru из ряда статистик, обладающих максимальной мощностью против заданной альтернативы. В примере 1.7, как и в примере 1.5, для альтернативы $We(0.1)$ стоит предпочесть статистику Кокса-Оукса и отклонить гипотезу H_0 , так как среди статистик с максимальной мощностью против данной альтернативы она обладает максимальной АОЭ;
- Увеличить объем выборки, тем самым увеличив мощности критериев, и повторять опыт. В примере 1.7 для альтернативы $We(0.1)$ ряд статистик достигают максимального значения мощности при $n = 10$, поэтому увеличение объема выборки имеет смысл лишь для увеличения мощностей остальных статистик из табл. 1.7.

Как для вычисления квантилей, так и для вычисления мощностей статистик методом Монте-Карло необходимо иметь возможность моделирования повторных выборок, подчиненных некоторому распределению (для моделирования квантилей выборочные данные извлекаются из экспоненциального распределения, для моделирования мощностей – из различных альтернативных семейств распределений, например Вейбулла). В разработанном

программном продукте предложены инструменты решения данной проблемы.

Алгоритмы моделирования н. о. р. с. в. хорошо изучены и в ряде случаев достаточно легко реализуемы. В случае экспоненциального и Вейбулла распределений для моделирования случайных величин используется метод обратной функции. Для моделирования некоторых распределений обычно используется процедура суммирования с. в., извлекаемых из других распределений. Например, для моделирования χ^2 -распределенных с. в. происходит суммирование квадратов с. в., имеющих нормальное распределение.

Для генерирования равномерно распределенных с. в. на интервале $[0, 1]$, часто обозначаемых в литературе, как $U[0,1]$, используются различные датчики. Такое равномерное распределение называют *стандартным*. Для равномерно распределенных случайных чисел в интервале от 0 до 1 выполняются следующие утверждения:

$$MX = \frac{1}{2}, \mathbb{D}X = \frac{1}{12}, X \in U[0, 1]. \quad (1.36)$$

Стоит учесть тот факт, что большинство датчиков генерируют лишь псевдослучайные числа, которые лишь имитируют с. в., имеющие равномерное распределение. Вычисление нового псевдослучайного числа в таких датчиках происходит по некоторому рекуррентному алгоритму. Поэтому такие датчики называются *алгоритмическими*. Последовательности, составленные из таких чисел, образуют петли, то есть обязательно существует цикл, повторяющийся бесконечное количество раз. Достоинством данных датчиков является быстроедействие, так как генераторы практически не требуют ресурсов памяти и очень компактны. Их недостатком можно назвать тот факт, что числа, сгенерированные этими датчиками, нельзя в полной мере назвать случайными, поскольку между ними имеется зависимость, а также наличие периодов в их последовательности.

Также, помимо алгоритмических датчиков, существуют так называемые *физические* датчики случайных величин, использующие для генерации чисел внешние источники. Такими источниками могут являться различные физические шумы. Хотя числа, генерируемые подобными датчиками, могут быть действительно случайными, они используются довольно редко по причине необходимости больших затрат для их реализации и возможных искажений распределения из-за внешних факторов.

При выборе датчика необходимо обращать внимание на качество генерируемых им чисел. В наши дни большинство средств разработки программного обеспечения содержат в себе датчики, отвечающие современным требованиям.

Глава 2. ПРОГРАММНЫЙ ПРОДУКТ

Главной целью данной работы является создание программного продукта, позволяющего выполнять проверку выборочных данных на экспоненциальность. В данной главе рассматриваются архитектура продукта, различные технологии, использованные при его разработке, а также примеры его использования на реальных задачах.

2.1 Структура продукта

Разработанный продукт подразделяется на две части:

1. основная,
2. упрощенная.

1.Основная часть. Данная часть является веб-приложением. В виде сайта она предоставляет услуги пользователям через сеть *Internet*. Как и любой другой веб-сервис, приложение имеет серверную и клиентскую части. Анализ статистических данных происходит на стороне сервера.

Для создания части приложения, выполняющейся на стороне сервера, используется объектно-ориентированный язык *Java*. Данный выбор объясняется популярностью языка в области веб-разработки, а также наличием большого количества различных сторонних фреймворков и библиотек в свободном доступе. При разработке продукта были использованы следующие библиотеки:

- Spring Framework – универсальный фреймворк с открытым исходным кодом, разработанный компанией Pivotal Software. В нем имеется множество инструментов для разработки разного рода программного обеспечения. Главным модулем фреймворка является контейнер IoC (Inversion of Control), управляющий жизненным циклом специальных объектов *Beans*. Главной причиной выбора данного фреймворка является наличие собственной MVC-платформы (Model-View-Controller), цель которой отделить бизнес-логику приложения от её визуализации.

- Hibernate – библиотека с открытым исходным кодом, разработанная компанией Red Hat. Библиотека реализует подход ORM (object-relational mapping – объектно-реляционное отображение), т. е. подход отображения объектно-ориентированной модели данных в реляционные базы данных. Использование данной библиотеки требуется для хранения квантилей и мощностей статистик в базе данных.
- Gson – библиотека для языка Java с открытым исходным кодом, разработанная компанией Google. Библиотека предназначена для сериализации объектов языка Java в текстовый формат сообщений *JSON* (JavaScript Object Notation), а также для десериализации сообщений обратно в объекты.

Для создания клиентской части был использован стандартный набор инструментов, применяемый при подобных разработках, в виде языка разметки *HTML*, языка описания стилей *CSS* и скриптового языка программирования *Javascript*. Также были использованы сторонние фреймворки и библиотеки:

- Bootstrap – фреймворк для веб-разработки, созданный компанией Twitter. Bootstrap содержит в себе *HTML* и *CSS* шаблоны, а также *Javascript* расширения, нацеленные на экономию времени и усилий разработчика.
- jQuery – библиотека, фокусирующаяся на взаимодействии *JavaScript* и *HTML*. Она помогает легко получать доступ к любому элементу страницы, обращаться к их атрибутам и содержимому.
- HighCharts – библиотека для отображения графических данных на веб-странице.

В приложении присутствуют:

- имитационный блок, предназначенный для генерации повторных выборок различных распределений,
- блок моделирования квантилей и мощностей статистик,

- блок, выполняющий проверку выборочных данных с помощью заданного списка статистик.

Преимущества использования веб-части продукта:

- Удобный визуальный интерфейс.
- Отсутствие необходимости скачивать специальное программное обеспечение. Для полноценного использования достаточно иметь установленный веб-браузер.
- Отсутствие необходимости локально хранить большие объемы таких дополнительных данных, как квантили и мощности статистик. Они хранятся в базе данных приложения.
- Постоянное пополнение базы данных.
- Техническая поддержка. Отсутствие необходимости следить за обновлениями, все изменения в процессе доработки приложения производятся на серверной стороне.
- Все вычисления выполняются на серверной стороне.

Недостатки использования веб-части продукта:

- Для использования приложения требуется доступ в сеть Internet.
- Сервисом нельзя пользоваться без регистрации и последующей авторизации.
- Затраты на пересылку данных на сервер и обратно клиенту.

Для эксплуатации приложения используются следующие сторонние программы:

- Tomcat – сервер, разработанный компанией Apache. Он является контейнером *сервлетов* (классов языка Java, расширяющих возможности сервера, которые предназначены для взаимодействия с пользователем по принципу запрос-ответ), т. е. программой, занимающейся системной поддержкой сервлетов и обеспечивающей их жизненный цикл. Tomcat необходим для запуска веб-приложения и развертывания его в сети.

- MySQL – свободная реляционная система управления базами данных, разработанная компанией Sun Microsystems. Программа обеспечивает поддержку базы данных для веб-приложения.

Данная часть разработанного продукта имеет структуру, типичную для большинства веб-приложений, разработанных с применением концепции MVC, описанной в [24]. Основной целью применения этой концепции заключается в разделении приложения на три составляющие:

- модель,
- представление,
- контроллер.

Контроллер необходим для взаимодействия с приложением и, используя модель и представление, генерирования некоторой реакции на действия пользователя. В качестве контроллера в разработанном приложении выступает класс *ApplicationController*.

В обязанности представления входит отображение данных полученных от модели. При этом представление не может влиять на модель. В качестве представления в приложении используется набор JSP-страниц.

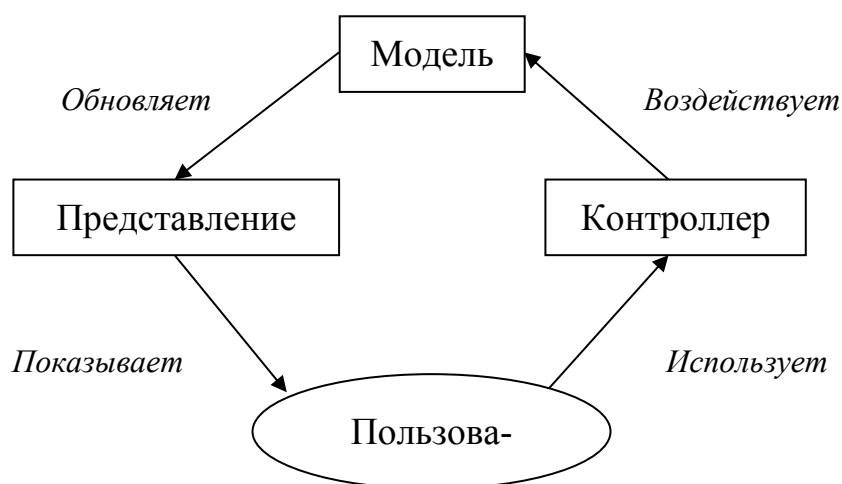


Рисунок 2.1. Диаграмма зависимостей при MVC-подходе для разработки программного обеспечения.

Под моделью обычно понимается часть, содержащая в себе функциональную бизнес-логику приложения. Представление и контроллер зависят от модели. Однако модель не зависит ни от представления, ни от контроллера. Модель предоставляет некоторый интерфейс контроллеру для взаимодействия, и представлению для доступа к данным. Диаграмма зависимостей составляющих программы при использовании MVC-подхода изображена на рис. 2.1.

Модель – самая крупная составляющая приложения, которая сама состоит из нескольких частей:

1. *ApplicationService* – интерфейс, предоставляющий доступ к модели для представления и контроллера. В качестве реализации интерфейса выступает класс *ApplicationServiceImpl*.

2. *Logic* – пакет, содержащий в себе реализации статистик и распределений вероятностей, а также различных вспомогательных математических функций и констант. В данном пакете за реализации статистик отвечает класс *ExpTestsExecutor*, за реализации с. в. распределений – класс *SelectionModulator*.

3. *Entities* – пакет, который содержит реализации различных сущностей, необходимых в приложении. В него входят классы объектов, предназначенных для сериализации библиотекой Gson и отправки информации клиентской части приложения:

- *ChartArray* – ассоциативный массив, использующийся для отображения графиков мощностей для статистик против разных альтернатив;
- *IterationsCountsSet* – множество возможных значений количества испытаний N для моделирования квантилей статистик. Также в структуре содержится значение N , используемое для моделирования по умолчанию;
- *QuantilesMap* – множество критических значений статистики для нескольких объемов повторных выборок и уровней значимости;

- *SelectionInfo* – минимальный набор необходимой информации о выборке для отображения в общем списке выборочных данных пользователя. В данную информацию входят уникальный идентификатор выборки, размер выборки и значение *hash*. Данное значение изменяется при добавлении новых значений в выборку, что позволяет следить за консистентностью данных;
- *SelectionData* – массив значений выборки;
- *TestResult* – информация о результате проверки выборочных данных на экспоненциальность с помощью некоторой статистики. Данная структура хранит в себе значение статистики для выборки и квантили;
- *TestResultsMap* – набор результатов проверки гипотезы по выборке с помощью ряда статистик;
- *FailedActionResult* – сущность для представления информации пользователю при возникновении ошибок на стороне сервера. Структура содержит в себе описание возникшей ошибки;
- *DistributionParams* – интерфейс для представления информации о заданных параметрах распределения. Данный интерфейс имеет классы реализаций для каждого распределения из имитационного блока.

Также в пакет *Entities* входят классы различных исключений, которые используются для обработки исключительных ситуаций в приложении. Исключительной ситуацией считаются некорректный ввод данных (например, значения объема выборок для моделирования квантилей), отсутствие доступа к базе данных, отсутствие необходимых данных в базе, и т. д.

4. *Hibernate* – пакет, предназначенный для доступа к базе данных. Он содержит классы объектов, отображаемых в базу, а также так называемые *DAO (Data Access Objects)*, которые обеспечивают доступ к этим объектам с помощью интерфейсов. В данном пакете приложения имеются следующие сущности для хранения в базе данных:

- *User* – структура для хранения основной информации о пользователе;

- *TestType* – структура предназначена для хранения основной информации о статистике (например необходимость упорядочивать выборки перед проверкой);
- *SignificanceLevel* – структура для хранения допустимых значений уровня значимости;
- *Quantile* – структура для хранения квантилей статистики (или одной квантили для односторонних критериев) при определенном объеме выборки и уровне значимости;
- *Selection* – структура, предназначенная для хранения выборок;
- *Power* – структура для хранения мощности статистики против определенной альтернативы.

5. *TaskManager* – класс, выступающий в роли диспетчера задач. Он позволяет одновременно выполнять несколько различных задач, а также наблюдать за их прогрессом. В качестве задачи может выступать моделирование квантилей или мощностей статистик. Диспетчер выделяет поток для выполнения некоторой задачи и сохраняет результаты её выполнения. В диспетчере имеется класс *ProcessObserver*, с помощью которого можно наблюдать за прогрессом выполнения задач. При этом реализации алгоритмов для задач находятся в пакете *Logic* в классах *ExpQuantilesModulator* и *PowerModulator*.

Принятый объектно-ориентированный подход и использование абстракций позволяет достаточно легко расширять веб-часть продукта дополнительными статистиками и альтернативными распределениями.

2.Упрощенная часть. Данная часть продукта является пакетом, разработанным для платформы языка R. Для его использования пакет необходимо скачать на локальную машину. Анализ данных происходит на стороне клиента.

Данная часть продукта состоит из набора различных функций, также разделенных на три блока: имитация распределений, моделирование кванти-

лей и мощностей статистик, проверка выборочных данных с помощью заданного списка статистик.

Для разработки пакета был выбран язык R, так как он создавался специально для статистической обработки данных и фактически стал стандартом для статистических программ. Для использования языка достаточно загрузить его дистрибутив с официального сайта разработчика и установить на локальную машину. Работу с R можно производить из командной строки. При этом для удобства использования языка разработан ряд графических интерфейсов и программных сред (например RStudio, R Commander). Кроме того, в ряде текстовых и кодовых редакторов предусмотрены специальные режимы для работы с R (например в ConTEXT).

Преимущества использования упрощенной части продукта:

- Сеть *Internet* нужна для того, чтобы загрузить пакет. После этого работа с пакетом производится вне сети.
- Для использования пакета не требуется регистрация.
- Простота применения.

Недостатки использования упрощенной части продукта:

- Отсутствие визуального интерфейса.
- Большие вычислительные затраты. Все функции выполняются на вычислительной машине пользователя.
- Необходимость самостоятельно загрузить его новую версию из сети при обновлении пакета.
- Необходимость хранения таблиц квантилей и мощностей статистик.
- Высокая вероятность возникновения необходимости самостоятельного моделирования квантилей и мощностей статистик из-за отсутствия централизованной пополняемой базы данных.

В языке R функции являются *функциями первого класса*, что позволяет передавать их в качестве аргумента другой функции, а также использовать

так называемые *замыкания функций*. В [19] замыкание определяется, как функция первого класса, в теле которой имеются ссылки на внешние переменные в их лексическом контексте.

Листинг 2.1. Пример функции-замыкания на языке R.

```
closure.factory.func ← function( k ) {  
  closure.func ← function( x ) x^k  
  return( closure.func )  
}  
  
square.func ← closure.factory.func( 2 )  
cube.func ← closure.factory.func( 3 )  
  
square.func( 5 ) # 25  
cube.func( 5 ) # 125
```

Пример 2.1. В листинге 2.1 приводится пример замыканий, возводящих некоторое число в заданную степень. Функция *closure.factory.func* называется *фабрикой*, потому что её результатом выполнения является новая функция *closure.func*, принимающая один параметр и являющаяся замыканием, так как она использует локальную переменную *k* из тела функции *closure.factory.func*. Переменной *square.func* присваивается результат выполнения *closure.factory.func* с параметром 2. Теперь значением *square.func* является функция, возводящая числа в квадрат. Аналогично, при вызове функции *cube.func* с некоторым числом в качестве параметра это число возводится в куб. При этом сигнатура полученных функций одинакова, что позволяет применять общий подход к их использованию. ■

Используя преимущества использования замыканий, в пакете выделяются несколько базовых функций, цель которых выделить общие участки кода, тем самым уменьшив его повторяемость, а также облегчить процесс добавления в него новых статистик и альтернативных распределений. Кроме того различные функции, доступные пользователю из пакета, делегируют ответственность этим базовым функциям за проверку ряда входящих параметров, а также упорядочивание необходимых данных. Примеры данных функций приводятся далее.

2.2. Имитационный блок

В разработанном продукте имеются средства генерирования случайных величин и повторных выборок, подчиненных различным распределениям. Эти инструменты необходимы для моделирования квантилей и мощностей. Среди распределений, доступных в данном продукте, имеются следующие распределения:

- экспоненциальное,
- Вейбулла,
- нормальное,
- Пирсона χ^2 ,
- Фишера F ,
- гамма.

Для моделирования квантилей статистик для проверки гипотезы экспоненциальности используются повторные выборки, подчиняющиеся экспоненциальному закону распределения. Все остальные распределения используются в качестве альтернатив для моделирования мощностей этих статистик.

В веб-приложении имеется возможность сохранять смоделированные повторные выборки в реляционную базу данных, а также загружать в нее собственные статистические данные через имеющийся API-интерфейс и через веб-форму на странице сайта.

В таблице 2.1 представлены некоторые функции API-интерфейса веб-приложения, которые представляют из себя набор инструментов для работы с выборочными данными. Данные функции носят лишь вспомогательный характер при работе с веб-приложением. Их основной целью является предоставление для пользователей возможности хранения и накопления выборочных данных на сервере.

Таблица 2.1. Функции API-интерфейса для работы с выборочными данными.

n	Название	Описание
1	<i>/addselection</i>	Добавление выборочных данных в базу приложения. В качестве параметров запроса выступают: <ul style="list-style-type: none"> - <i>sel_name</i> – идентификатор для выборки; - <i>sel_values</i> – значения выборки, перечисленные через запятую.
2	<i>/appendtoselection</i>	Добавление дополнительных значений к уже существующей выборке. Параметры данного запроса совпадают с <i>/addselection</i> .
3	<i>/removeselection</i>	Удаление выборки пользователя из базы данных по имени. Запрос требует один параметр <i>sel_name</i> .
4	<i>/selection</i>	Загрузка выборки пользователя из базы данных приложения. Запрос требует один параметр <i>sel_name</i> .
5	<i>/modulateselection</i>	Генерирование повторной выборки, извлекаемой из некоторого распределения. В качестве параметров запроса выступают: <ul style="list-style-type: none"> - <i>sel_size</i> – объем извлекаемой выборки. Данный параметр должен иметь положительное значение; - <i>sel_type</i> – уникальный идентификатор распределения. Для каждого из доступных в приложении распределений имеются дополнительные параметры (например для экспоненциального распределения имеется возможность указать параметр масштаба λ); - <i>sorted</i> – необязательный параметр, который указывает на необходимость упорядочить с. в. по возрастанию. - <i>name</i> – идентификатор смоделированной выборки для сохранения в базу данных. Он является необязательным, и в случае его отсутствия выборка не сохраняется.

В разработанном пакете на языке R имеется одна базовая функция *base.rand.selection* для выделения общих частей кода алгоритмов моделирования повторных выборок. Данная функция принимает три аргумента:

- *count* – объем выборки.

- *value.func* – функция для моделирования с. в. нужного распределения. Данная функция не принимает аргументов и является замыканием. Пример реализации такой функции будет приведен далее.
- *sorted* – указание о необходимости упорядочить с. в. по возрастанию.

Представим краткие сведения об использованных в работе распределениях вероятностей.

1. *Экспоненциальное распределение.* Плотность экспоненциального (показательного) распределения определяется следующим образом:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (2.1)$$

где λ – параметр масштаба. Данное распределение обладает следующими свойствами:

$$\mathbb{M}X = \frac{1}{\lambda}, \quad \mathbb{D}X = \frac{1}{\lambda^2}, \quad (2.2)$$

где $\mathbb{M}X$ – математическое ожидание с.в., $\mathbb{D}X$ – дисперсия с. в., имеющей *Exp*-распределение.

В случае, когда $\lambda = 1$, распределение называется *стандартным экспоненциальным (стандартным показательным)*.

Экспоненциальное распределение является частным случаем гамма-распределения $\Gamma(\lambda, \alpha)$ при $\alpha = 1$, где λ – параметр масштаба, а α – параметр формы.

Случайные величины, подчиняющиеся данному распределению, вычисляются по формуле:

$$x = -\frac{1}{\lambda} \ln(1 - u), \quad (2.3)$$

где u – с. в., имеющая непрерывное равномерное распределение на интервале $[0, 1]$.

В разработанном пакете на языке R имеются следующие функции для моделирования экспоненциального распределения:

- *exp.rand.value(scale=1)* – функция возвращает случайную величину, имеющую экспоненциальное распределение с заданным параметром масштаба *lambda* (по умолчанию *scale = 1*).
- *exp.rand.selection(count, scale=1, sorted=false)* – функция, предназначенная для моделирования повторных выборок объема *count*, извлекаемых из экспоненциального распределения с параметром масштаба *scale* (по умолчанию 1). Также есть возможность при моделировании указать на необходимость упорядочить случайные числа по возрастанию (параметр *sorted*, по умолчанию *false*), если выборка моделируется для проверки с помощью критерия, который использует порядковые статистики.

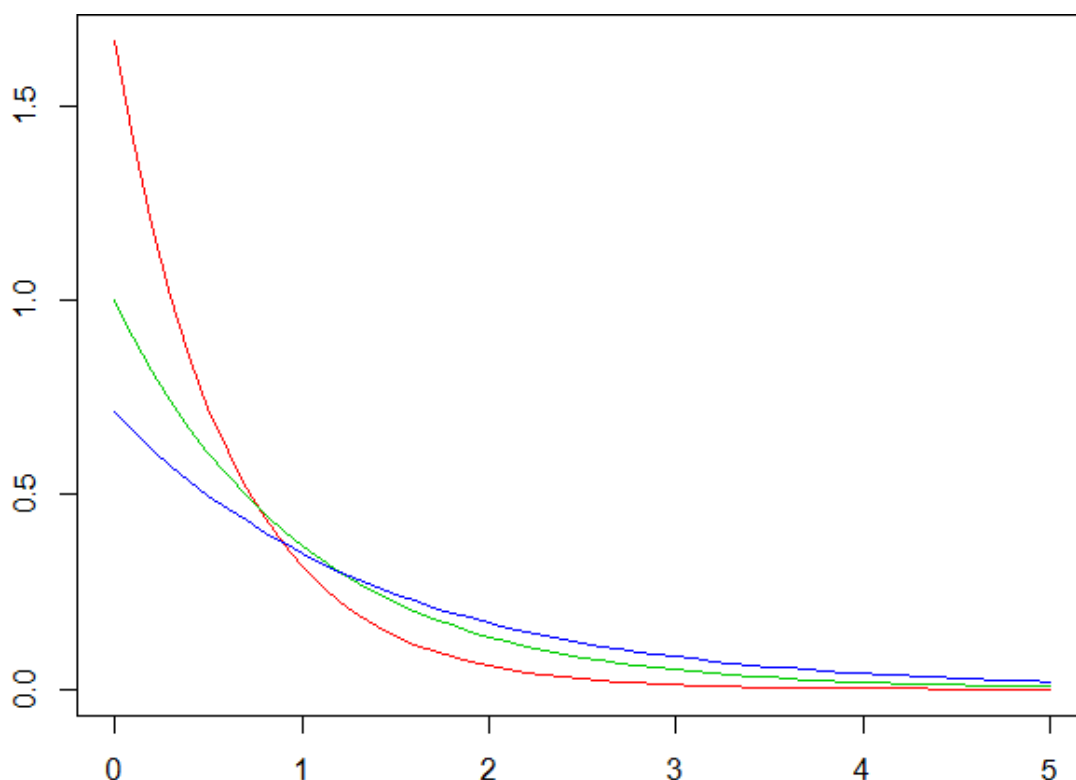


Рисунок 2.2. Графики функций плотности экспоненциального распределения для $\lambda = 0.6$ (синий), для $\lambda = 1$ (зеленый) и для $\lambda = 1.4$ (красный).

2. *Распределение Вейбулла.* Плотность распределения Вейбулла определяется следующим образом:

$$f(x) = \begin{cases} k\lambda^k x^{k-1} e^{-(\lambda x)^k}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (2.4)$$

где λ – параметр масштаба, k – параметр формы. При $k = 1$ распределение (2.3) является экспоненциальным. Данное распределение обладает следующими свойствами:

$$\mathbb{M}X = \lambda \Gamma\left(1 + \frac{1}{k}\right), \mathbb{D}X = \lambda^2 \Gamma\left(1 + \frac{2}{k}\right) - (\mathbb{M}X)^2 \quad (2.5)$$

где $\mathbb{M}X$ – математическое ожидание с.в., $\mathbb{D}X$ – дисперсия с. в., имеющей *We*-распределение.

Случайные величины, имеющие распределение Вейбулла, вычисляются по формуле:

$$x = \frac{1}{\lambda} \sqrt[k]{-\ln(1 - u)}, \quad (2.6)$$

где u – случайная величина, имеющая равномерное распределение на интервале $[0, 1]$.

Разработанный пакет на языке R содержит следующие функции для моделирования распределения Вейбулла:

- *weibull.rand.value(scale=1, shape=1)* – по аналогии с функцией *exp.rand.value*, данная функция генерирует случайные величины, имеющие распределение Вейбулла с заданными параметром масштаба *scale* (по умолчанию 1) и параметром формы *shape* (по умолчанию 1).
- *weibull.rand.selection(count, scale=1, shape=1, sorted=false)* – аналогична функции *exp.rand.selection*, но при этом моделирует повторные выборки распределения Вейбулла. К списку параметров добавился параметр формы *shape* (по умолчанию 1).

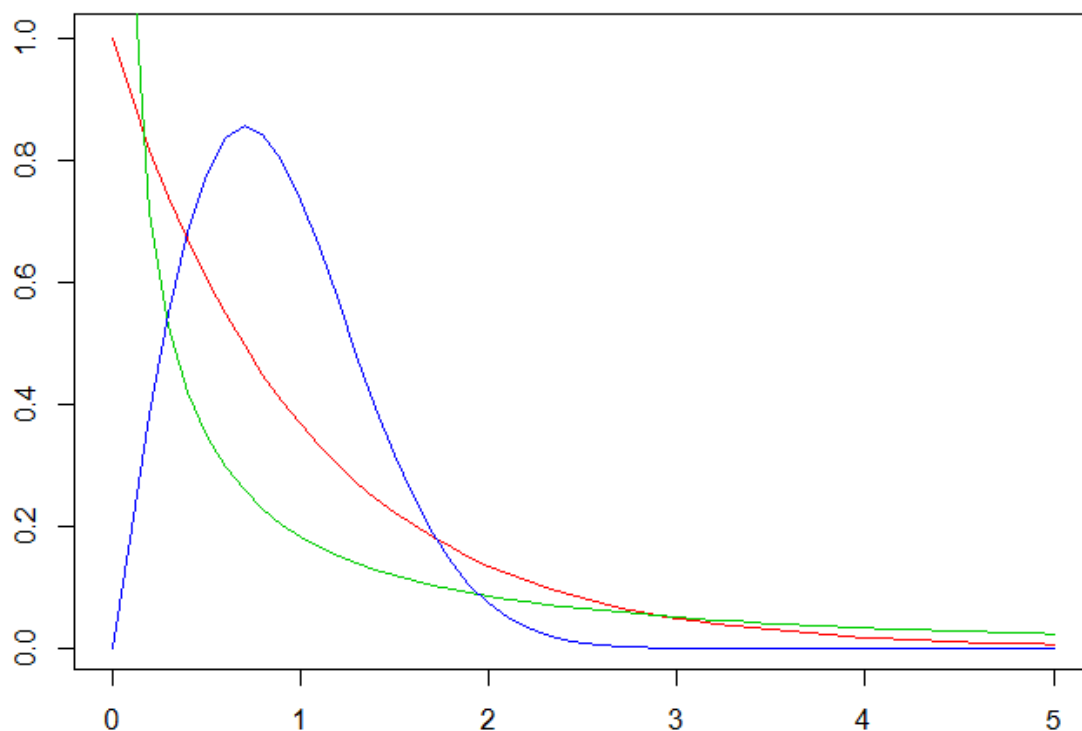


Рисунок 2.3. Графики функций плотности распределения Вейбулла для $\lambda = 1, k = 1$ (красный), для $\lambda = 1, k = 0.5$ (зеленый) и для $\lambda = 1, k = 2$ (синий).

3. *Нормальное распределение.* Функция плотности нормального распределения имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.7)$$

где $\mu = M\xi$ – математическое ожидание (среднее значение), $\sigma = \sqrt{D\xi}$ – среднеквадратическое отклонение с. в. ξ .

Стандартным называется нормальное распределение с параметрами $\mu = 0$ и $\sigma = 1$. В этом случае плотность распределения принимает вид

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (2.8)$$

Для генерирования случайных величин стандартного нормального распределения используется *преобразование Бокса-Мюллера* [14].

Суть данного преобразования состоит в следующем. Пусть u_1 и u_2 – независимые случайные величины, имеющие непрерывное равномерное распределение на интервале $[-1, 1]$. Если величина $s = u_1^2 + u_2^2$ не принадле-

жит интервалу $(0, 1]$, то необходимо сгенерировать величины заново. Если $s \in (0, 1]$, то получаем две случайных величины, имеющих стандартное нормальное распределение, по формулам:

$$z_1 = u_1 f, \quad z_2 = u_2 f, \quad (2.9)$$

где $f = \sqrt{-2 \frac{\ln s}{s}}$. Полученные в (2.9) случайные величины z_1 и z_2 являются независимыми.

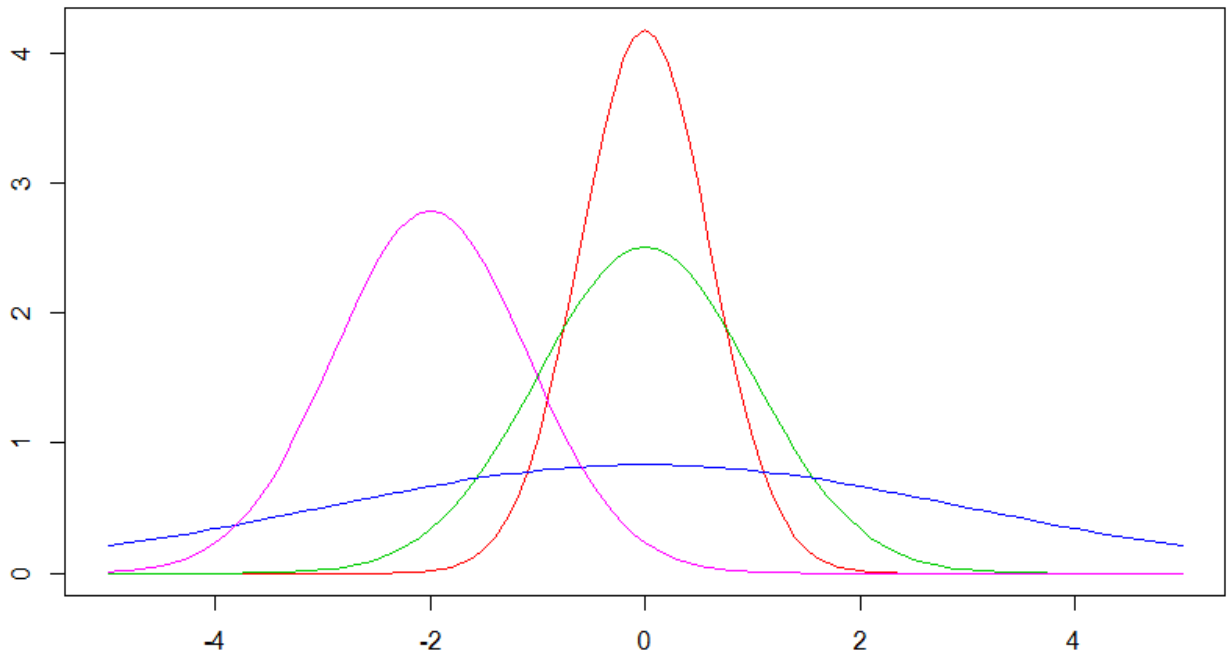


Рисунок 2.4. Графики функций плотности нормального распределения для $\mu = 0, k = 1$ (зеленый), для $\mu = 0, k = 0.6$ (красный), для $\mu = 0, k = 3$ (синий) и для $\mu = -2, k = 0.9$ (фиолетовый).

Далее от них можно перейти к нормальному распределению с математическим ожиданием μ и среднеквадратическим отклонением с. в. σ по формуле

$$t = \mu + \sigma z, \quad (2.10)$$

где z — случайные величины, имеющие стандартное нормальное распределение.

Разработанный пакет на языке R содержит следующие функции для моделирования нормального распределения:

- *normal.rand.value(m=0, d=1)* – данная функция генерирует случайные величины, имеющие нормальное распределение с заданными математическим ожиданием m (по умолчанию 0) и среднеквадратическим отклонением k (по умолчанию 1).
- *normal.rand.selection(count, m=0, d=1, sorted=false)* – функция моделирует повторные выборки, извлекаемые из нормального распределения с математическим ожиданием m (по умолчанию 0) и среднеквадратическим отклонением k (по умолчанию 1) объема $count$.

4. *Распределение Пирсона χ^2* . Данное распределение является распределением суммы квадратов k независимых стандартных нормальных случайных величин, т. е. если z_1, \dots, z_k – совместно независимые стандартные нормальные с. в., то величина

$$x = z_1^2 + \dots + z_k^2 \quad (2.11)$$

имеет χ^2 -распределение с k степенями свободы.

χ^2 -распределение является частным случаем гамма-распределения $\Gamma(\lambda, \alpha)$ при $\lambda = 2$ и $\alpha = \frac{k}{2}$, где λ – параметр масштаба, а α – параметр формы, и имеет функцию плотности

$$f(x) = \Gamma\left(2, \frac{k}{2}\right) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad (2.12)$$

где $\Gamma(\lambda, \alpha)$ – гамма-распределение, а $\Gamma(x)$ – гамма-функция. Данное распределение обладает следующими свойствами:

$$\mathbb{M}X = k, \mathbb{D}X = 2k \quad (2.13)$$

где $\mathbb{M}X$ – математическое ожидание с.в., $\mathbb{D}X$ – дисперсия с. в., имеющей нормальное распределение.

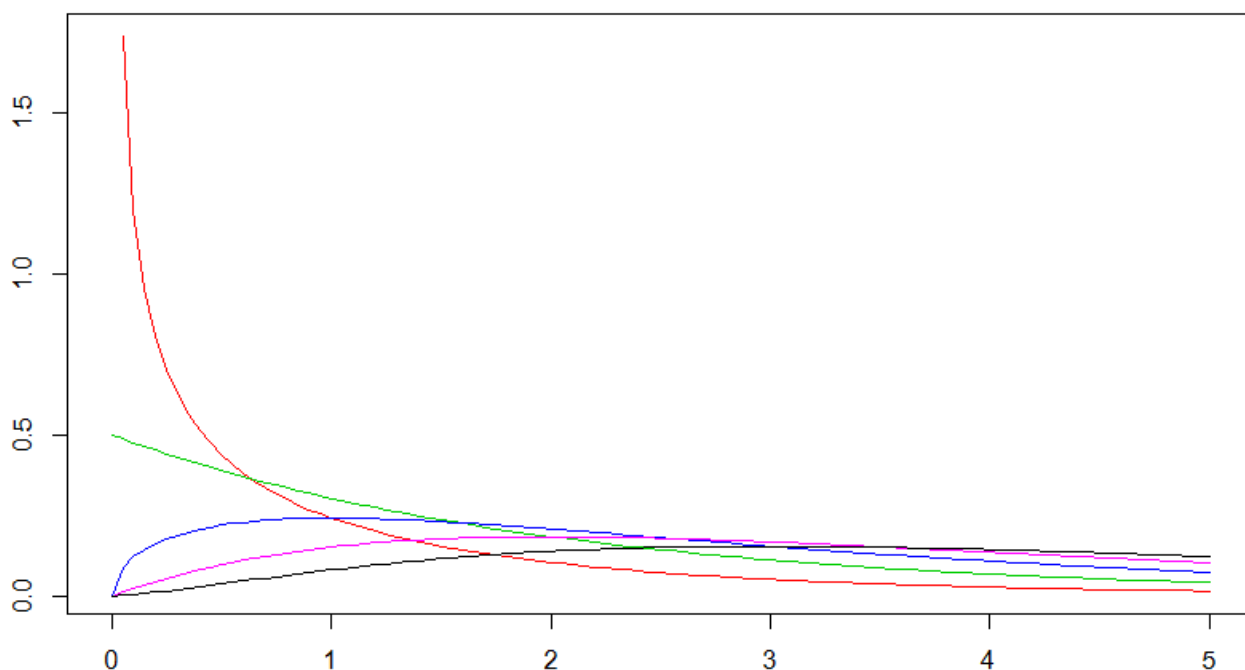


Рисунок 2.5. Графики функций плотности χ^2 -распределения для $k = 1$ (красный), для $k = 2$ (зеленый), для $k = 3$ (синий), для $k = 4$ (фиолетовый) и для $k = 5$ (черный).

Если $k = 2$, то распределение χ^2 совпадает с экспоненциальным распределением $Exp(\lambda)$ с параметром масштаба $\lambda = 0.5$.

Разработанный пакет на языке R содержит следующие функции для моделирования χ^2 -распределения:

- *chi.squared.rand.value(k=1)* – данная функция генерирует случайные величины, имеющие χ^2 -распределение, где параметр k (по умолчанию 1) – количество степеней свободы распределения.
- *chi.squared.rand.selection(count, k=1, sorted=false)* – функция моделирует повторные выборки объема *count*, извлекаемые из χ^2 -распределения с k степенями свободы.

5. *Распределение Фишера F*. Распределение Фишера является двухпараметрическим и имеет вид

$$F = \frac{Y_1/k_1}{Y_2/k_2}, \quad (2.14)$$

где Y_1 и Y_2 – две независимые случайные величины, имеющие распределения $\chi^2(k_1)$ и $\chi^2(k_2)$ соответственно, k_1 и k_2 – степени свободы этих распределений. Их также называют степенями свободы распределения Фишера: k_1 – числителя, k_2 – знаменателя. Данное распределение обладает следующими свойствами:

$$\mathbb{M}X = \frac{k_2}{k_2-2}, \text{ если } k_2 > 2, \quad (2.15)$$

$$\mathbb{D}X = \frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}, \text{ если } k_2 > 4, \quad (2.16)$$

где $\mathbb{M}X$ – математическое ожидание с.в., $\mathbb{D}X$ – дисперсия с. в., имеющей F -распределение.

Случайную величину распределения Фишера также можно выразить через стандартное нормальное распределение. Пусть x_1, \dots, x_{k_1} и y_1, \dots, y_{k_2} – случайные величины, распределенные по стандартному нормальному закону, тогда величина

$$F = \frac{k_2 \sum_{i=1}^{k_1} x_i^2}{k_1 \sum_{i=1}^{k_2} y_i^2} \quad (2.17)$$

имеет распределение Фишера.

Функция плотности распределения Фишера определяется следующим образом:

$$f(x) = \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \cdot \frac{x^{\frac{k_1}{2}-1}}{(x+1)^{\frac{k_1+k_2}{2}}}, \quad (2.18)$$

где $x > 0$, $\Gamma(x)$ – гамма-функция, k_1 и k_2 – числа степеней свободы F -распределения.

Разработанный пакет на языке R содержит следующие функции для моделирования распределения Фишера:

- *fisher.rand.value(k1=1, k2=1)* – данная функция генерирует случайные величины, имеющие *F*-распределение, где параметры k_1 и k_2 – степени свободы распределения.
- *fisher.rand.selection(count, k1=1, k2=1, sorted=false)* – функция моделирует повторные выборки объема *count*, извлекаемые из *F*-распределения с k_1 и k_2 степенями свободы.

Листинг 2.2. Реализация функций *fisher.rand.value* и *fisher.rand.selection*.

```
base.rand.selection ← function( count, value.func, sorted=F ) {
  X ← vector( "numeric", count )

  for( j in 1:count ){
    X[j] ← value.func()
  }

  if( sorted ){
    X ← sort( X )
  }
  return( X )
}

fisher.rand.value ← function( k1=1, k2=1 ) {
  k1.value ← chi.squared.rand.value( k1 ) / k1
  k2.value ← chi.squared.rand.value( k2 ) / k2
  return( k1.value / k2.value )
}

fisher.rand.selection ← function( count, k1=1, k2=1, sorted=F ) {
  value.func ← function() fisher.rand.value( k1, k2 )
  return( base.rand.selection( count, value.func, sorted ) )
}
```

Реализации функций *base.rand.selection*, *fisher.rand.value* и *fisher.rand.selection* приведены в листинге 2.2. Для функции *fisher.rand.selection* в качестве замыкания *value.func*, необходимого в базовой функции *base.rand.selection* для моделирования с. в., используется вызов функции *fisher.rand.value* с заданными параметрами *k1* и *k2*.

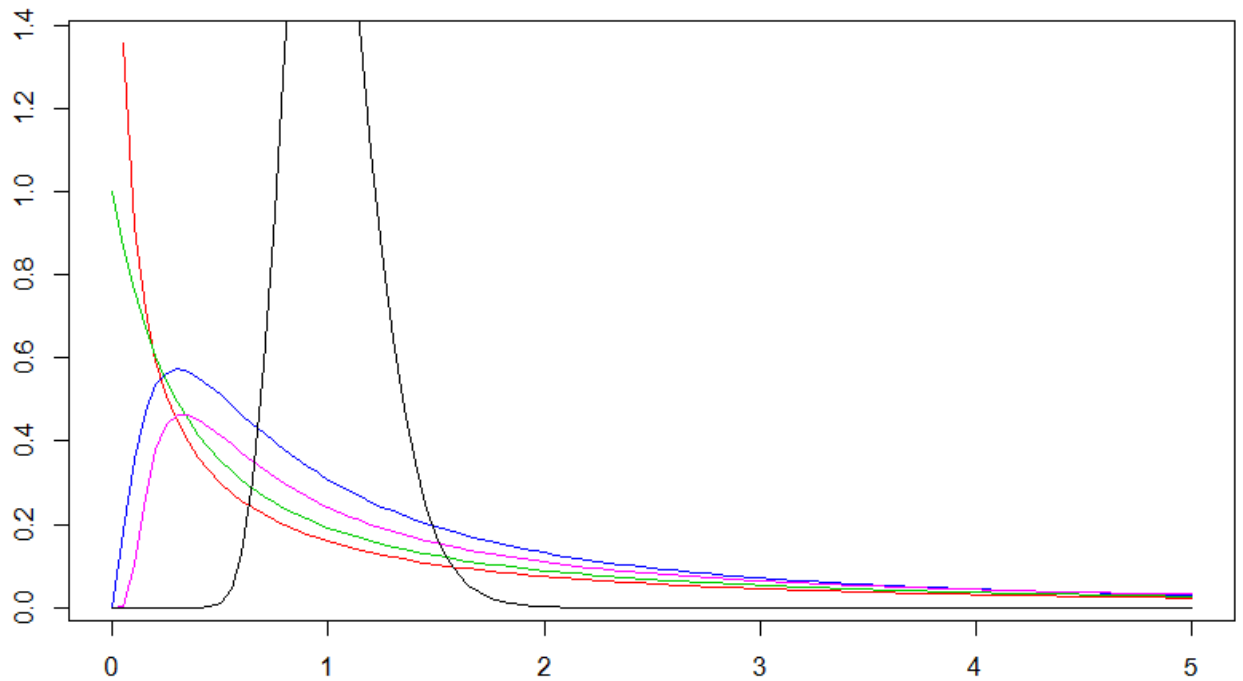


Рисунок 2.6. Графики функций плотности распределения Фишера для $k_1 = 1, k_2 = 1$ (красный), для $k_1 = 2, k_2 = 1$ (зеленый), для $k_1 = 5, k_2 = 2$ (синий), для $k_1 = 100, k_2 = 1$ (фиолетовый) и для $k_1 = 100, k_2 = 100$ (черный).

5. *Гамма-распределение.* Данное распределение задается плотностью вероятности, имеющей вид

$$f(x) = \begin{cases} x^{\alpha-1} \frac{e^{-\frac{x}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)}, & x \geq 0, \\ 0, & x < 0 \end{cases} \quad (2.19)$$

где $\Gamma(x)$ – гамма-функция, α и β – параметры гамма-распределения.

В литературе данное распределение часто обозначается, как $\Gamma(\alpha, \beta)$. Оно обладает следующими свойствами:

$$\mathbb{M}X = \alpha\beta, \mathbb{D}X = \alpha\beta^2, \quad (2.20)$$

где $\mathbb{M}X$ – математическое ожидание с.в., $\mathbb{D}X$ – дисперсия с. в., имеющей гамма-распределение.

Имеется множество алгоритмов моделирования гамма-распределения. В данной работе будет рассмотрен один из самых простых алгоритмов. Он

заключается в суммировании α с. в., имеющих экспоненциальное распределение с параметром масштаба $1/\beta$:

$$x = \sum_{i=1}^{\alpha} y_i, \quad (2.21)$$

где $y_i \in \text{Exp}\left(1/\beta\right)$, $i = 1, \dots, \alpha$. Недостатком данного метода является возможность моделирования гамма-распределения только для целых значений α . Такое распределение также называется распределением Эрланга.

Таблица 2.2. Параметры запросов функции */modulateselection* из таблицы 2.1 для разных распределений

n	Распределение	Значение параметра <i>sel_type</i>	Дополнительные параметры запроса
1	Экспоненциальное	«exp»	<ul style="list-style-type: none"> - <i>scale</i> – параметр масштаба λ (по умолчанию 1).
2	Вейбулла	«weibull»	<ul style="list-style-type: none"> - <i>scale</i> – параметр масштаба λ (по умолчанию 1), - <i>shape</i> – параметр формы k (по умолчанию 1).
3	Нормальное	«normal»	<ul style="list-style-type: none"> - <i>m</i> – математическое ожидание распределения μ (по умолчанию 0), - <i>d</i> – среднеквадратическое отклонение распределения σ (по умолчанию 1).
4	Пирсона χ^2	«chi.squared»	<ul style="list-style-type: none"> - <i>k</i> – параметр, задающий количество степеней свободы (по умолчанию 1).
5	Фишера	«fisher»	<ul style="list-style-type: none"> - <i>k1</i> – количество степеней свободы числителя (по умолчанию 1), - <i>k2</i> – количество степеней свободы знаменателя (по умолчанию 1).
6	Гамма	«gamma»	<ul style="list-style-type: none"> - <i>a</i> – параметр α (по умолчанию 1), - <i>b</i> – параметр β (по умолчанию 1).

Разработанный пакет на языке R содержит следующие функции для моделирования гамма-распределения:

- `gamma.rand.value(a=1, b=1)` – данная функция генерирует случайные величины, имеющие гамма-распределение с параметрами α и β ;
- `gamma.rand.selection(count, a=1, b=1, sorted=false)` – функция моделирует повторные выборки объема *count*, извлекаемые из гамма-распределения с параметрами α и β .

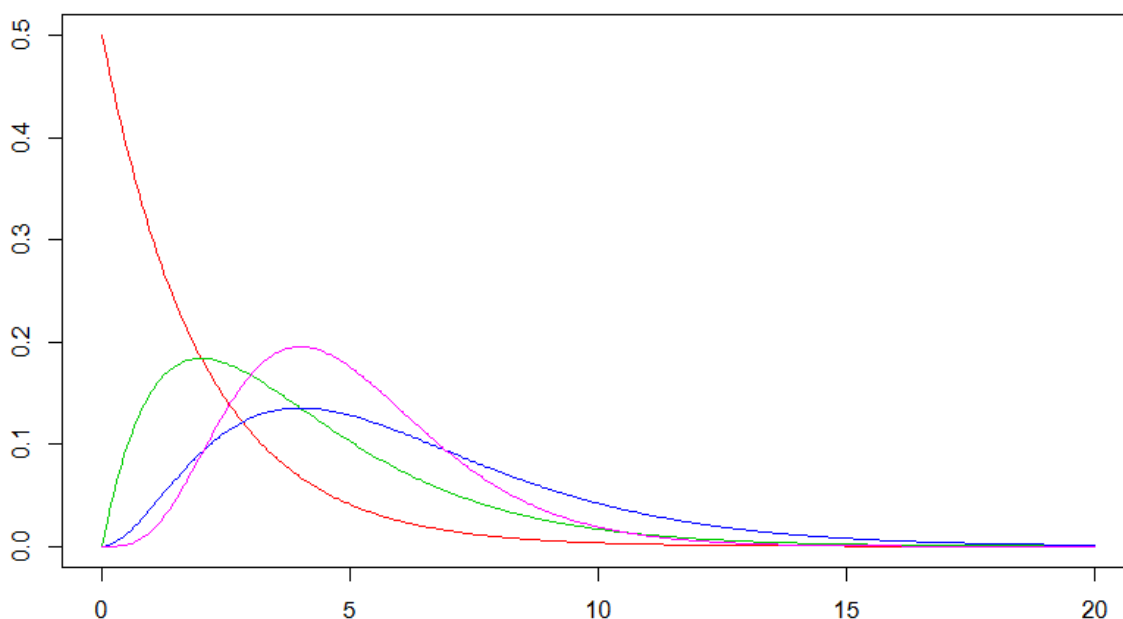


Рисунок 2.7. Графики функций плотности распределения Фишера для $\alpha = 1, \beta = 2$ (красный), для $\alpha = 2, \beta = 2$ (зеленый), для $\alpha = 3, \beta = 2$ (синий) и для $\alpha = 5, \beta = 1$ (фиолетовый).

2.3. Статистики

В разделе 1.3 приводится весь список используемых в программном продукте статистик для проверки гипотезы экспоненциальности. Данный список открыт для дополнения другими статистиками. Ввиду общего подхода по использованию не возникнет проблем с добавлением новой статистики в продукт, что обеспечивает его расширяемость.

Пакет на языке R содержит следующие базовые функции для выполнения проверки гипотезы H_0 :

- *base.test.value* – функция используется для создания функций, реализующих статистику, и является самой простой среди базовых статистик, выполняя лишь проверку переданной выборки, а также сортируя выборку для критериев, использующих порядковые статистики,
- *base.test.modulate* – функция обобщает алгоритмы моделирования квантилей для статистик, принимая в качестве параметров объем повторных выборок, уровень значимости, количество испытаний, а также реализацию статистики. Реализация данной функции представлена в листинге 2.2,
- *base.test.quantiles* – базовая функция, предназначенная обеспечивать доступ уже смоделированным квантилям, либо вызывать переданную в параметрах функцию моделирования,
- *base.test* – функция для обобщения алгоритмов проверки гипотезы H_0 по выборочным данным, принимая в качестве параметров повторную выборку, уровень значимости, реализацию алгоритма статистики, а также функцию доступа к критическим значениям статистики.
- *base.test.power* – функция предназначена для обеспечения общего подхода к моделированию мощностей статистик против различных альтернатив и принимает в качестве параметров объем генерируемых выборок, извлекаемых из альтернативного распределения, функцию для моделирования выборок, функцию для проверки гипотезы H_0 с помощью выбранной статистики, а также функцию доступа к квантилям этой статистики,
- *base.test.power.values* – базовая функция для доступа к уже имеющимся значениям мощностей статистики против определенной альтернативы. Функция принимает замыкания для доступа к мощностям по каждому из стандартных уровней значимости (0.01, 0.05 и 0.1) и функцию для моделирования мощности, которая вызывается, если требуемое значение мощности отсутствует.

Для каждой из статистики имеются следующие функции:

- [название статистики].test.value;

- [название статистики].test.modulate;
- [название статистики].test.quantiles;
- [название статистики].test;
- [название статистики].test.[название альтернативы].power;
- [название статистики].test.[название альтернативы].power.values.

Главной функцией пакета является *test.for.exponentiality*, которая выполняет проверку гипотезы H_0 по выборочным данным с помощью всех имеющихся в пакете статистик. Если альтернатива не задана, то функция выполняет проверку на согласованность выборочных данных гипотезе. Если альтернатива задана, то функция выполняет проверку гипотезы H_0 с помощью ряда наиболее мощных статистик среди имеющихся в программном продукте против альтернативы H_1 . Для задания альтернативы необходимо передать в качестве аргумента функции название распределения. Для проверки гипотезы H_0 с помощью ряда определенных критериев необходимо передать в качестве аргумента массив с названиями требуемых статистик.

В API-интерфейсе веб-части продукта имеются следующие функции:

- */testtypes* – функция возвращает названия всех имеющихся в приложении статистик,
- */signlevels* – функция возвращает набор всех уровней значимости, используемых для проверки гипотезы H_0 ,
- */addsignlevel* – функция для добавления нового уровня значимости
- */removesignlevel* – функция удаления уровня значимости из списка всех доступных пользователю уровней значимости,
- */modulatequantile* – функция для запуска моделирования квантилей статистики с заданными в параметрах уровнем значимости, объемом выборки и количеством испытаний,
- */getquantiles* – функция для загрузки квантилей для выбранной статистики с заданными параметрами,

- */modulatepower* – функция для запуска моделирования мощности статистики против заданной в списке параметров альтернативы.
- */testselection* – функция для проверки гипотезы экспоненциальности по выборочным данным.

Листинг 2.3. Реализация функции `base.test.modulate`

```
base.test.modulate ← function( count, n, alpha, test.func ){
  if ( count <= 0 ){
    return( "Wrong iteration count!" )
  }
  if( n <= 0 ){
    return( "Wrong selection size!" )
  }
  if( is.null( alpha ) ){
    return( "Wrong significance level!" )
  }
  for ( i in 1:length( alpha ) ){
    if ( alpha[i] <= 0 | alpha[i] >=0.5 ){
      return( "Wrong significance level!" )
    }
  }

  Vals ← vector( "numeric", count )

  for( i in 1:count ){
    X ← exp.rand.selection( n, 1, F )
    Vals[i] ← test.func( X )
  }

  Vals ← sort( Vals )

  alpha.size ← length( alpha )
  result ← vector( "numeric", 2 * alpha.size )

  for ( i in 1:alpha.size ){
    result[2*i - 1] ← Vals[alpha[i] * count]
    result[2*i] ← Vals[(1 - alpha[i]) * count]
  }

  return( result )
}
```

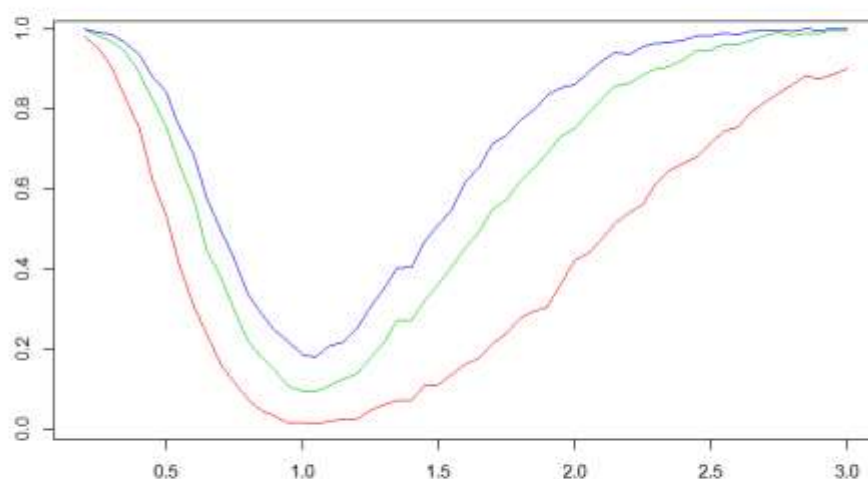


Рисунок 2.8. Графики мощностей статистики Джини против We -альтернативы для различных уровней значимости. Ось ординат графиков отображает значение мощности статистики, ось абсцисс – значение параметра формы k распределения.

На рис. 2.8 изображены графики мощностей статистики Джини G_n против альтернативы Вейбулла для различных уровней значимости λ . Красным цветом изображен график мощностей для $\lambda = 0.01$, зеленым – для $\lambda = 0.05$, синим – для $\lambda = 0.1$. Анализ рисунка 2.8 показывает, что при увеличении уровня значимости статистики её мощность увеличивается. При этом не стоит забывать о том, что уровень значимости равен объему ошибок первого рода.

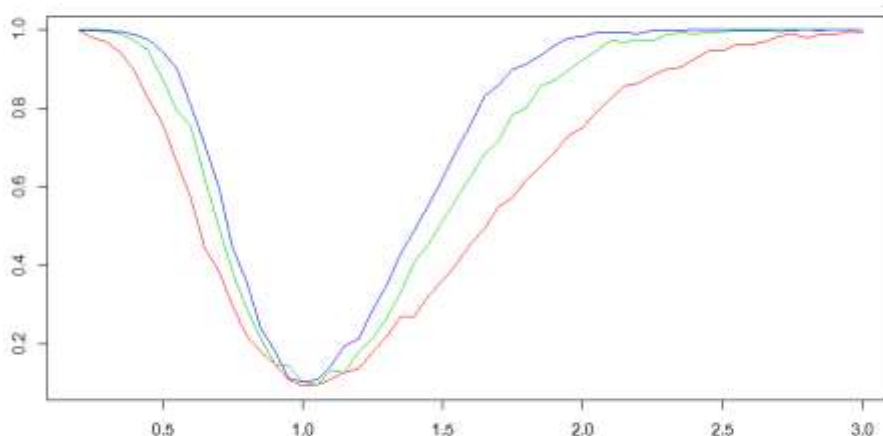


Рисунок 2.9. Графики мощностей статистики Джини против We -альтернативы для различных объемов выборки n . Ось ординат графиков отображает значение мощности статистики, ось абсцисс – значение параметра формы k распределения.

На рис. 2.9 изображены графики мощностей статистики G_n против альтернативы Вейбулла для разных объемов выборки n . Красным цветом изображен график мощностей для $n = 10$, зеленым – для $n = 15$, синим – для $n = 20$. Увеличение мощности статистики G_n при росте n свидетельствует о его состоятельности.

Frocini test		Greenwood test		Moran test		Shapiro-Wilk test		Sherman test	
Result	True	Result	True	Result	True	Result	True	Result	False
Value	0.28862	Value	2.5886	Value	-0.29338	Value	0.13247	Value	0.48218
First	0.11858	First	1.3531	First	-0.36252	First	0.029429	First	0.24129
Second	0.37538	Second	2.6363	Second	0.34793	Second	0.13636	Second	0.46763

Рисунок 2.10 Результаты проверки гипотезы в веб-браузере.

Пример 2.2. На рис. 2.10 изображен результат проверки на состоятельность гипотезы H_0 по выборочным данным с помощью статистик Фроцини, Гринвуда, Морана, Шапиро-Уилка и Шермана. В качестве возвращаемой клиенту информации по каждой из статистик выступают:

- значение статистики для заданной выборки,
- квантили статистики при заданном уровне значимости и объеме выборки,
- результат проверки гипотезы (попадает ли значение статистики для выборки в интервал между квантилями).

По полученным результатам можно склониться к принятию гипотезы H_0 , так как четыре статистики из пяти её принимают. ▀

Ввиду различной алгоритмической сложности следует сравнить время моделирования квантилей для различных критериев. В табл. 2.3 приводится данное сравнение для объема выборочных данных $n = 100$. В этой таблице также приведено сравнение двух алгоритмов для статистики Джини из (1.19) и (1.20).

Как показывают результаты, вариант вычисления значения статистики под номером (1.19) значительно уступает по времени (примерно в 12 раз) более удобному для моделирования варианту из (1.20). Также можно обратить внимание на тот факт, что сложность алгоритма не всегда говорит о его качестве. Например, критерий Кокса-Оукса часто является наиболее мощным

критерием против различных альтернатив, а также имеет максимальные значения АОЭ по Бахадуру, но при этом время моделирования квантилей для этой статистики меньше многих других из табл. 2.3.

В основном все статистики имеют схожую алгоритмическую сложность. Но имеются такие статистики, для которых время моделирования квантилей сравнительно больше (например, критерии Фортиана-Гране и Кимбера-Мичела). При этом данные статистики не зарекомендовали себя, как «более лучшие» для проверки гипотезы H_0 . Для статистики Джексона выполнение моделирования занимает в среднем в шесть раз больше времени, чем для остальных критериев.

Сложность алгоритма может сыграть ключевую роль при выборе некоторого критерия из имеющихся, так как могут возникнуть ситуации, когда нет достаточного времени для моделирования квантилей и мощностей статистики.

Таблица 2.3. Значения времени работы функций моделирования квантилей для различных статистик. $n = 100$.

Статистика	Название функции	Номер формулы	Время работы (в сек.)
Бартлетта	bartlett.test.modulate	1.26	13.26
Кокса-Оукса	coxoakes.test.modulate	1.18	14.88
Фортиана-Гране	fortianagrane.test.modulate	1.22	21.04
Фроцини	frozini.test.modulate	1.23	16.93
Джини	gini.test.value	1.20	16.67
	gini.test.value.bad	1.19	205.84
Гринвуда	greenwood.test.value	1.17	13.66
Джексона	jackson.test.value	1.21	100.89
Кимбера-Мичела	kimbermichel.test.value	1.30	22.89
Морана	moran.test.value	1.16	13.61
Рао	rao.test.value	1.31	18.99
Шапиро-Уилка	shapirowilk.test.value	1.25	13.43
Шермана	sherman.test.value	1.24	13.48

На время моделирования мощностей влияет не только сложность статистики, но и сложность алгоритма моделирования альтернативы. В табл. 2.4 приводится сравнение моделирования мощностей статистики Джексона против различных альтернатив для объема выборок $n = 10$. Анализ табл. 2.4 показывает, что алгоритм генерации с. в. может существенно сказываться на времени работы моделирования мощности. Для распределений, моделирование которых основано на суммировании с. в., подчиняющихся другим распределениям, время работы алгоритма растет с каждым увеличением параметра, по которому происходит суммирование.

Таблица 2.4. Значения времени работы функций моделирования мощностей статистики Джини против различных альтернатив. $n = 10$.

Альтернатива	Время работы (в сек.)
Вейбулла	0.41
Нормальное	0.59
Пирсона χ^2 для $k = 1$	0.66
Пирсона χ^2 для $k = 5$	1.70
Пирсона χ^2 для $k = 10$	3.30
Фишера для $k_1 = 1, k_2 = 1$	0.96
Фишера для $k_1 = 5, k_2 = 1$	2.00
Фишера для $k_1 = 5, k_2 = 5$	3.20
Фишера для $k_1 = 10, k_2 = 1$	3.35
Фишера для $k_1 = 10, k_2 = 10$	5.85
Гамма для $\alpha = 1$	0.45
Гамма для $\alpha = 5$	0.85
Гамма для $\alpha = 10$	1.36

Например, при увеличении параметра $\alpha = 1$ до значения $\alpha = 5$ при моделировании гамма-распределения, время выполнения алгоритма увеличилось примерно в два раза. Причинами, по которым время выполнения моделирования не увеличилось в пять раз, является тот факт, что затраты на упорядочивание повторных выборок (статистика Джексона использует порядковые статистики) и на вычисление значений статистики.

2.4. Примеры использования программного продукта

В примерах 1.1, 1.3 была рассмотрена задача из [5, с. 257] о проверке гипотезы экспоненциальности для моментов появления отказов автомобиля во время стендовых виброиспытаний. В данной задаче для выборочных данных гипотеза экспоненциальности принимается. Рассмотрим еще несколько примеров использования программного продукта на примерах из различных источников.

В [12] рассматривается задача расчета надежности аппаратуры гидроакустического канала связи. В ней при оценке надежности различных микросхем, резисторов, конденсаторов и других приборов в качестве распределения вероятностей их отказов рассматривается экспоненциальное распределение, так как только для этого закона имеются необходимые справочные данные об интенсивности отказов электронных схем. При этом возникает необходимость в проверке экспоненциальности интервалов отказов. В качестве инструмента для проверки гипотезы об экспоненциальности в данной задаче можно использовать разработанный в данной задаче продукт.

В [1] рассматривается задача о проверке постоянности функции интенсивности отказов системы кондиционирования воздуха в самолетах Боинг-720. Данная проверка выполняется для определения того, имеется ли тенденция изменения интервалов между отказами. Увеличение интервалов может свидетельствовать о росте опыта обслуживания персонала, устранения дефектных узлов. Уменьшение интервалов может быть следствием износа, старения, или плохого технического обслуживания.

Как известно, функция интенсивности отказов постоянна тогда и только тогда, когда распределение интервалов между отказами имеет экспоненциальное распределение.

В табл. 2.5 приводятся результаты наблюдений. Капитальный ремонт был проведен на самолетах через 2000 часов после начала испытаний. В [1]

проверка осуществляется для всего парка самолетов, а также для каждого самолета в отдельности.

Для случая, когда результаты испытаний для самолетов рассматриваются, как одно целое, в [1] при помощи изучения гистограммы эмпирической функции распределения гипотеза отвергается в пользу распределения убывающей функцией интенсивности (УФИ).

Таблица 2.5. Интервалы между отказами системы кондиционирования самолетов Боинг-720. Капитальный ремонт отмечен в таблице символом +.

Номер самолета												
7907	7908	7909	7910	7911	7912	7913	7914	7915	7916	7917	8044	8045
194	413	90	74	55	23	97	50	359	50	130	487	102
13	14	10	57	320	261	51	44	9	254	493	18	209
41	58	60	48	56	87	11	102	12	5		100	14
29	37	186	29	104	7	4	72	270	283		7	57
33	100	61	502	220	120	141	22	603	35		98	54
181	65	49	12	239	14	18	39	3	12		5	32
	9	14	70	47	62	142	3	104			85	67
	169	24	21	246	47	68	15	2			91	59
	447	56	29	176	225	77	197	438			43	134
	184	20	386	182	71	80	188				230	152
	36	79	59	33	246	1	79				3	27
	201	84	27	+	21	16	88				130	14
	118	44	+	15	42	106	46					230
	+	59	153	104	20	206	5					66
	34	29	26	35	5	82	5					61
	31	118	326		12	54	36					34
	18	25			120	31	22					
	18	156			11	216	139					
	67	310			3	46	210					
	57	76			14	111	97					
	62	26			71	39	30					
	7	44			11	63	23					
	22	23			14	18	13					
	34	62			11	191	14					
		+			16	18						
		130			90	163						
		208			1	24						
		70			16							
		101			52							
		208			95							

Результаты проверок различными статистиками всего множества интервалов отказов приведено в табл. 2.6. Анализ данной таблицы показывает, что для большинства критериев (критерии Гринвуда, Фроцини, Шермана, Шапиро-Уилка, Джини, Кокса-Оукса и Фортиана-Гране) гипотеза экспоненциальности отклоняется, и лишь для критериев Морана, Кимбера-Мичела и Рао гипотеза принимается. Поэтому можно предположить, что гипотеза H_0 не согласуется с выборочными данными из задачи из [1], и отклонить её.

Таблица 2.6. Результаты проверки гипотезы экспоненциальности по всем интервалам отказов с помощью ряда критериев, представленных в данной работе. $\alpha = 0.05$, $n = 203$

Статистика	Значение статистики	Квантили статистики		Результат
Морана	-0.0565	-0.0938	0.0888	принимается
Гринвуда	2.309	1.7941	2.361	отклоняется
Фроцини	0.4914	0.3821 (односторонняя)		отклоняется
Шермана	0.4102	0.3398	0.3947	отклоняется
Шапиро-Уилка	0.0061	0.0037	0.0058	отклоняется
Кимбера-Мичела	0.0448	0.0615 (односторонняя)		принимается
Джини	0.5492	0.4670	0.5330	отклоняется
Рао	2.5149	3.9315(односторонняя)		принимается
Кокса-Оукса	-0.1461	-0.1440	0.1464	отклоняется
Фортиана-Гране	1.1317	0.8821	1.0951	отклоняется

Отклонение гипотезы об экспоненциальности для сгруппированных интервалов отказов всех самолетов не означает, что распределение интервалов отказов для каждого самолета в отдельности не может быть экспоненциальным.

Если бы все исследуемые самолеты имели одинаковую интенсивность отказов, то интервалы, сгруппированные для различных самолетов, имели бы экспоненциальное распределение. С другой стороны, если каждому из самолетов в отдельности соответствует различная, хотя и постоянная интенсивность отказов, то сгруппированные интервалы должны иметь убывающую интенсивность отказов.

Далее можно проверить гипотезу H_0 для каждого самолета в отдельности. В [1] для поиска тенденции изменения интенсивности отказов используется критерий Манна.

Таблица 2.7. Результаты проверки гипотезы экспоненциальности для каждого самолета с помощью ряда критериев, представленных в данной работе. $\alpha = 0.05$. Принятие гипотезы с помощью статистики обозначается символом «+», отклонение – символом «-».

	Статистики								
	Gr_n	Fr_n	Sh_n	SW_n	KM_n	G_n	R_n	CO_n	FG_n
7907	+	+	+	+	+	+	+	+	+
7908	+	+	+	+	+	+	+	+	+
7908*	+	+	+	-	+	+	+	-	+
7909	+	+	+	+	+	+	+	+	+
7909*	-	+	+	+	+	+	+	-	-
7910	-	-	-	-	-	-	+	+	-
7910*	+	+	+	+	+	+	+	+	+
7911	+	+	+	+	+	+	+	+	+
7911*	+	-	+	+	+	+	+	+	+
7912	+	+	-	+	+	-	+	+	+
7913	+	+	+	+	+	+	+	+	+
7914	+	+	+	+	+	+	+	+	+
7915	+	-	-	+	-	+	-	-	+
7916	+	+	+	+	+	+	+	+	+
7917	+	+	+	+	+	+	+	+	+
8044	+	+	+	+	+	+	+	+	+
8045	+	+	+	+	+	+	+	+	+

Для проверки гипотезы H_0 для интервалов отказов отдельных самолетов используем те же критерии, что и в табл. 2.6. Будем рассматривать данные об отказах для самолетов до и после капитального ремонта отдельно, так как их интенсивность отказов могла измениться. Результаты проверки представлены в табл. 2.7. Анализ данной таблицы показывает, что для большинства самолетов (7907, 7908, 7909, 7910 после ремонта, 7911, 7913, 7914, 7916, 7917, 8044, 8045) гипотеза H_0 принимается всеми использованными статистиками. Также для самолетов с номерами 7908 (после ремонта), 7909 (после ремонта), 7911 (после ремонта) и 7912 гипотеза принимается большинством

критериев. Для данных самолетов имеет смысл принять гипотезу экспоненциальности.

Для самолета с номером 7910 до ремонта большинство статистик при проверке гипотезы H_0 дают отрицательный результат. Следовательно, для этого самолета следует отклонить гипотезу.

Для самолета с номером 1715 четыре статистики принимают гипотезу H_0 и пять статистик отклоняют ее. В такой ситуации крайне трудно принять окончательное решение. Если имеется возможность повторить испытание, следует ею воспользоваться. Также можно сузить класс альтернатив и принимать решение, исходя из различных характеристик используемых статистик.

Глава 3. СПЕЦИАЛЬНЫЕ ВОПРОСЫ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ

Веб-интерфейс является инструментом для взаимодействия пользователя с веб-приложением через браузер. По этой причине обеспечение качественного веб-интерфейса играет важную роль при разработке программного продукта.

Веб-интерфейс приложения включает в себя набор страниц HTML, а также каскадные таблицы CSS, отвечающие за стили отображаемых данных, и наборы скриптов на языке Javascript, отвечающие за поведение страниц. Страницы HTML могут содержать в себе различные формы, списки, таблицы, кнопки, поля для ввода текстовых данных и т. д.

Основные исследования в данной области направлены на улучшение доступности, удобства и простоты интерфейсов, а также увеличение безопасности данных. Например, в качестве доступности можно рассматривать хорошую разборчивость текста интерфейса для людей со слабым зрением, а также наличие нескольких доступных языков для локализации в приложении. При проектировании интерфейса часто важную роль играет простота, так как излишне усложненный интерфейс может привести к возникновению различных проблем у пользователя при использовании приложения. Также одним из важных качеств хорошего интерфейса является предсказуемость поведения приложения в ответ на действия пользователя.

Многие разработчики рассматривают веб-интерфейс, как лишь некоторое дополнение к функциональной части приложения, часто не выделяя должных ресурсов на его развитие. При этом для пользователя интерфейс является главной частью веб-сервиса, поэтому некачественный интерфейс резко снижает функциональность и качество системы в целом. Общее впечатление от использования приложения складывается в основном из достоинств и недостатков интерфейса. Отсутствие желания у разработчика программного

обеспечения к разработке качественного и понятного интерфейса может привести к весьма негативным последствиям.

На впечатление пользователя очень сильно влияет время отклика сервера на отправленный запрос. Долгое время загрузки может значительно сократить количество желающих пользоваться сервисом.

Компания *Aberdeen Group* провела исследование, согласно которому задержка загрузки страницы на одну секунду ведет за собой следующие последствия:

- уменьшение количества просмотров страницы на 11%;
- снижение удовлетворенности покупателей на 16%;
- снижение конверсии посетителей в покупателей на 7%.

По этой причине добавлять новые возможности, которые могут существенно замедлить работу с приложением, нужно лишь при крайней необходимости.

Также важным свойством интерфейса, на которое пользователи часто обращают внимание, является кроссбраузерность, т. е. поддержка одинакового внешнего вида и одинаковой функциональности при работе в различных браузерах. Помимо этого на оценку качества интерфейса может повлиять возможность работы с ним через мобильные устройства.

В рамках данной работы используется множество технологий, позволяющих улучшить качество веб-интерфейса. Например, *Bootstrap Framework* содержит в себе набор различных стандартных классов стилей для более удобного отображения элементов страницы. Также Bootstrap имеет инструменты для обеспечения кроссбраузерности приложения.

Веб-интерфейс разработанного программного продукта отличается своей простотой и понятностью. Если пользователь не авторизован, при входе на сайт приложение перенаправит пользователя на страницу авторизации. Также для неавторизованного пользователя имеется доступ к странице регистрации. Страницы авторизации и регистрации приводятся на рис. 3.1.

Вход

Регистрация

Рисунок 3.1. Формы регистрации и авторизации

После авторизации пользователю становятся доступными страницы для работы со статистиками и выборочными данными. В верхней части всех страниц приложения имеется навигационное меню. Данное меню изображено на рис. 3.2. При нажатии на имя пользователя с правой части меню появляется список с некоторыми доступными для пользователя действиями:

- выйти из учетной записи. При нажатии на данную кнопку сервис перенаправляет пользователя на страницу авторизации.
- загрузить в базу данных приложения уже имеющуюся у пользователя выборку. Значения выборки можно загрузить из файла, или ввести в текстовое поле. Также необходимо задать идентификатор выборки. Модальное окно для загрузки выборочных данных приводится на рис. 3.3.
- смоделировать повторную выборку, извлекающуюся из некоторого распределения. Данное распределение задается в модальном окне для моделирования. Помимо этого необходимо задать идентификатор и желаемый объем для моделируемой выборки. Данное окно представлено на рис. 3.4.

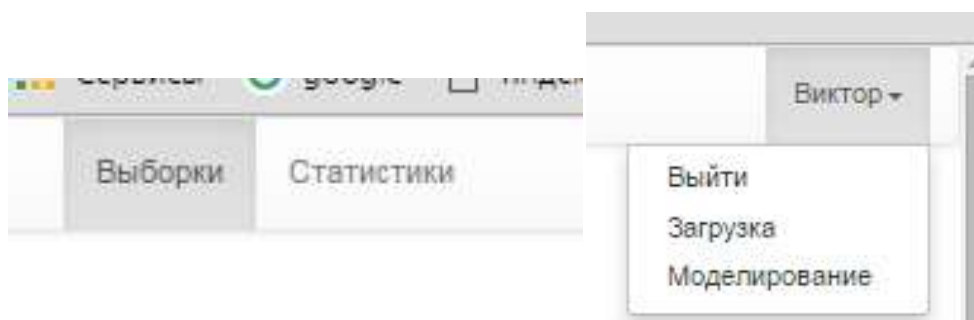


Рисунок 3.2. Левая и правая части навигационного меню страниц приложения

The screenshot shows a window titled "Добавить существующие данные" (Add existing data). It contains two input fields: "Идентификатор:" (Identifier) and "Значения:" (Values). Below the "Значения:" field is a button labeled "Загрузить из файла" (Load from file). At the bottom right of the window is a blue button labeled "Добавить" (Add).

Рисунок 3.3. Окно добавления существующей выборки в базу данных приложения

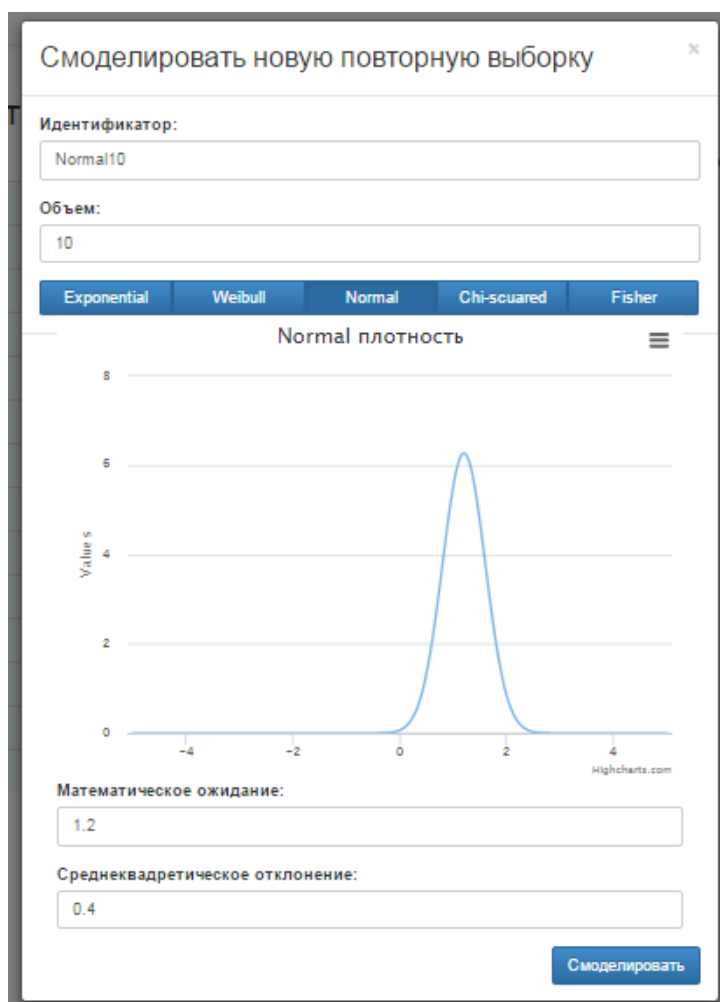


Рисунок 3.4. Окно моделирования повторной выборки в базу данных приложения

При переходе по вкладке «Статистики» навигационного меню, открывается страница со всеми доступными в продукте статистиками для проверки

гипотезы экспоненциальности. При выборе одной из статистик появляется формула статистики, а также таблица её квантилей.

При отсутствии квантилей для некоторого объема выборки и уровня значимости имеется возможность запустить моделирование этих квантилей нажатием на нужную ячейку таблицы. Текущий прогресс моделирования квантилей отображается в данной ячейке.

12	1.2432	3.3037
13	1.2636	3.3017
14	<div> <div style="width: 33%;">33%</div> </div>	
15	1.2892	3.2683
16	1.3065	3.2045

Рисунок 3.5. Отображение прогресса моделирования квантилей

Также имеется возможность изучить мощности статистики против различных альтернатив.

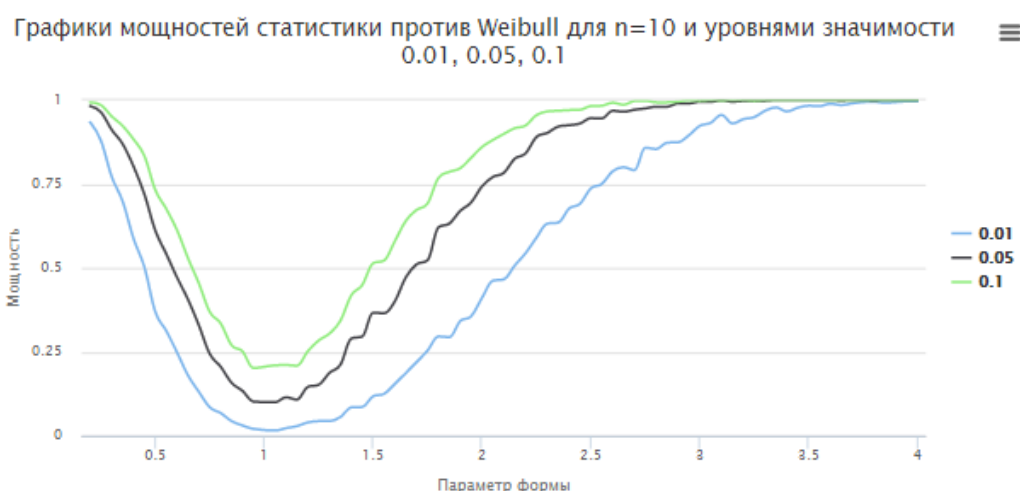


Рисунок 3.6. Графики мощностей статистики против заданной альтернативы

При переходе по вкладке «Выборки» навигационного меню, открывается страница с таблицей всех имеющихся у пользователя выборок. Имеется возможность посмотреть значения каждой из них, а также выполнить проверку дынных выборок на экспоненциальность. Проверку можно выполнять на согласованность экспоненциальному распределению, а также против некоторого семейства альтернатив. При выборе альтернативного распределения отображаются графики мощностей использованных при проверке статистик

для их сравнения, а также таблица со значениями асимптотически оптимальной эффективности статистик для выбранной альтернативы. Данная отображаемая информация изображена па рис. 3.7. При анализе этой информации можно принять или отклонить гипотезу H_0 .

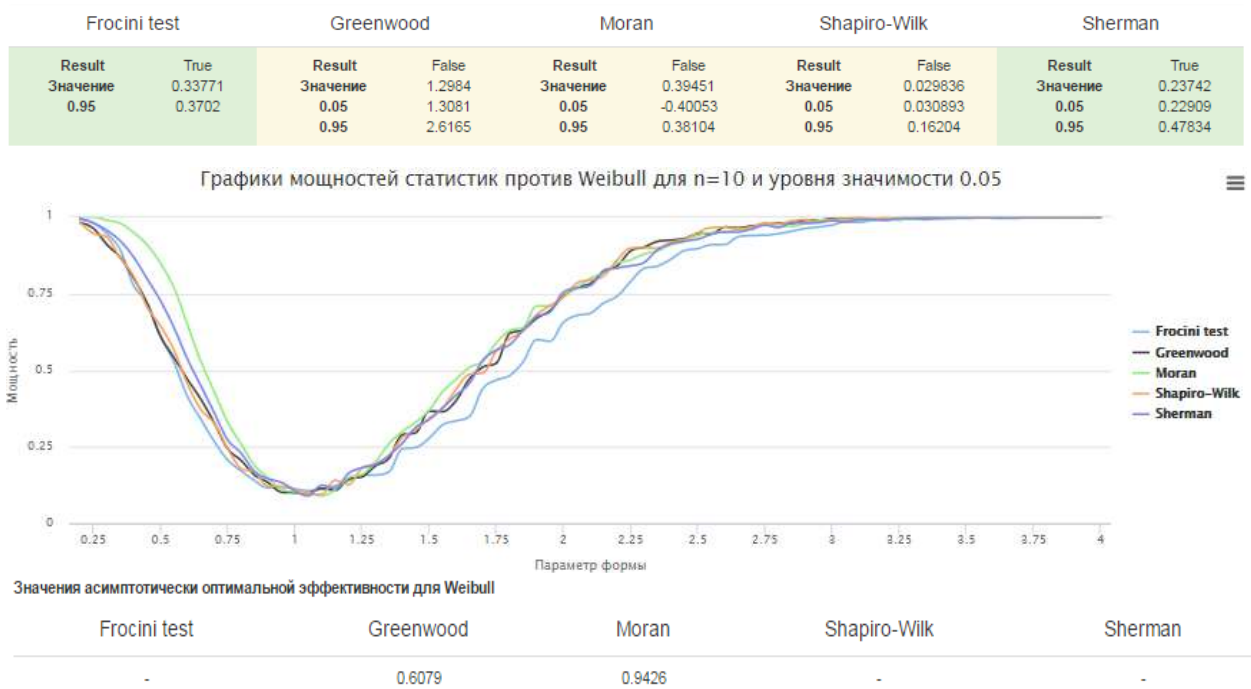


Рисунок 3.7. Информация, отображаемая при проверке выборочных данных на экспоненциальность.

Веб-интерфейс разработанного приложения довольно прост в применении. Пи этом для его использования пользователь должен обладать минимальным необходимым объемом знаний из области математической статистики.

ЗАКЛЮЧЕНИЕ

Наличие в разработанном продукте большого количества различных критериев, а также инструментов для сравнения их характеристик позволяет снизить вероятность принятия неверного решения в задачах проверки гипотезы экспоненциальности. Инструменты моделирования квантилей статистик имеют точность несколько знаков после запятой. Такой точности оказывается достаточно для высказывания различных предположений. Наличие в имитационном блоке инструментов генерирования с. в. большого количества распределений позволяет моделировать мощность критериев против различных альтернатив. Также значения асимптотически оптимальной эффективности по Бахадуру для некоторых статистик, имеющиеся в программном продукте, могут служить аргументами при выборе более предпочтительного критерия для проверки.

При рассмотрении ряда критериев из главы 1.3 были изучены следующие принципы для проверки гипотезы экспоненциальности:

- при проверке на согласованность данных не стоит принимать окончательный ответ на основе лишь формальной проверки на экспоненциальность по какому-либо одному из выбранных критериев;
- при заданной альтернативе следует делать выбор в пользу одного из критериев, основываясь на имеющихся данных о значениях АОЭ по Бахадуру этих критериев;
- при отсутствии значений некоторых квантилей или мощностей для статистик имеется возможность их моделирования с помощью метода Монте-Карло;
- если есть возможность увеличения объема выборки n , то при возникновении неопределенности при принятии решения эту возможность следует использовать.

Проверка программного продукта на различных задачах показала, что при его использовании в большинстве случаев бывает достаточно аргументов для принятия конечного решения.

Разработанный программный продукт открыт для расширения дополнительными критериями и альтернативными распределениями. В перспективе можно расширять функциональную базу продукта. Например, можно добавить алгоритмы моделирования эмпирической эффективности статистик.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Барлоу Р., Прошан Ф. Математическая теория надежности. М.: Советское радио, 1969. – 488 с.
2. Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д. Математические методы в теории надежности. – М.: Наука, 1965. – 524 с.
3. Григорьев Ю. Д. Гипотеза экспоненциальности: методологический аспект. Заводская лаборатория. Диагностика материалов. 2016. Т.82. № 1. С. 69-80.
4. Ивченко Г. И., Медведев Ю. И. Математическая статистка: Учебное пособие для вузов – М.: Высш. шк. 1984. – 248 с.
5. Капур Л., Ламберсон Л. Надежность и проектирование систем. – М.: Мир, 1980. – 606 с.
6. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
7. Кокс Д. Р., Оукс Д. Анализ данных типа времени жизни. – М.: Финансы и статистика, 1988. – 192 с.
8. Никитин Я. Ю. Асимптотическая эффективность непараметрических критериев. – М.: Наука, 1995. – 240 с.
9. Рао С. Р. Линейные статистические методы и их применения. – М.: Наука, 1968. – 548 с.
10. Чирина А. В. Асимптотическая эффективность критериев экспоненциальности, свободных от параметра масштаба // Дисс... канд. физ.-мат. наук, СПбГУ, 2005.
11. Чистяков В. П. Курс теории вероятностей. — М.: Наука, 1982. – 233 с.
12. Филиппов Б. И., Малахова Е. А. Расчет надежности аппаратуры гидроакустического канала связи. – Сборник научных трудов НГТУ, 2015, No.4(82), 67-91.

13. Baringhaus, L., Henze, N. Tests of fit for exponentiality based on a characterization via the mean residual life function// Statistical Papers, 2000, 41, 225–236.
14. Box G., Muller M. A Note on the Generation of Random Normal Deviates, The Annals of Mathematical Statistics (1958), Vol. 29, No. 2 pp. 610–611
15. Epstein B. Testing for the validity of the assumption that the underlying distribution of life is exponential. Technometrics, 1960, 2, 1-2, 83-101, 167-183.
16. Epps T. W., Pulley L. B. A test for exponentiality vs. monotone hazard alternatives derived from the empirical characteristic functions. JRSS, 1986, B48, 206-213.
17. Fortiana G., Grane A. A scale-free goodness-of-fit statistic for the exponential distribution based on maximum correlations. Journ. Statist. Plann. Infer., 2002, 108, 85-97.
18. Frozini B. V. On the distribution and power of a goodness-of-fit statistic with parametric and nonparametric applications. Amsterdam-Oxford-New York: North-Holland. Publ. Comp., 1987, 133-154
19. Gafter N. A definition of closures. 2007, Neal Gafter's Blog, URL: <http://gafter.blogspot.ru/2007/01/definition-of-closures.html>
20. Gail M. H., Gastwirth J. L. A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. JRSS, 1978, B40, No. 3, 350-357.
21. Greenwood M. The statistical study of infectious diseases. JRSS, 1946, Ser. A, 109, 85-110.
22. Jackson O. A. Y. An analysis of departures from the exponential distribution. JRSS, 1967, B29, 540-549.
23. Kimber, A.C. Tests for the exponential, Weibull and Gumbel distributions based on the stabilized probability plot. Biometrika, 1985, 72, 3, 661-663.

24. Leff, A., Rayfield, J. T. Web-Application Development Using the Model/View/Controller Design Pattern. IEEE Enterprise Distributed Object Computing Conference, 2001, 118–127.
25. Lilliefors, H. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown// J. Amer. Stat. Ass., 1969, 64, 387-389.
26. Mann H. B. Nonparametric tests against trend. *Econometrica*, 1945, v. 13, 245-259.
27. Moran P. The random division of an interval. Part II. *JRRS*, 1951, B13, No. 1, 147-150.
28. Sherman B. A random variable related to the spacing of sample values. *Ann. Math. Stat.*, 1950, 421, 339-361.
29. Shapiro S., Wilk M. An analysis of variance test for the exponential distribution complete samples. *Technometrics*, 1972, 14, 355-370.
30. Stephens M. A. On the W test for exponentiality with origin known. *Technometrics*, 1978, 20, 353-357.

ПРИЛОЖЕНИЕ А. ТАБЛИЦЫ КВАНТИЛЕЙ СТАТИСТИК

Таблица 1. Квантили для статистики Кокса-Оукса.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	-0.3943	0.5148	-0.5714	0.6010	-0.9281	0.7324
11	-0.3869	0.4998	-0.5512	0.5828	-0.8763	0.7104
12	-0.3744	0.4789	-0.5354	0.5587	-0.8545	0.6859
13	-0.3614	0.4643	-0.5169	0.5484	-0.8186	0.6659
14	-0.3532	0.4482	-0.5002	0.5270	-0.7907	0.6574
15	-0.3458	0.4357	-0.4928	0.5161	-0.7752	0.6464
16	-0.3382	0.4217	-0.4762	0.4971	-0.7512	0.6188
17	-0.3331	0.4050	-0.4624	0.4823	-0.7353	0.6013
18	-0.3189	0.3997	-0.4491	0.4744	-0.6988	0.5971
19	-0.3124	0.3869	-0.4477	0.4568	-0.6747	0.5787

Таблица 2. Квантили для статистики Фортиана-Гране.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	-0.5393	1.0781	-0.4795	1.1710	-0.3829	1.3731
11	-0.5579	1.0815	-0.5014	1.1728	-0.4053	1.3628
12	-0.5764	1.0819	-0.5226	1.1757	-0.4314	1.3569
13	-0.5924	1.0953	-0.5356	1.1784	-0.4468	1.3521
14	-0.6082	1.0972	-0.5550	1.1801	-0.4645	1.3489
15	-0.6178	1.0987	-0.5656	1.1825	-0.4768	1.3429
16	-0.6272	1.0982	-0.5783	1.1857	-0.4917	1.3385
17	-0.6411	1.1018	-0.5896	1.1853	-0.5030	1.3315
18	-0.6529	1.1026	-0.6022	1.1884	-0.5175	1.3256
19	-0.6584	1.1065	-0.6135	1.1902	-0.5325	1.3204

Таблица 3. Квантили для статистики Фроцини.

n	$U_{0.1}$	$U_{0.05}$	$U_{0.01}$
	0.90	0.95	0.99
10	0.3299	0.3771	0.4705
11	0.3308	0.3779	0.4701
12	0.3337	0.3789	0.4729
13	0.3324	0.3792	0.4734
14	0.3349	0.3803	0.4779
15	0.3361	0.3806	0.4783
16	0.3366	0.3809	0.4721
17	0.3388	0.3807	0.4760
18	0.3390	0.3811	0.4717
19	0.3387	0.3813	0.4797

Таблица 4. Квантили для статистики Джини.

n	$U_{0.1}$	$U_{0.05}$	$U_{0.01}$
	0.90	0.95	0.99
10	0.6249	0.6571	0.7235
11	0.6158	0.6509	0.7126
12	0.6127	0.6439	0.6982
13	0.6054	0.6372	0.6891
14	0.6039	0.6313	0.6847
15	0.5991	0.6276	0.6788
16	0.5966	0.6229	0.6706
17	0.5928	0.6177	0.6640
18	0.5901	0.6141	0.6612
19	0.5877	0.6153	0.6576

Таблица 5. Квантили для статистики Гринвуда.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	1.3778	2.3570	1.3088	2.6320	1.2168	3.3560
11	1.3972	2.3626	1.3272	2.6476	1.2182	3.3136
12	1.4211	2.3695	1.3551	2.6415	1.2421	3.2995
13	1.4330	2.3781	1.3680	2.6311	1.2700	3.2831
14	1.4564	2.3869	1.3844	2.6459	1.2754	3.2729
15	1.4641	2.3675	1.3981	2.6205	1.2911	3.2685
16	1.4738	2.3864	1.4058	2.6324	1.3038	3.2414
17	1.4935	2.3607	1.4275	2.6267	1.3265	3.2247
18	1.5057	2.3876	1.4387	2.5976	1.3347	3.1936
19	1.5083	2.3668	1.4453	2.5818	1.3413	3.1768

Таблица 6. Квантили для статистики Морана.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	-0.2668	0.3286	-0.4016	0.3812	-0.6829	0.4539
11	-0.2682	0.3135	-0.3885	0.3659	-0.6313	0.4452
12	-0.2577	0.3016	-0.3705	0.3519	-0.6253	0.4304
13	-0.2424	0.3392	-0.3586	0.3392	-0.5898	0.4141
14	-0.2414	0.2940	-0.3445	0.3283	-0.5416	0.4052
15	-0.2389	0.2747	-0.3353	0.3195	-0.5475	0.3995
16	-0.2331	0.2631	-0.3241	0.3094	-0.5375	0.3815
17	-0.2254	0.2546	-0.3146	0.3025	-0.4848	0.3758
18	-0.2136	0.2444	-0.3069	0.2957	-0.4851	0.3682
19	-0.2183	0.2451	-0.3016	0.2863	-0.4591	0.3607

Таблица 7. Квантили для статистики Рао.

n	$U_{0.1}$	$U_{0.05}$	$U_{0.01}$
	0.90	0.95	0.99
10	2.7732	3.6995	5.9060
11	2.7326	3.7161	6.2492
12	2.7326	3.7389	6.1802
13	2.7012	3.7216	5.8212
14	2.7385	3.7172	6.3083
15	2.7246	3.7834	6.0894
16	2.7477	3.7563	6.2644
17	2.7860	3.7559	6.0034
18	2.7094	3.7620	6.2186
19	2.7240	3.7706	6.1426

Таблица 8. Квантили для статистики Шапиро-Уилка.

n	$U_{0.1}$		$U_{0.05}$		$U_{0.01}$	
	0.10	0.90	0.05	0.95	0.01	0.99
10	0.0378	0.1370	0.0311	0.1645	0.0211	0.2366
11	0.0362	0.1246	0.0302	0.1498	0.0207	0.2052
12	0.0351	0.1125	0.0292	0.1367	0.0206	0.1963
13	0.0330	0.0051	0.0283	0.1253	0.0203	0.1753
14	0.0324	0.0989	0.0275	0.1166	0.0202	0.1593
15	0.0311	0.0915	0.0265	0.1082	0.0195	0.1464
16	0.0298	0.0844	0.0256	0.1012	0.0192	0.1407
17	0.0285	0.0807	0.0248	0.0944	0.0187	0.1309
18	0.0277	0.0756	0.0241	0.0888	0.0185	0.1225
19	0.0273	0.0718	0.0235	0.0839	0.0183	0.1145