

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»

На правах рукописи



Исаева Елена Валерьевна

Восстановление функции плотности и оценивание параметров регрессионных
зависимостей на основе вейвлет-анализа

Специальность 05.13.17 –

«Теоретические основы информатики»

Диссертация на соискание учёной степени

кандидата технических наук

Научный руководитель:
доктор технических наук, доцент
Тимофеев Владимир Семенович

Новосибирск – 2022

Оглавление

Введение.....	6
Глава 1 Методы восстановления функции плотности и оценивания параметров регрессионных моделей.....	13
1.1 От анализа Фурье к вейвлет-анализу.....	13
1.1.1 Преобразование Фурье.....	14
1.1.2 Вейвлет и его свойства.....	17
1.1.3 Непрерывное вейвлет преобразование.....	19
1.1.4 Дискретное вейвлет преобразование.....	20
1.2 Методы оценивания функции плотности распределения	22
1.2.1 Понятие плотности распределения.....	22
1.2.2 Метод максимального правдоподобия.....	23
1.2.3 Метод моментов	24
1.2.4 Метод «Гребенка»	26
1.2.5 Гистограмма	28
1.2.6 Оценка Розенבלата – Парзена	28
1.2.7 Адаптивные оценки плотности	30
1.3 Классические методы оценивания параметров регрессионных моделей... ..	32
1.3.1 Задача регрессионного анализа.....	32
1.3.2 Метод максимального правдоподобия.....	33
1.3.3 Метод наименьших квадратов	35
1.4 Устойчивые методы оценивания параметров регрессионных моделей	36
1.5 Адаптивные методы оценивания параметров регрессионных моделей	38
1.5.1 Оценки, минимизирующие вектор остатков в L_p - метрике	38
1.5.2 Метод адаптивного оценивания параметров на основе обобщенного лямбда-распределения	38
1.6 Обзор программных систем оценивания параметров регрессионных уравнений и восстановления функции плотности.....	39
1.7 Обоснование цели и задачи исследования.....	43

Глава 2 Построение и исследование оценок функции плотности распределения с помощью преобразования Фурье и вейвлет-анализа	47
2.1 Оценка функции плотности распределения на основе различных ортонормированных базисов	47
2.2 Оценивание плотности распределения с помощью преобразования Фурье	49
2.3 Оценивание функции плотности распределения с помощью вейвлетов....	50
2.3.1 Оценивание функции плотности распределения с помощью ортогональных вейвлетов.....	51
2.3.2 Оценивание функции плотности распределения с помощью неортогональных вейвлетов.....	56
2.4 Исследование качества оценок функции плотности на основе преобразования Фурье	64
2.5 Исследование качества оценок функции плотности на основе ортогональных вейвлетов.....	68
2.5.1 Вейвлет Хаара	68
2.5.2 Вейвлет LITTLEWOOD & PALEY	75
2.5.3 Вейвлет Морле	80
2.6 Исследование качества оценок функции плотности на основе неортогональных вейвлетов.....	86
2.6.1 Вейвлет DOG.....	86
2.6.2 Вейвлет «Мексиканская шляпа»	94
2.7 Сравнительный анализ вейвлет-оценок плотности распределения	101
Выводы по главе 2.....	102
Глава 3 Алгоритмы адаптивного оценивания параметров регрессионных моделей с использованием вейвлетов и преобразования Фурье.....	104
3.1 Алгоритм оценивания параметров регрессионных моделей	104
3.2 Оценивание параметров регрессионных моделей с использованием преобразования Фурье	106

3.3	Оценивание параметров регрессионных моделей с использованием ортогональных вейвлетов.....	108
3.4	Оценивание параметров регрессионных моделей с использованием неортогональных вейвлетов.....	109
3.5	Исследование алгоритмов адаптивного оценивания параметров регрессионных моделей с использованием вейвлетов и преобразования Фурье	111
	Выводы по главе 3.....	116
Глава 4 Программная система адаптивного оценивания неизвестных параметров регрессионных моделей и восстановления функции плотности.....		
4.1	Структура программного комплекса.....	118
4.2	Режим работы программного комплекса	122
4.2.1	Восстановление функции плотности распределения	124
4.2.1.1	Режим «Оценивание функции плотности на основе преобразования Фурье».....	124
4.2.1.2	Режим «Оценивание функции плотности на основе вейвлет-анализа»	125
4.2.2	Оценивание параметров регрессионных моделей	127
4.2.2.1	Режим моделирования	127
4.2.2.2	Режим оценивания параметров.....	129
	Выводы по главе 4.....	130
Глава 5 Прогноз влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений		
5.1.	Применение методов оценивания регрессионных зависимостей при прогнозировании температурных изменений	131
5.2.	Постановка задачи	132
5.3.	Оценки параметров уравнениям температурных изменений	135
	Выводы по главе 5.....	140
	Заключение	141

Список литературы	142
Приложение А Интерфейс программного комплекса	152
Приложение Б Фактические данные температурного мониторинга состояния грунтов.....	159
Приложение В Свидетельства о регистрации программы для ЭВМ.....	162
Приложение Г Акты внедрения	163

Введение

Актуальность темы и степень ее разработанности. Методы теоретической и прикладной статистики нашли широкое применение в различных сферах, среди которых можно выделить оптимизацию сложных технологических процессов, сертификацию технических систем и изделий, геофизические, биологические, генетические и социологические исследования. Возникающие в реальной жизни задачи нередко заставляют исследователя сталкиваться с необходимостью поиска зависимости между входными данными, задающими условия функционирования, и выходными данными, которые характеризуют изучаемый объект. Решение таких задач может быть выполнено путем построения регрессионных моделей, и один из этапов состоит в оценивании неизвестных параметров. Такой подход позволяет не просто восстанавливать исходную зависимость, но и выполнять прогнозирование поведения изучаемого объекта.

Классические методы оценивания неизвестных параметров регрессионных зависимостей позволяют получать достаточно корректные и качественные результаты только при условии, что имеются достоверные предположения о свойствах случайной компоненты. Одним из них является метод максимального правдоподобия. Этот метод основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия. Метод максимального правдоподобия был рекомендован и значительно популяризирован Р. Фишером между 1912 и 1922 годом, хотя ранее он был использован К.Ф. Гауссом и П.С. Лапласом. Применение данного метода возможно при условии, что имеется достоверная информация о виде распределения случайных ошибок наблюдения. Предположение о нормальности распределения ошибок наблюдения позволяет применить метод наименьших квадратов и тем самым упростить поиск оценок. Метод был впервые применен К.Ф. Гауссом в 1795 году, а американский математик Р.А. Эдрейн в 1808-м рассмотрел его теоретико-вероятностные приложения. Работы А.А. Маркова в начале XX века позволили включить метод наименьших квадратов в теорию прикладной математической статистики, в которой он является

важной и естественной частью. На практике в большинстве случаев распределение случайной ошибки нельзя считать нормальным, и оценки, полученные таким образом, не позволяют сделать статистически корректные выводы и результаты. Это обстоятельство заставляет исследователей быть не только осторожными при использовании таких оценок, но и искать другие подходы к решению вопроса оценивания неизвестных параметров регрессионной зависимости. В этом случае на помощь исследователю приходят адаптивные методы оценивания параметров регрессионных зависимостей. В данной области можно отметить работы R.V. Hogg, R.V. Lenth, В.И. Мудров, В.И. Денисова. Огромное разнообразие встречающихся распределений случайной ошибки привело к идее восстановления неизвестной функции плотности на основе теории вейвлетов и преобразования Фурье с последующим применением метода максимального правдоподобия. Выбор в пользу такого подхода обусловлен широким использованием вейвлет-анализа и преобразования Фурье для аппроксимации различных функций. Основателем анализа Фурье является французский математик Жан Батист Жозеф Фурье, который сформулировал основы этой теории к 1807 году. С появлением все большего количества экспериментальных данных, которые на тот момент времени обрабатывались преобразованием Фурье, стало понятно: метод имеет ограниченные возможности в поиске закономерностей в этих данных, что привело к возникновению теории вейвлетов. Вейвлет-анализ является сравнительно новым направлением развития в прикладной математике. В 80-х годах Александр Гроссман и Жан Морле в ходе анализа сейсмических и акустических сигналов столкнулись с необходимостью введения нового термина – «вейвлет». До появления термина Альфред Хаар работал над системой базисных функций, которая обладает определенными свойствами и специфичными для вейвлетов признаками, а именно: локальной областью, ортогональностью, единичной нормой, нулевым средним, автомодельностью. В развитии теории вейвлет-анализа можно выделить работы И Добеши, К Чуи, Н.М. Астафьевой, О.В. Нагорнова.

Цель и задачи исследования. Целью настоящего исследования является разработка математического и алгоритмического обеспечения для восстановления

функции плотности с использованием вейвлет-анализа и преобразования Фурье и адаптивного оценивания параметров линейно-параметризованных регрессионных моделей. Для достижения этой цели были поставлены и решены следующие задачи:

1) вычислить нормы для следующих материнских вейвлетов: LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа» и построить системы базисных функций;

2) разработать и исследовать алгоритмы восстановления функции плотности с использованием ортогональных и неортогональных вейвлетов, а также на основе преобразования Фурье сформулировать рекомендации относительно их использования;

3) разработать и исследовать алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием ортогональных и неортогональных вейвлетов и преобразования Фурье;

4) разработать программный комплекс для адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения и использовать его для решения задачи мониторинга температурных изменений состояния грунта.

Область исследования. Содержание диссертации соответствует п. 5 области исследований «Разработка и исследование моделей и алгоритмов анализа данных, обнаружение закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» паспорта специальности 05.13.17 – «Теоретические основы информатики» (в области технических наук).

Методы исследования. Для решения поставленных задач использовались базовые понятия математического анализа, линейной алгебры, вычислительной математики, а также теория вероятностей, математическая статистика, регрессионный анализ, вейвлет-анализ, Фурье-анализ, методы оптимизации и методы статистического моделирования.

Достоверность и обоснованность научных положений, выводов и рекомендаций обеспечивается корректным применением аналитических методов

исследования, соответствием выводов и результатов вычислительных экспериментов известным теоретическим положениям.

Научная новизна работы состоит в следующем:

1) сформулирован и доказан ряд утверждений, обеспечивающих построение системы базисных функций для восстановления функции плотности распределения;

2) предложены новые алгоритмы, приводящие к повышению качества восстановления функции плотности, на основе ортогональных и неортогональных вейвлетов и преобразования Фурье, а также определены наилучшие значения параметра сглаживания для каждого из материнских вейвлетов;

3) разработаны алгоритмы, повышающие точность оценивания параметров линейно-параметризованных регрессионных моделей с использованием вейвлет-анализа и преобразования Фурье; получены выражения для логарифмической функции правдоподобия и ее производных на основе предложенных оценок функции плотности.

Теоретическая значимость состоит в развитии теории вейвлет-анализа с целью обеспечения возможности использования ненормированных вейвлетов для построения системы базисных функций, а также расширения возможностей регрессионного анализа за счет использования новых оценок функции плотности. В частности, получены выражения для логарифмической функции правдоподобия и ее производных. Такой подход позволяет восстанавливать регрессионные зависимости при условии отсутствия достоверной информации о виде распределения случайных ошибок наблюдения.

Практическая значимость работы заключается в разработанных алгоритмах оценивания параметров линейно-параметризованных регрессионных моделей, когда функция плотности восстанавливается на основе вейвлетов или преобразования Фурье. Разработан программный комплекс, в котором реализованы предложенные алгоритмы оценивания параметров регрессионных моделей и восстановления функции плотности, и зарегистрирован в виде объекта интеллектуальной собственности как программа для ЭВМ (№ гос. рег. 2022613412

от 14.03.2022) [49]. Применение программного комплекса позволило решить прикладную задачу технического характера о прогнозе влияния геокриологических последствий глобального потепления на устойчивость и долговечность жилых зданий и сооружений.

Реализация результатов работы. Результаты диссертационной работы нашли свое применение в ООО «Мерзлотный инженерно-строительный центр», в учебном процессе НГТУ и МАОУ «Инженерный лицей НГТУ», о чем имеются соответствующие акты внедрения.

Основные положения, выносимые на защиту. На защиту вынесены следующие положения:

1) утверждения об ортонормированности вейвлетов LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа», позволяющие выполнять построение системы базисных функций;

2) алгоритмы восстановления функции плотности на основе различных ортогональных и неортогональных вейвлетов и преобразования Фурье, а также результаты исследования качества оценивания функции плотности;

3) алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием восстановленных функций плотности, включая выражения логарифмической функции правдоподобия и ее производных, а также результаты их сравнительного анализа;

4) программный комплекс адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения, а также результаты решения задачи о прогнозе влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений.

Личный вклад автора в совместных публикациях заключается в:

1) доказательстве утверждений, касающихся свойств ортонормированности материнских вейвлетов и построении на их основе различных систем базисных функций;

2) построении и исследовании алгоритмов восстановления функции плотности распределения случайной величины на основе различных ортогональных и неортогональных вейвлетов;

3) разработке и исследовании алгоритмов адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием вейвлетов: Морле, LITTLEWOOD & PALEY, DOG, «Мексиканская шляпа» и на основе преобразования Фурье;

4) реализации программного комплекса WTiRM V1.0, позволяющего восстанавливать регрессионные зависимости и выполнять оценивание функции плотности, и его использовании для решения прикладных задач.

Апробация работы. Основные результаты работы докладывались и обсуждались на Всероссийской научной конференции молодых ученых «Наука. Технология. Инновации», г. Новосибирск, 2001 г., 2003 г., на Российской научно-технической конференции «Обработка информации и математическое моделирование», г. Новосибирск, 2020 г., 2022 г., на Международной конференции «Информационные технологии в бизнесе и производстве» (ITBI 2020), г. Новосибирск, 2020 г.

Публикации. Основные результаты исследований по теме диссертации опубликованы в 10 печатных работах [27-29, 42, 43, 49, 58- 60, 103], из них 3 работы опубликованы в изданиях, рекомендованных ВАК [58-60], 1 работа опубликована в издании, индексируемом в базе данных Scopus [103], получено свидетельство о государственной регистрации программы для ЭВМ [49].

Структура и объём работы. По структуре настоящая работа состоит из введения, пяти глав основного содержания, заключения, списка литературы, состоящего из 107 источников, четырёх приложений. Диссертация изложена на 165 страницах основного текста, содержит 41 рисунок и 27 таблиц.

Краткое содержание работы. В первой главе описываются базовые определения преобразования Фурье и вейвлет-анализа. Рассмотрены методы оценивания вектора неизвестных параметров уравнения регрессионной зависимости, которые пользуются наибольшей популярностью. К таким методам

относятся классические, устойчивые и адаптивные методы. Проведен обзор существующих программных систем, которые можно использовать для оценивания параметров регрессионной зависимости и восстановления функции плотности. Приводится обоснование цели и задач исследования.

Во второй главе рассматривается задача построения алгоритмов для оценивания функции плотности распределения на основе различных ортонормированных базисов, таких как тригонометрическая система функций, система функций на основе ортогональных и неортогональных вейвлетов. Для построения базисных функций были вычислены нормы для вейвлетов Морле, LITTLEWOOD & PALEY, DOG, «Мексиканская шляпа». Сформулирован и доказан ряд утверждений, касающихся свойств ортонормированности материнских вейвлетов. Предложены способы, позволяющие выполнить поправку на ортонормированность материнских вейвлетов. Проведены исследования точности оценивания функции плотности с использованием рассмотренных базисов.

Третья глава посвящена разработке алгоритмов адаптивного оценивания параметров линейно-параметризованных регрессионных моделей на основе преобразования ортогональных и неортогональных вейвлетов, а также преобразования Фурье. Получены выражения для логарифмических функций правдоподобия и их производных. Исследована точность оценивания неизвестных параметров регрессионных зависимостей с помощью разработанных алгоритмов.

Четвертая глава включает в себя описание программного комплекса WTiRM V1.0, в котором реализованы разработанные алгоритмы оценивания неизвестных параметров регрессионных моделей и восстановления функции плотности.

В пятой главе показано применение разработанных алгоритмов оценивания неизвестных параметров регрессионной зависимости в реальной технической задаче прогнозирования температурного состояния грунтов, что позволяет своевременно реагировать на нарушения мерзлотно-грунтовых условий для зданий и сооружений, построенных в условиях крайнего севера.

Глава 1 Методы восстановления функции плотности и оценивания параметров регрессионных моделей

В этой главе рассматриваются базовые понятия Фурье анализа, а также основы вейвлет-анализа, практические аспекты их применения при решении задач, связанных с обработкой экспериментальных данных. Проанализированы наиболее известные подходы для поиска оценок вектора неизвестных параметров уравнения регрессии, а именно классические методы, такие как метод максимального правдоподобия, метод наименьших квадратов, а также устойчивые и адаптивные методы. Выполнен обзор существующих программных средств, применимых для восстановления функции плотности распределения и для оценивания параметров регрессионных зависимостей.

1.1 От анализа Фурье к вейвлет-анализу

Анализ Фурье является хорошо разработанным инструментом, который играет одну из центральных ролей в математическом анализе и его приложениях. Преобразование Фурье и ряды Фурье имеют фундаментальное значение и особенно привлекательны за счет своих свойств ортогональности и простоты их выражения на основе двух функций $\sin x$ и $\cos x$. Теория вейвлет-анализа появилась в связи с необходимостью обработки большого количества экспериментальных данных, в которых стандартный и хорошо развитый метод преобразований Фурье показал свою ограниченность. Например, вейвлет преобразование позволяет анализировать свойства сигналов одновременно и в физической, и в частотной области, в отличие от преобразования Фурье. Свою роль сыграло и бурное развитие вычислительной техники, что позволило численно решать такие задачи, которые до этого были просто нереализуемыми. Так же, как и в анализе Фурье, вейвлет-анализ состоит из двух разделов: вейвлет-преобразование и вейвлет-ряды. Вейвлет преобразование определяется как свертка с учетом растяжения некоторой функции (базисный вейвлет), а вейвлет-ряд выражается через единственную функцию с помощью двух

операций: двоичных сжатий и целочисленных сдвигов. В отличие от анализа Фурье, вейвлет-преобразование тесно связано с вейвлет-рядом.

1.1.1 Преобразование Фурье

Обозначим через $L^2(0, 2\pi)$ пространство всех измеримых функций $f(t)$, определенных на интервале $(0, 2\pi)$, и таких, что выполняется следующее условие [10-12, 20, 23, 34, 36, 70]:

$$\int_0^{2\pi} |f(t)|^2 dt < \infty. \quad (1.1)$$

Функции из $L^2(0, 2\pi)$ периодически продолжаемы на всю вещественную ось $R = (-\infty, \infty)$, а именно $f(t) = f(t - 2\pi)$ для всех t [19, 34, 70].

Любую $f(t)$ из пространства $L^2(0, 2\pi)$ можно представить рядом Фурье [4, 20, 22, 32, 70]

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{int}, \quad (1.2)$$

где константные значения c_n – коэффициенты Фурье, которые определяются формулой:

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{int} dt.$$

Сходимость ряда (1.2) дает следующее равенство [10, 20, 70]:

$$\lim_{M, N \rightarrow +\infty} \int_0^{2\pi} \left| f(t) - \sum_{n=-M}^N c_n e^{int} \right|^2 dt = 0.$$

Напомним понятия скалярного произведения и нормы в пространстве $L^2(R)$, которые будут встречаться в дальнейшем [4, 20, 33, 70]:

1) скалярное произведение в пространстве $L^2(R)$ определяется следующим образом:

$$\langle p, q \rangle = \int_{-\infty}^{+\infty} p(t)q^*(t)dt, \quad (1.3)$$

где символ звездочка означает операцию комплексного сопряжения, $p(t), q(t) \in L^2(R)$;

2) норма $\|p\|$ в пространстве $L^2(R)$ определяется в следующем виде

$$\|p\| = \sqrt{\langle p, p \rangle}. \quad (1.4)$$

Обратим внимание, что $w_n(t) = e^{int}$, $n = \dots, -1, 0, 1, \dots$ удовлетворяет двум условиям [4, 33, 35, 36, 70]:

$$1) \langle w_m, w_n \rangle = 0 \text{ для всех } m \neq n;$$

$$2) \|w_n\| = 1.$$

Таким образом система функций $w_n(t)$ образует ортонормированный базис в пространстве $L^2(0, 2\pi)$, построенный с помощью масштабного преобразования единственной функции $w(t) = e^{it}$ и определяется в виде $w_n(t) = w(nt)$. Исходя из свойств $w_n(t)$ следует, что для коэффициентов ряда Фурье выполняется равенство Парсеваля [4, 20, 70]:

$$(2\pi)^{-1} \int_0^{2\pi} |f(t)|^2 dt = \sum_{n=-\infty}^{+\infty} |c_n|^2.$$

Принимая во внимание (1.1), можно сказать, что каждая 2π – периодическая интегрируемая с квадратом функция может быть получена комбинацией целочисленных растяжений функции $w(t) = \cos t + i \sin t$, которая является синусоидальной волной.

Рассмотрим функцию $f(t) \in L^2(R)$ как аналоговый сигнал, тогда преобразование Фурье [23, 33, 70]

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt,$$

где $\hat{f}(\omega)$ – спектральная информация об аналоговом сигнале $f(t)$, ω – частота. Преобразование Фурье является основой спектрального анализа и позволяет характеризовать частотную составляющую измеряемого сигнала.

Аналогично равенству Парсеваля для рядов Фурье, равенство Парсеваля, которое описывает связь между функциями из $L^2(R)$ и их преобразованием Фурье, выражается формулой [20, 70]

$$\langle f, g \rangle = (2\pi)^{-1} \langle \tilde{f}(t), \tilde{g}(t) \rangle,$$

где $f(t), g(t) \in L^2(R)$.

Для функций, которые равны нулю вне заданного интервала, можно вычислить ряд Фурье на любом интервале, включающем точки, где функция не равна нулю. Преобразование Фурье так же определено для такой функции. При увеличении длины интервала, на котором вычисляется ряд Фурье, коэффициенты ряда начинают напоминать преобразование Фурье. Такой вид называют дискретным преобразованием Фурье [10 – 12, 20, 70]:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i n k}{N}}, \quad (1.5)$$

где N – количество значений сигнала, измеренных за период, и количество компонент разложения. x_n – измеренные значения сигнала, $n = 0, \dots, N-1$. X_k – комплексная амплитуда синусоидального сигнала, $k = 0, \dots, N-1$, k – индекс частоты. Частота k – го сигнала равна $\frac{k}{T}$, где T – период времени, в течение которого брались данные. Коэффициенты дискретного преобразования Фурье представляют собой последовательность конечной длины, и по ним можно точно восстановить исходную последовательность. Обратное дискретное преобразование задается с помощью выражения [10 – 12, 20, 70]

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i n k}{N}}.$$

Дискретное преобразование обладает таким свойством, как линейность. Если двум последовательностям x_n и y_n с одинаковым периодом соответствуют наборы

X_k и Y_k , то последовательность $\{ax_n + by_n\}$ будет соответствовать спектру $aX_k + bY_k$, где a и b – произвольные константы. Так же дискретное преобразование Фурье обладает свойством симметрии.

Дискретное преобразование Фурье может быть сильно упрощено, если использовать свойства симметрии и периодичность коэффициентов, разделив вычисления в (1.5) разделить на две меньшие части:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i n k}{N}} = \sum_{m=0}^{\frac{N}{2}-1} x_{2m} e^{-\frac{2\pi i (2m)k}{N}} + e^{-\frac{2\pi i k}{N}} \sum_{m=0}^{\frac{N}{2}-1} x_{2m+1} e^{-\frac{2\pi i (2m)k}{N}}, \quad (1.6)$$

где x_{2m} и x_{2m+1} соответственно четные и нечетные значения сигнала. Получилось две суммы, которые похожи на меньшие дискретные преобразования Фурье, со значениями с четным номером и со значениями с нечетным номером. Соотношение (1.6) определяет быстрое преобразование Фурье [10 – 12, 20, 70].

1.1.2 Вейвлет и его свойства

В случае рассмотрения пространства $L^2(R)$ измеримых функций $f(t)$, определенных на всей числовой оси, становится очевидным, что пространства $L^2(0, 2\pi)$ и $L^2(R)$ совершенно различные. Синусоидальная волна не принадлежит $L^2(R)$ и, следовательно, семейство $w_n(t)$ не может быть базисом функционального пространства $L^2(R)$. Поэтому возникает необходимость определить достаточно простые функции для конструирования базиса пространства $L^2(R)$.

Вейвлетом называется нетривиальная функция $\psi(t)$ из пространства $L^2(R)$, удовлетворяющая следующим свойствам [10, 31, 39, 44, 45, 70].

1. Локализация. Базисные вейвлет-функции должны быть определены на конечном интервале как во временной, так и в частотной области [10 – 12, 70].

2. Нулевое среднее. Исходная вейвлет-функция осциллирует вокруг оси времени и имеет нулевую площадь

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (1.7)$$

3. Ограниченность. Базисная вейвлет-функция обладает конечной энергией

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty. \quad (1.8)$$

4. Автомодельность (самоподобие) базиса [10 – 12, 70]. В случае сжатия (растяжения) и сдвига материнского вейвлета $\psi(t)$ число осцилляций не изменяется.

Простейшим примером вейвлета является функция, предложенная А. Хааром и впоследствии получившая название вейвлета Хаара (HAAR - вейвлет) [74 – 77, 81, 100 – 102]:

$$\psi(t) = \begin{cases} 1, & t \in [0, 0.5), \\ -1, & t \in [0.5, 1), \\ 0, & t \notin (0, 1]. \end{cases} \quad (1.9)$$

Вейвлеты на основе производных функции Гаусса [10, 11, 69, 70, 107] выражаются соотношением

$$\psi(t) = (-1)^n \frac{d^n}{dt^n} e^{-\frac{t^2}{2}} \quad (1.10)$$

и обладают наилучшими показателями локализации как во временной, так и в частотной области. В случае $n = 1$ получается WAVE-вейвлет [20, 39, 69, 104, 105]:

$$\psi(t) = -te^{-\frac{t^2}{2}}. \quad (1.11)$$

При $n = 2$ получаем вейвлет, называемый «Мексиканской шляпой» или МНАТ-вейвлет [20, 39, 69, 104, 105]:

$$\psi(t) = (1 - t^2)e^{-\frac{t^2}{2}}. \quad (1.12)$$

Функция Гаусса образует также DOG - вейвлет как разность Гауссианов [20, 39, 69, 104, 105]:

$$\psi(t) = e^{-\frac{t^2}{2}} - \frac{1}{2}e^{-\frac{t^2}{8}}. \quad (1.13)$$

Основные достоинства данных вейвлетов – симметрия или антисимметрия при нечетных значениях производных для вейвлета Гаусса.

Широкое распространение получили вейвлет Морле [10, 11, 12, 20, 70]

$$\psi(t) = e^{-\frac{t^2}{2}} \cos 5t \quad (1.14)$$

и вейвлет LITTLEWOOD & PALEY [10, 11, 12, 20, 70]

$$\psi(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}. \quad (1.15)$$

Так же как и в случае $L^2(0, 2\pi)$, где одна функция $w(t) = e^{it}$ порождает целое пространство, каждая из функций $\psi(t)$ (1.6) – (1.10) может быть использована для порождения всего пространства $L^2(R)$ путем целочисленных сдвигов, а именно $\psi(t - k)$, $k \in Z$.

1.1.3 Непрерывное вейвлет-преобразование

Рассмотрим пространство $L^2(R)$ функций $f(t)$, определенных на всей действительной оси R и обладающих конечной нормой (1.6). Базис функционального пространства $L^2(R)$, построенный с помощью непрерывных масштабных преобразований и переносов вейвлета $\psi(t)$ с произвольными значениями базисных параметров, описывается следующим соотношением [10 – 12, 31, 32, 71]:

$$\Psi_{ab}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad a, b \in R, \quad \psi \in L^2(R). \quad (1.16)$$

Множитель $|a|^{-1/2}$ обеспечивает независимость нормы базисных функций от масштабирующего числа [44, 45, 48, 31, 32, 69, 71]. Тогда прямое вейвлет-преобразование будет представлено в виде [10 – 12, 44]:

$$W(a,b) = \int_{-\infty}^{+\infty} f(t) \psi_{ab}(t) dt,$$

где $a, b \in R, a \neq 0$ и выполняется условие (1.7).

Вычислив коэффициенты $W(a,b)$, можно восстановить исходный вид функции $f(t)$ с помощью обратного интегрального вейвлет-преобразования, которое задается выражением [10 – 12, 31, 32, 71]:

$$f(t) = \frac{1}{\sqrt{|C_\psi|}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W(a,b) \psi_{ab}(t) \frac{da db}{a^2},$$

где постоянная $C_\psi < \infty$, зависит только от ψ и задается соотношением

$$C_\psi = 2\pi \int_{-\infty}^{+\infty} |\xi|^{-1} |\widehat{\psi}(\xi)|^2 d\xi. \quad (1.17)$$

Формулу (1.17) можно использовать для восстановления $f(t)$, если известно её вейвлет-преобразование $W(a,b)$, или как способ записи в виде суперпозиции вейвлетов $\psi_{ab}(t)$, коэффициенты в этой суперпозиции точно заданы через вейвлет-преобразование $f(t)$.

1.1.4 Дискретное вейвлет-преобразование

Параметры a и b , приведенные в (1.16), непрерывно определены в R , что является избыточным, так как в пределе имеется бесконечное множество копий порождающего вейвлета. Вследствие полученной избыточности возникает необходимость дискретизации параметров, которая осуществляется следующим образом [10 – 12, 31, 32, 78, 81-83]:

$$a = a_0^m, \quad b = nb_0 a_0^m, \quad a_0 > 1, \quad b_0 > 0, \quad m, n \in Z.$$

Здесь b_0 – начальное значение, которое выбирается произвольно, поэтому можно принять $b_0 = 1$. Параметр смещения имеет зависимость от масштаба, из чего

следует, что чем больше масштаб, тем больше сдвиг. В дискретном представлении базис пространства $L^2(R)$ представляется так [10 – 12, 31, 32, 78, 81-83]:

$$\Psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi(a_0^{-m}t - n), \quad (1.18)$$

где $m, n \in Z$, $\psi(t) \in L^2(R)$. Отметим, что дискретизация параметров выполняется с сохранением ортогонализации преобразования, которая является следствием ортонормированности $\Psi_{m,n}(t)$. Согласно определению, ортонормированная система функций должна удовлетворять двум условиям [4, 44, 46, 70, 92, 93, 95]:

$$1) \langle \Psi_{m,n}, \Psi_{i,j} \rangle = \delta_{m,i} \cdot \delta_{n,j}, \quad (1.19)$$

$$2) \|\Psi_{m,n}\| = 1, \quad (1.20)$$

где $\delta_{m,n} = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases}$, $m, n, i, j \in Z$.

Из соотношения (1.18) и условия (1.20) получаем

$$\|\Psi_{m,n}(t)\| = \left\| a_0^{-\frac{m}{2}} \psi(a_0^{-m}t - n) \right\| = \left\{ a_0^{-m} \int_{-\infty}^{+\infty} |\psi(a_0^{-m}t - n)|^2 dt \right\}^{\frac{1}{2}} = \|\psi(t)\|, \quad (1.21)$$

где $\psi(t) \in L^2(R)$. Таким образом, выполнение условия (1.20) семейством $\Psi_{m,n}(t)$ зависит от выбора порождающего их материнского вейвлета $\psi(t)$. Если материнский вейвлет нормированный, то и система функций $\Psi_{m,n}(t)$ также обладает этим свойством. Примером простейшего ортонормированного вейвлета является вейвлет Хаара [70, 76, 88, 89, 91, 95, 101]. Тогда прямое вейвлет-преобразование будет представлено в виде [10-12, 20, 70]

$$C_{m,n} = \int_{-\infty}^{+\infty} f(t) \Psi_{m,n}(t) dt,$$

где $m \in Z, n \in Z$. Существование обратного преобразования для дискретного вейвлет-преобразования определяется с использованием неравенства Рисса [20, 70]. Если для базиса на основе выражения (1.18) существуют две положительные постоянные A и B , $0 < A \leq B < \infty$, такие что

$$A \|\{C_{m,n}\}\|^2 \leq \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \|C_{m,n} \psi_{m,n}(t)\|^2 \leq B \|\{C_{m,n}\}\|^2,$$

$$\|\{C_{m,n}\}\|^2 = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} |C_{m,n}|^2 < \infty,$$

то функция $\psi(t) \in L^2(R)$ называется R -функцией, а последовательность $\{\psi_{m,n}(t)\}$ – базисом Рисса. В этом случае возможна полная реконструкция функции, которая выражается разложением в ряд

$$f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{m,n} \psi_{m,n}(t),$$

где $m \in Z, n \in Z$.

С целью создания эффективных алгоритмов для восстановления $f(t)$ частотную ось разбивают на диапазоны, использующие степени 2 для масштабного параметра a . В этом случае семейство базисных функций $\psi_{m,n}(t)$ образует ортогональный базис [23, 33, 35, 45, 70].

1.2 Методы оценивания функции плотности распределения

1.2.1 Понятие плотности распределения

Числовая величина ζ , значения которой зависят от элементарных исходов $w \in \Omega$ (Ω – пространство элементарных событий), называется случайной величиной [14, 15, 34, 38]. Распределение вероятностей случайной величины задано, если определены вероятности $P\{a \leq \zeta \leq b\}$ всевозможных событий $\{a \leq \zeta \leq b\}$, каждое из которых означает, что ζ принимает одно из значений в соответствующих пределах: $a \leq \tilde{x} \leq b, a, b \in R$ [1, 2, 6, 9].

Случайная величина ζ имеет непрерывное распределение вероятностей [6, 9], если для любых a и b ($a \leq b$)

$$P\{a \leq \zeta \leq b\} = \int_a^b f(\tilde{x}) d\tilde{x},$$

где $f(\tilde{x})$ – некоторая неотрицательная интегрируемая функция, $\int_{-\infty}^{+\infty} f(\tilde{x}) d\tilde{x} = 1$,

называемая плотностью распределения случайной величины ζ [14, 15, 25, 34].

Основные свойства функции плотности распределения [25, 34, 38, 50]:

- 1) $f(\tilde{x}) \geq 0$;
- 2) $\int_{-\infty}^{+\infty} f(\tilde{x}) d\tilde{x} = 1$;
- 3) $\int_a^b f(\tilde{x}) d\tilde{x} = P\{a \leq \zeta \leq b\}$.

Функция, определенная как $F(\tilde{x}) = P(\zeta < \tilde{x})$, $\tilde{x} \in R$ называется функцией распределения случайной величины ζ [14, 15, 25, 34].

В практическом применении функция плотности распределения $f(\tilde{x})$ является непрерывной почти всюду на области определения, поэтому справедливо равенство $f(\tilde{x}) = F'(\tilde{x})$, $\tilde{x} \in R$. Плотность распределения для непрерывных случайных величин является одной из основных характеристик случайной величины, она определяет закон распределения случайной величины. Оценивание функции плотности распределения является важнейшей задачей для решения которой используют параметрические методы оценивания или непараметрические методы.

1.2.2 Метод максимального правдоподобия

Метод максимального или наибольшего правдоподобия [2, 21, 51, 52] предложен Р. Фишером [66]. С помощью этого метода производится точечная оценка неизвестных параметров априорно известного закона распределения случайной величины.

Если функция распределения $F(\tilde{x}, \theta) = P(\zeta < \tilde{x})$ имеет плотность $f(\tilde{x}, \theta)$, то функцией правдоподобия непрерывной случайной величины ζ называют функцию аргумента θ :

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta).$$

Логарифмическая функция правдоподобия определяется соотношением

$$\ln L(y, \theta) = \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) = \sum_{i=1}^n \ln (f(x_i, \theta)).$$

В качестве оценки параметра θ принимают такое его значение $\theta^* = \theta^*(x_1, x_2, \dots, x_n)$, при котором логарифмическая функция правдоподобия достигает максимума. Оценку θ^* называют оценкой максимального правдоподобия [2, 21, 51, 52].

В случае если плотность распределения непрерывной случайной величины ζ определяется двумя неизвестными параметрами θ_1, θ_2 , то функция правдоподобия является функцией двух независимых аргументов

$$L(x_1, x_2, \dots, x_n, \theta_1, \theta_2) = f(x_1, \theta_1, \theta_2) f(x_2, \theta_1, \theta_2) \dots f(x_n, \theta_1, \theta_2),$$

а логарифмическая функция правдоподобия принимает вид

$$\ln L(y, \theta) = \ln \left(\prod_{i=1}^n f(x_i, \theta_1, \theta_2) \right) = \sum_{i=1}^n \ln (f(x_i, \theta_1, \theta_2)).$$

Оценки, полученные методом максимального правдоподобия, состоятельны, распределены асимптотически нормально при больших значениях n ($n \rightarrow \infty$). Недостаток метода состоит в том, что при его использовании требуется информация о законе распределения случайной величины.

1.2.3 Метод моментов

Метод моментов, предложенный английским статистиком Карлом Пирсоном в 1894 г. [47], заключается в приравнивании определенного числа выборочных моментов к соответствующим теоретическим, которые являются функциями неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$. Рассматривая количество моментов, равное

числу k неизвестных параметров, подлежащих определению, и решая полученные уравнения относительно этих параметров, получаем искомые оценки. Иначе говоря, оценки параметров $\theta_1, \theta_2, \dots, \theta_k$ являются решениями систем уравнений

$$\alpha_i(\theta_1, \theta_2, \dots, \theta_k) = \hat{\alpha}_i, i = i_1, \dots, i_k.$$

На практике можно использовать универсальное GL-распределение [53], которое позволяет описать такие распределения, как нормальное, экспоненциальное, Вейбулла, гамма-, бета- и другие [30, 53]. Функция GL - распределения зависит от четырех параметров $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ и может быть определена в терминах квантилей распределения [87] выражением

$$Q_{\varepsilon_i}(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{1}{\lambda_2} \left[\frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4} \right], 0 \leq u \leq 1, \quad (1.22)$$

$$t = Q_{\varepsilon_i}(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4), i = 1, \dots, n.$$

Функция плотности принимает вид [86]

$$h(t) = \frac{\lambda_2}{u^{\lambda_3-1} + (1-u)^{\lambda_4-1}}, 0 \leq u \leq 1.$$

Функция плотности GL - распределения либо не имеет экстремума, либо имеет один или два экстремума [87], что определяется значением параметров λ_3 и λ_4 . Согласно [87] GL-распределение можно полностью определить его первыми четырьмя моментами. Поэтому для вычисления оценок параметров распределения можно использовать метод моментов [34, 40, 41, 52], выраженный следующей пошаговой процедурой.

Шаг 1. Вычислить выборочные моменты для выборки $\{Z_i | i = \overline{1, n}\}$ [34]:

$$\hat{\alpha}_1 = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \hat{\alpha}_2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2, \quad \hat{\alpha}_3 = \frac{1}{n\hat{\sigma}^3} \sum_{i=1}^n (Z_i - \bar{Z})^3, \quad \hat{\alpha}_4 = \frac{1}{n\hat{\sigma}^4} \sum_{i=1}^n (Z_i - \bar{Z})^4;$$

Шаг 2. Приравнять значения этих моментов к значениям теоретических для GL - распределения, которые вычисляются по формулам [86]:

$$\alpha_1 = \lambda_1 + \frac{A}{\lambda_2}, \quad \alpha_2 = \frac{1}{\lambda_2^2} (B - A^2),$$

$$\alpha_3 = \frac{1}{\lambda_2^3 \sigma^3} (C - 3AB + 2A^3), \quad \alpha_4 = \frac{1}{\lambda_2^4 \sigma^4} (D - 4AC + 6A^2B - 3A^4),$$

где $A = \frac{1}{\lambda_3(1+\lambda_3)} + \frac{A}{\lambda_4(1+\lambda_4)},$

$$B = \frac{1}{\lambda_3^2(1+2\lambda_3)} + \frac{A}{\lambda_4^2(1+2\lambda_4)} - \frac{2}{\lambda_3\lambda_4} \beta(1+\lambda_3, 1+\lambda_4),$$

$$C = \frac{1}{\lambda_3^3(1+3\lambda_3)} + \frac{A}{\lambda_4^3(1+3\lambda_4)} - \frac{3}{\lambda_3^2\lambda_4} \beta(1+2\lambda_3, 1+\lambda_4) + \frac{3}{\lambda_4^2\lambda_3} \beta(1+\lambda_3, 1+2\lambda_4),$$

$$D = \frac{1}{\lambda_3^4(1+4\lambda_3)} + \frac{A}{\lambda_4^4(1+4\lambda_4)} - \frac{4}{\lambda_3^3\lambda_4} \beta(1+3\lambda_3, 1+\lambda_4) + \frac{6}{\lambda_3^2\lambda_4^2} \beta(1+2\lambda_3, 1+2\lambda_4) +$$

$$+ \frac{4}{\lambda_4^3\lambda_3} \beta(1+\lambda_3, 1+3\lambda_4).$$

Шаг 3. Решить систему уравнений

$$\alpha_i = \hat{\alpha}_i, \quad i = 1, \dots, 4 \quad (1.23)$$

и найти оценки параметров $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.

Отметим, что система уравнений (1.23) имеет решение, если для распределения существуют и конечны первые четыре момента [34]. Такое условие выполняется для распределений, в семействе GL - распределения [87].

1.2.4 Метод «Гребенка»

Пусть $F(\tilde{x})$ – непрерывная функция, где $\tilde{x} \in R$, тогда плотность распределения $f(\tilde{x}) = F'(\tilde{x})$. В качестве оценки функции плотности возьмем $\hat{f}_n(\tilde{x}) = \hat{F}'_n(\tilde{x})$, где $\hat{F}_n(\tilde{x})$ – некоторая оценка функции распределения. Возьмем в качестве $\hat{F}_n(\tilde{x})$ – эмпирическую функцию распределения. Производная от эмпирической функции приводит к следующему результату:

$$\hat{f}_n(\tilde{x}) = \hat{F}'_n(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n \delta(\tilde{x} - x_i), \quad (1.24)$$

$$\text{где } \delta(\tilde{x}) = C'(\tilde{x}) = \begin{cases} 0, & \tilde{x} \neq 0, \\ \infty, & \tilde{x} = 0 \end{cases}, \quad C(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0 \end{cases}.$$

Дельта-функция Дирака $\delta(\tilde{x})$ [33], а также порождающая её функция Хевисайда $C(t)$ [33] относятся к классу обобщенных функций, которые определяются как предел последовательности дифференцируемых функций. В частности, $\delta(\tilde{x})$ можно представить в виде:

$$\tilde{\delta}(t, \lambda) = \frac{1}{h} k\left(\frac{t}{h}\right),$$

где h – параметр размытости и $k(t)$ – функция, удовлетворяющая следующим ограничениям:

- 1) $k(t) > 0$;
- 2) $k(t) = k(-t)$;
- 3) $\int_{-\infty}^{+\infty} k(t) dt = 1$.

В качестве $k(t)$ можно взять любую непрерывную симметричную функцию плотности распределения. Если $h \rightarrow 0$, то $\tilde{\delta}(t, \lambda) = \delta(t)$. Обобщенные функции обладают рядом свойств $\delta(t)$:

$$\begin{aligned} \text{а) } \delta(-t) &= \delta(t); & \text{в) } \int_{-\infty}^{+\infty} \delta(t) dt &= 1; \end{aligned} \tag{1.25}$$

$$\begin{aligned} \text{б) } \delta(ct) &= \frac{1}{c} \delta(t); & \text{г) } \int_{-\infty}^{+\infty} \varphi(x) \delta(t-x) dt &= \varphi(t). \end{aligned}$$

Рассмотрим оценку $\hat{f}_n(\tilde{x})$, определяемую соотношением (1.24), которая представляет собой сумму бесконечных всплесков в точках выборочных наблюдений x_i , $i = 1, \dots, n$ – «гребенка». Используя свойство (1.25 г) для $\delta(t)$, можно показать, что $\hat{f}_n(\tilde{x})$ является несмещенной оценкой $f(\tilde{x})$, но не является состоятельной [25, 34, 38, 51].

1.2.5 Гистограмма

Запишем плотность в виде [23, 35]

$$f(\tilde{x}) = F'(\tilde{x}) = \lim_{\Delta\tilde{x} \rightarrow 0} \frac{F(\tilde{x} + \Delta\tilde{x}) - F(\tilde{x})}{\Delta\tilde{x}}$$

и в качестве оценки $f(\tilde{x})$ возьмем

$$\hat{f}_n(\tilde{x}) = \frac{F_n(\tilde{x} + \Delta\tilde{x}) - F_n(\tilde{x})}{\Delta\tilde{x}} = \frac{1}{n\Delta\tilde{x}} \sum_{i=1}^n C(\tilde{x} + \Delta\tilde{x} - x_i) - C(\tilde{x} - x_i) = \frac{p_{\tilde{x}}}{n\Delta\tilde{x}},$$

где $p_{\tilde{x}}$ – количество выборочных значений, попавших в интервал $(\tilde{x}, \tilde{x} + \Delta\tilde{x}]$.

В результате приходим к следующему алгоритму оценки плотности.

Шаг 1. Разбить интервал определения $\zeta \in [a, b]$ на m непересекающихся интервалов D_1, \dots, D_m шириной $\Delta\tilde{x}$ каждый.

Шаг 2. Подсчитать количество выборочных значений p_d , попавших в интервал D_d .

Шаг 3. Получить оценку $\hat{f}_n(\tilde{x}) : \frac{p_d}{n\Delta\tilde{x}}, \tilde{x} \in D_d, d = 1, \dots, m$, которая называется гистограммой [1, 2, 9, 25].

Исследования показывают, что если $\Delta\tilde{x}$ не стремится к нулю, то гистограмма будет смещенной оценкой. Гистограмма проста в вычислительном отношении и имеет наглядный графический вид, что во многом и определяет ее популярность.

1.2.6 Оценка Розенבלата – Парзена

Рассмотрим в качестве оценки плотности оценку вида $\hat{f}_n(\tilde{x}) = \hat{F}'_n(\tilde{x})$, где

$$\hat{F}_n(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{\tilde{x} - x_i}{h_n}\right),$$

а в качестве $K(t)$ можно взять любую непрерывную и

симметричную функцию распределения, h_n - параметр размытости ($h_n \rightarrow 0$ при $n \rightarrow \infty$). В результате получим

$$\hat{f}_n(\tilde{x}) = \hat{F}'_n(\tilde{x}) = \frac{1}{n} \sum_{i=1}^n K' \left(\frac{\tilde{x} - x_i}{h_n} \right) = \frac{1}{n h_n} \sum_{i=1}^n k \left(\frac{\tilde{x} - x_i}{h_n} \right), \quad (1.26)$$

где $k(t) = K'(t)$ – плотность распределения $K(t)$, которую называют ядерной функцией, или ядром. С другой стороны, используя свойство (1.25 г) дельта-функции, можно записать

$$f(\tilde{x}) = \int \delta(\tilde{x} - t) f(t) dt = \int \delta(\tilde{x} - t) d\hat{F}(t) = \int \lim_{\lambda \rightarrow 0} \delta(\tilde{x} - t, h) d\hat{F}(t).$$

Применяя метод подстановки, получаем

$$\hat{f}_n(\tilde{x}) = \frac{1}{h_n} \int k \left(\frac{\tilde{x} - x_i}{h_n} \right) dF_n(t) = \frac{1}{n h_n} \sum_{i=1}^n k \left(\frac{\tilde{x} - x_i}{h_n} \right). \quad (1.27)$$

Из свойств дельта-функции можно сделать качественный вывод, что в формировании оценки $\hat{f}_n(\tilde{x})$ в точке \tilde{x} участвуют только те выборочные значения, которые находятся в ближайшей окрестности точки \tilde{x} (т. е. оценка носит локальный характер). Такие оценки плотности, предложенные М. Розенблатом (1957) и исследованные Э. Парзенем (1962), получили название оценок плотности Розенבלата – Парзена [82]. Оценки (1.26) и (1.27) были получены в соответствии с методом подстановки как производные от оценок сглаженных эмпирических функций распределения, поэтому их свойства изучены исходя из свойств сглаженных эмпирических функций распределения. Оценка является [2, 9, 25]:

1) асимптотически несмещенной;

2) состоятельной;

3) $\sqrt{n h_n} \left[\hat{f}_n(\tilde{x}) - E \hat{f}_n(\tilde{x}) \right] \Leftrightarrow \Phi(0, \sigma^2(\tilde{x}))$ – имеет асимптотически нормальное

распределение с нулевым средним и дисперсией

$$\sigma^2(\tilde{x}) = f(\tilde{x}) \int k^2(\tilde{x}) d\tilde{x},$$

где $\int k^2(\tilde{x}) d\tilde{x} < \infty$.

В качестве ядер $k(t)$ можно использовать следующие [55]:

1) нормальное $k(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$;

2) Лапласа $k(t) = \frac{1}{2} e^{-|x|}$;

3) Фишера $k(t) = \frac{1}{2\pi} \left(\frac{\sin \frac{x}{2}}{\frac{x}{2}} \right)$;

4) Коши $k(t) = \frac{1}{\pi} \left(\frac{1}{(1+x)^2} \right)$;

5) логистическое $k(t) = \frac{e^{-x}}{(1+e^{-x})^2}$;

6) Епанечникова $k(t) = \frac{3 \left(1 - \frac{x^2}{5} \right) 1}{4\sqrt{5}}, |x| \leq \sqrt{5}$;

7) равномерное $k(t) = \frac{1}{2}, |x| \leq 1$;

8) треугольное $k(t) = 1 - |x|, |x| \leq 1$.

Смещение и дисперсия оценки $\hat{f}_n(\tilde{x})$ зависят от вида ядра $k(t)$ и параметра размытости h_n .

1.2.7 Адаптивные оценки плотности

Оптимальные параметры размытости зависят от неизвестной плотности и ее производных, т. е. они неизвестны. Более того, нахождение оптимального параметра размытости является более сложной задачей, чем оценивание плотности. В практических приложениях могут интересовать значения оценки плотности не в среднем, а в некоторой области определения ξ , например, в области модального значения случайной величины в задачах распознавания образов, где параметр размытости будет небольшим, или на хвостах распределения в задачах надежности, где параметр размытости будет большим. Эти соображения не добавляют

оптимизма при выборе конкретного алгоритма оценки плотности и приводят к идее построения адаптивных ядерных оценок [65], у которых параметр размытости и ядро настраиваются на основе имеющейся выборки. Существует два метода построения адаптивных оценок плотности. Первый связан с построением оценки оптимального параметра размытости, полученного из теоретических соображений, он рассматривается как функционал от неизвестного распределения. Моделирование показывает, что такой подход дает неплохие результаты при больших объемах выборки ($n > 400$). Вторым методом заключается в следующем. Разобьем выборку \tilde{X}_n на две подвыборки \tilde{X}_{n_1} и \tilde{X}_{n_2} , $n_1 + n_2 = n$. Назовем \tilde{X}_{n_1} – обучающей выборкой, по которой построим оценку плотности, а \tilde{X}_{n_2} – контрольной, по которой оценим некоторый функционал качества оценки. Найдем минимум оценки функционала качества по параметру размытости и ядрам из некоторого набора ядер. При большом объеме выборки (около тысячи и более наблюдений) такой подход вполне оправдан. В реальных задачах при небольшом, как правило, объеме выборки предложен метод перекрестного экзамена. Из выборки убирается наблюдение x_i , которое используется в качестве контрольного. По оставшимся $(n - 1)$ наблюдениям строится оценка. Затем x_i возвращается в выборку, вместо него в качестве контрольного берется следующее наблюдение. Эту процедуру повторяют n раз. В n контрольных наблюдениях производится оценка функционала качества и усреднение по ним. В качестве функционала можно взять информационный функционал:

$$J = \int \ln[l(t)] f(t) dt = \int \ln[l(t)] dF(t),$$

где $l(t)$ – некоторая плотность распределения. Функционал J принимает максимальное значение при $l(t) = f(t)$. Рассмотрим $l(t) = \hat{f}_n(t, h_n, k(t))$ как функцию параметра размытости h_n и ядра $k(t)$. Поиск оптимальных h_n и $k(t)$ сводится к следующей задаче:

$$(h_n^*, k^*(\tilde{x})) = \arg \max_{h_n, k(\tilde{x})} J(h_n, k(\tilde{x})) = \arg \max_{h_n, k(\tilde{x})} \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left[\frac{1}{(n-1)h_n} \sum_{j=1}^{n-1} k \left(\frac{x_i - x_j}{h_n} \right) \right] \right\}.$$

Моделирование адаптивных оценок на основе этого метода, может быть выполнено как для выборок большого объема, так и для выборок небольших объемов.

1.3 Классические методы оценивания параметров регрессионных моделей

Очень часто при решении задач, возникающих в реальной жизни, требуется определить, как зависит наблюдаемая случайная величина от одной или нескольких других величин. Самый общий случай такой зависимости – статистическая зависимость. Для зависимых случайных величин имеет смысл рассмотреть математическое ожидание одной из них при фиксированном значении другой (других). Такое условное математическое ожидание показывает, как влияет на среднее значение первой величины изменение значений второй. Разумеется, наблюдать это среднее значение невозможно, но можно наблюдать значения первой случайной величины при разных значениях остальных. Эту зависимость можно представить, как вход и выход некоторого объекта – «черный ящик» [14, 16]. Входные данные, или «факторы», как правило, известны. На выходе получается результат преобразования входных данных по каким-либо правилам. Задача определения зависимости между входными факторами и выходными называется задачей регрессионного анализа [13, 16, 21].

1.3.1 Задача регрессионного анализа

Решение задачи регрессионного анализа заключается в выборе структуры регрессионной модели, а затем в оценивании ее параметров с помощью какого-либо известного метода оценивания.

Выберем некоторую регрессионную модель, считающуюся истинной моделью исследуемого объекта:

$$y = X\theta + \varepsilon, \quad (1.28)$$

где $X = \begin{bmatrix} \varphi_1(x_{11}) & \dots & \varphi_q(x_{1q}) \\ \dots & \dots & \dots \\ \varphi_1(x_{n1}) & \dots & \varphi_q(x_{nq}) \end{bmatrix}$ – матрица значений действительных функций,

$rg(X) = q$; $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ – оцениваемые неизвестные параметры модели; $\varphi_i(x)$ – известные вещественные функции; x_{ij} – значения входных факторов в n наблюдениях; $y = (y_1, y_2, \dots, y_n)^T$ – вектор значений отклика; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ – вектор случайных ошибок, q – число неизвестных параметров; n – число проведенных экспериментов.

Будем считать, что вектор ошибок наблюдений ε_i состоит из независимых одинаково распределенных случайных величин с функцией плотности $h(t)$, для которых выполняются условия [2, 21]

$$E(\varepsilon_i) = 0, \quad D(\varepsilon_i) = \sigma^2 < \infty. \quad (1.29)$$

Требуется по имеющимся значениям отклика и входных факторов выполнить наиболее точное оценивание вектора неизвестных параметров регрессионного уравнения (1.28).

1.3.2 Метод максимального правдоподобия

Рассмотрим задачу оценивания параметров регрессионного уравнения (1.22). Оценки неизвестных параметров можно найти с помощью метода максимального правдоподобия (ММП) [2, 21, 51, 52], предложенного Р. Фишером в 1937 году [66]. Он предписывает выбирать неизвестные параметры так, чтобы максимизировать функцию правдоподобия случайного вектора. Поскольку вектор случайной ошибки независимый, следовательно, функция правдоподобия принимает вид

$$L(y, \theta) = \prod_{i=1}^n h(y_i, \theta), \text{ где } h(y_i, \theta) \text{ – известная плотность распределения } y_i, i = 1 \dots n, \text{ и}$$

при заданном y зависит только от θ . Построение функции правдоподобия для y

возможно только в случае, если известно распределение случайной ошибки ε .
Логарифмическая функция правдоподобия определяется соотношением

$$\ln L(y, \theta) = \ln \left(\prod_{i=1}^n h(y_i, \theta) \right) = \sum_{i=1}^n \ln(h(y_i, \theta)). \quad (1.30)$$

Метод максимального правдоподобия [2, 22, 51, 52] позволяет находить оценки $\hat{\theta}$ неизвестных параметров $\theta = (\theta_1, \dots, \theta_q)^T$, максимизируя по этим параметрам логарифмическую функцию правдоподобия (1.30).

Оценка $\hat{\theta}$, полученная методом максимального правдоподобия, обладает многими положительными свойствами. Она асимптотически несмещена (при $n \rightarrow \infty$) среди всех оценок (линейных и нелинейных) и имеет наименьшую ковариационную матрицу, которая известна как нижняя граница Рао – Крамера [2]. Для оценок $\hat{\theta}$, полученных по ММП, выполняется следующее неравенство: $L(y, \hat{\theta}) \geq L(y, \tilde{\theta})$, $\tilde{\theta}$ - любая другая оценка [2].

Предположим, что вектор ошибок ε состоит из независимых случайных величин, распределенных по экспоненциальному закону $\varepsilon \sim e^\lambda$ [51, 56], где $\lambda > 0$. Тогда плотность распределения

$$h(t) = \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n (y_i - X_i \theta)}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

где X_i – i -я строка матрицы X . Логарифмическая функция правдоподобия с неизвестными параметрами λ и θ принимает вид

$$\ln(L(y, \theta)) = n \ln \lambda - \lambda \sum_{i=1}^n (y_i - X_i \theta).$$

Оценки λ и θ определяются, как решение следующей системы уравнений

$$\frac{\partial \ln(L(y, \theta))}{\partial \theta_i} = 0, \quad i = 1, \dots, q,$$

где $\hat{\sigma}^2 = \lambda^{-2}$ – оценка дисперсии [52].

Если предположить, что плотность независимых случайных величин ε_i имеет вид

$$h(t) = \frac{1}{2\sigma} e^{-\frac{|t|}{\sigma}},$$

т. е. если ε_i имеют распределение Лапласа, то при любом фиксированном σ^2 максимум функции правдоподобия достигается при наименьшем значении суммы $\sum_{i=1}^n |y_i - X_i\theta|$ абсолютных отклонений. Этот метод был разработан в 60-х годах XX века [47] и получил название метода наименьших модулей.

1.3.3 Метод наименьших квадратов

Предположим, что вектор ошибок ε регрессионного уравнения (1.28) состоит из независимых случайных величин с нормальным распределением $N(0, \sigma^2)$. Функция правдоподобия принимает вид

$$L(y, \theta) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - X_i\theta)^2}.$$

Очевидно, что при любом фиксированном σ^2 максимум функции правдоподобия достигается при наименьшем значении суммы квадратов ошибок $\sum_{i=1}^n (y_i - X_i\theta)^2 = \sum_{i=1}^n \varepsilon_i^2$. Метод максимального правдоподобия в случае, когда вектор ошибок и вектор наблюдаемых откликов состоит из независимых и одинаково распределенных случайных величин с нормальным распределением, называется методом наименьших квадратов (МНК). Этот метод был разработан К. Гауссом в 1794 – 1795 годах [47].

Применение МНК возможно при выполнении предположения (1.29) и заключается в решении задачи минимизации остаточной суммы квадратов [16, 25, 52, 66]:

$$\min_{\theta} S(\theta),$$

где

$$S(\theta) = (y - X\theta)^T (y - X\theta) = y^T y - 2\theta^T X^T y + \theta^T X^T X \theta. \quad (1.31)$$

В дальнейшем дифференцируя $S(\theta)$ по θ получим

$$\frac{\partial S(\theta)}{\partial \theta} = -2X^T y + 2X^T X \theta.$$

Решение системы уравнений $\frac{\partial S(\theta)}{\partial \theta} = 0$ позволяет найти точку минимума

$$X^T X \theta = X^T y,$$

$$\hat{\theta} = (X^T X)^{-1} X^T y.$$

Оценка максимального правдоподобия для σ^2 , она же точка максимума по σ^2 функции правдоподобия, равна

$$\hat{\sigma}^2 = \frac{1}{n - q} (y - X\hat{\theta})^T (y - X\hat{\theta}).$$

Оценки неизвестных параметров, полученные с помощью МНК, являются несмещенными, состоятельными и эффективными (теорема Гаусса – Маркова) [51]. Отметим, что существование аналитического решения и наличие таких оптимальных свойств является основными преимуществами МНК [21, 52]. Однако серьезным недостатком метода является сильная чувствительность оценок к грубым ошибкам наблюдения (выбросам).

1.4 Устойчивые методы оценивания параметров регрессионных моделей

Классические методы оценивания, представленные выше, являются неустойчивыми к появлению грубых ошибок наблюдений – выбросов. Поэтому особый интерес вызывают устойчивые методы оценивания параметров, в частности оценка Хьюбера [53, 72], знаковый метод [8, 17], ранговый метод [68], метод наименьшей медианы квадратов [53, 64] и другие. Эти методы оценивания оказываются более устойчивыми к появлению грубых ошибок наблюдений, что подтверждается в различных исследованиях [8, 17, 53].

Для определения устойчивых оценок параметров регрессионной модели (1.28) можно воспользоваться М-оценками вектора неизвестных параметров θ , минимизируя функцию

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho(y_i - X_i \hat{\theta}), \quad (1.32)$$

где $\rho(y_i - X_i \hat{\theta})$ – функция потерь. Выбор вида функции потерь определяет вид оценок. В случае $\rho(y_i - X_i \hat{\theta}) = (y_i - X_i \hat{\theta})^2$ получим МНК-оценки, а при $\rho(y_i - X_i \hat{\theta}) = |y_i - X_i \hat{\theta}|$ – оценки метода наименьших модулей.

Необходимые условия минимума функции (1.32) следуют из равенства нулю частных производных по параметрам θ_j :

$$\sum_{i=1}^n \gamma(y_i - X_i \hat{\theta}) \varphi_j(x_{ij}) = 0, \quad j = 1, \dots, q,$$

где $\gamma(y_i - X_i \hat{\theta}) = \rho'(y_i - X_i \hat{\theta})$. Чтобы сделать М-оценку независимой от коэффициента масштаба распределения, можно перейти к задаче

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho\left(\frac{y_i - X_i \hat{\theta}}{\lambda}\right), \quad (1.33)$$

где $\rho\left(\frac{y_i - X_i \hat{\theta}}{\lambda}\right)$ – функция потерь, $\lambda > 0$ – параметр масштаба.

Необходимые условия минимума функции (1.33) следуют из равенства нулю частных производных по параметрам θ_j :

$$\sum_{i=1}^n \gamma\left(\frac{y_i - X_i \hat{\theta}}{\lambda}\right) \varphi_j(x_{ij}) = 0, \quad j = 1, \dots, q,$$

где $\gamma\left(\frac{y_i - X_i \hat{\theta}}{\lambda}\right) = \rho'\left(\frac{y_i - X_i \hat{\theta}}{\lambda}\right)$. Параметр масштаба λ можно находить вместе с θ ,

а можно считать известным, используя какую-либо оценку. В качестве устойчивой оценки λ для нормального распределения используют медиану абсолютных

отклонений (MAD-оценку) [18]: $\hat{\lambda} = \frac{\text{med}|y_i - X_i \hat{\theta}|}{0.67449}$.

1.5 Адаптивные методы оценивания параметров регрессионных моделей

Адаптивное оцениванию параметров регрессионных моделей можно выполнять построением квазиправдоподобных оценок при использовании некоторого параметризованного по форме семейства распределений [61-63, 84].

1.5.1 Оценки, минимизирующие вектор остатков в L_ν -метрике

Согласно [18], оценка вектора неизвестных параметров может быть найдена путем минимизации функции

$$\sum_{i=1}^n |y_i - X_i \hat{\theta}|^\nu, \quad (1.34)$$

которая задает М-оценку с функцией $\rho(t) = |t|^\nu$. При $\nu \geq 1$ у функции (1.34) существует единственный локальный минимум, причем он совпадает с глобальным. В случае $0 < \nu < 1$ функция (1.34) может иметь несколько локальных минимумов. Оценки, полученные таким образом, минимизируют по θ (1.34) в L_ν -метрике и называются L_ν -оценками. Отметим, что при $\nu = 1$ полученные оценки совпадают с оценками по методу наименьших модулей; при $\nu = 2$ – с МНК-оценками. L_ν -оценки не зависят от параметра масштаба [18], а значит, оценивание параметров можно проводить независимо от оценивания параметра масштаба.

1.5.2 Метод адаптивного оценивания параметров на основе обобщенного лямбда-распределения

Один из алгоритмов адаптивного оценивания параметров регрессии на основе GL-распределения предложен в работе [19]. Принимая во внимание, что ошибки наблюдений являются независимыми, логарифмическую функцию правдоподобия можно представить в виде

$$\ln(L(y, \theta)) = \sum_{i=1}^n \ln(h(t, \theta)), \quad (1.35)$$

где функция плотности GL-распределения определяется соотношением (1.28). Тогда алгоритм адаптивного оценивания параметров регрессии на основе GL-распределения описывается следующей пошаговой процедурой [19, 57]:

Шаг 1. Определить начальное значение вектора неизвестных параметров $\hat{\theta}^0 = (\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_q^0)^T$ уравнения (1.35); $l=0$, где l – номер итерации.

Шаг 2. Вычислить значения остатков $e_i = y_i - X_i \hat{\theta}^l$ регрессионной зависимости.

Шаг 3. Выполнить идентификации распределения остатков с использованием GL-распределения, соотношение (1.22).

Шаг 4. Определить значение логарифмической функции правдоподобия $\ln(L(e, \hat{\theta}^l))$.

Шаг 5. Найти значение оценки вектора неизвестных параметров $\hat{\theta}^{l+1} = \arg \max_{\theta} L(e_1, \dots, e_n, \theta^l)$.

Шаг 6. Итерационный процесс завершается, если $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| < \delta$, где δ - заданная погрешность вычисления. Если же $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| > \delta$, то выполняется переход на шаг 2 при условии $l = l + 1$.

1.6 Обзор программных систем оценивания параметров регрессионных уравнений и восстановления функции плотности

Рассмотрим статистические программные пакеты, которые пользуются наибольшей популярностью и имеют развитый функционал.

Пакет программ распространяется частной компанией Minitab Inc. [76] и предназначен для обработки статистических данных. В программах из этого пакета можно визуализировать, анализировать, сравнивать данные для реализации бизнес-задач. Minitab имеет удобный интерфейс, позволяющий легко составлять отчеты, проводить анализ, форматировать таблицы и графики. Пакет Minitab включает

следующий набор статистических инструментов: управление процессом статистической обработки данных; оценка мощности и объема выборки; планирование экспериментов; матричные функции; анализ измерительных систем, анализ надежности / выживаемости, анализ временных рядов и прогнозирование, многомерный анализ. В случае выполнения регрессионного анализа в комплектации статистического аппарата пакета Minitab имеется возможность для решения задач множественной линейной и нелинейной регрессии, где оценки неизвестных параметров вычисляются по МНК. Восстановление функции плотности распределения выполняется с помощью гистограммного метода оценивания или метода максимального правдоподобия. Встроенных процедур для построения оценок функции плотности на основе вейвлет-анализа и преобразования Фурье пакет Minitab не имеет.

Пакет STADIA (Statistical Dialogue System) [7, 37] – универсальный статистический пакет, разработанный специалистами Московского государственного университета им. М.В. Ломоносова совместно с НПО «Информатика и компьютеры». Пакет ориентирован на массового пользователя и предоставляет широкий набор методов статистического анализа данных: описательная статистика, дисперсионный, корреляционный и спектральный анализ, сглаживание, прогнозирование, линейная и нелинейная регрессия, кластерный и факторный анализ, методы контроля качества, анализ и замена пропущенных значений. Отметим, что при решении задач множественной линейной и нелинейной регрессии в STADIA оценки неизвестных параметров вычисляются по МНК. В пакете STADIA восстановление функции плотности распределения выполняется с помощью гистограммного метода оценивания или метода максимального правдоподобия. Также возможно построение и редактирование двух, трёх и многомерной графики: зависимости, прогнозы, диаграммы рассеяния, карты, гистограммы, столбиковые, башенные и круговые диаграммы, установка размеров, надписей по осям и под рисунком. Отметим, что методы спектрального анализа, представленные в данном пакете, предназначены для цифровой обработки сигналов.

Пакет STATISTICA [73] разработан компанией StatSoft и предназначен для статистического и графического анализа данных, содержит инструменты для разработки пользовательских приложений в различных областях. Выбор доступных аналитических инструментов выполняется с помощью альтернативного пользовательского интерфейса. Эффективность работы как с таблицами данных на локальном диске, так и с удаленными хранилищами данных обеспечивается гибкой технологией доступа к данным. Пакет предоставляет пользователям следующие возможности статистического анализа данных: исследование корреляций между переменными; диаграмма рассеяния, матричная диаграмма рассеяния; быстрые основные статистики и блоковые статистики (интерактивные средства, позволяющие одним щелчком мыши вычислять основные статистики и строить графики в любой момент в течение сеанса работы); интерактивный калькулятор вероятностных распределений (позволяет интерактивно исследовать структуру распределений, например зависимость вероятности от параметров); анализ многомерных откликов, многомерное шкалирование; анализ при помощи временных рядов и прогнозирование временных зависимостей, в том числе анализ сезонных колебаний. Оценивание неизвестных параметров регрессионных зависимостей, в этом пакете выполняется, как правило, МНК или ММП на основе нормального распределения. Оценивание функции плотности распределения выполняется с помощью гистограммного подхода или метода максимального правдоподобия. В пакете STATISTICA нет готовых решений по восстановлению функции плотности на основе различных вейвлетов и преобразования Фурье.

Разработанное компанией SPSS Inc программное обеспечение SPSS (Statistical Package for the Social Sciences) [26] предназначено для сложного статистического анализа, планирования и создания бизнес-отчетности. Все продукты IBM SPSS интегрированы в целостную систему, что позволяет без труда переходить из одной программы в другую. Пакет позволяет осуществлять управление данными и содержит наиболее распространенные методы статистического анализа данных: проведение описательной статистики; построение линейных и нелинейных моделей; осуществление преобразования

данных; проведение факторного, кластерного и дисперсионного анализа; вычисление корреляций. В целом отметим: в комплектации статистического аппарата в пакете SPSS полностью отсутствуют непараметрические методы регрессионного анализа, имеется возможность для решения задач множественной линейной и нелинейной регрессии, где оценки неизвестных параметров вычисляются МНК. Построение оценки функции плотности распределения выполняется на основе построения гистограмм или метода максимального правдоподобия. Пакет SPSS не имеет встроенных процедур для оценивания функции плотности с помощью различных вейвлетов и преобразования Фурье.

Пакет SAS (Statistical Analysis System) [99] – профессиональный статистический пакет, разработанный компанией SAS Institute Inc. Первая версия пакета была выпущена в 1976 году. Последняя версия SAS Statistics позиционируется как статистический пакет для аналитической работы на современных предприятиях. Пакет эффективно работает с данными различных типов, а встроенные инструменты статистического анализа могут быть применены для решения широкого круга задач. Использование пакета SAS позволяет создавать, редактировать данные, управлять ими, выполнять математический и статистический анализ, графическое оформление, генерирование отчетов. Рассмотрим три направления, где это программное обеспечение поможет в достижении эффективных решений. Пакет предоставляет пользователям следующие возможности статистического анализа данных: смешанные модели; анализ категориальных данных; байесовский анализ; многомерный анализ; анализ выживаемости; психометрический анализ; анализ пропущенных значений. Оценивание неизвестных параметров регрессионных зависимостей в данном пакете выполняется с помощью МНК. Восстановление функции плотности распределения выполняется на основе гистограммного подхода или метода максимального правдоподобия. Имеющиеся встроенные инструменты спектрального анализа предназначены для цифровой обработки сигналов.

Описанные выше универсальные пакеты, такие как SPSS, STATISTICA, Stadia, Minitab, не ориентированы на специфическую предметную область и могут

применяться для анализа данных из различных областей деятельности. Многопрофильность универсальных пакетов позволяет провести лишь пробный анализ различных типов данных с использованием широкого диапазона статистических методов. Пакет SAS отличается от универсальных тем, что позволяет работать со сверхбольшими объемами данных, применять узкоспециализированные методы анализа, создавать собственную систему обработки данных, но достаточно сложен в освоении для непрофессионалов. К недостатку всех представленных пакетов можно отнести небольшое число методов оценивания неизвестных параметров регрессионных зависимостей, которые ограничиваются, как правило, МНК и ММП на основе нормального распределения. Как было отмечено, эти методы дают хороший результат только в случае нормального распределения ошибок, что на практике наблюдается достаточно редко. В некоторых областях анализируемые данные настолько специфичны, что к ним следует применять особые методы оценивания параметров, не представленные в данных пакетах. Отметим, что указанные выше пакеты не имеют функционала для быстрого и простого использования преобразования Фурье и вейвлет-анализа для оценивания функции плотности и восстановления неизвестных параметров регрессионных зависимостей.

1.7 Обоснование цели и задачи исследования

Анализ основных методов и подходов для решения задач, связанных с оцениванием неизвестных параметров регрессионных моделей, показал, что в большинстве своем существующие методы позволяют получать достаточно корректные и качественные результаты только при условии, что имеются достоверные предположения о свойствах случайной компоненты. Одним из таких способов определения неизвестных оценок является метод максимального правдоподобия. Применение этого метода возможно при условии, что имеется достоверная информация о виде распределения случайных ошибок наблюдения. Предположение о нормальности распределения ошибок наблюдения позволяет

применить метод наименьших квадратов и тем самым упростить поиск оценок. Отмечено, что метод наименьших квадратов теряет устойчивость при наличии в выборке выбросов. В этом случае целесообразно использовать устойчивые методы оценивания. Однако их реализация может быть затруднена либо решением сложных оптимизационных задач, либо выбором вида функции потерь. Использование квазиравдоподобных методов допускается при условии, что существует возможность оценить функцию плотности распределения ошибок по имеющимся данным. Оценка функции плотности распределения может быть получена как с помощью параметрических методов оценивания, так и не параметрических. Параметрические методы требуют задания функции плотности. Среди непараметрических методов наибольшую известность получила ядерная оценка Розенבלата – Парзена. Однако ее качество сильно зависит от значения параметра сглаживания, неоптимальный выбор которого может приводить к резким флуктуациям графика восстанавливаемой функции плотности, что особенно критично на малых выборках и делает невозможным её дальнейшее использование при решении задач, связанных с оцениванием неизвестных параметров регрессионных моделей. Это обстоятельство заставляет исследователей быть осторожными при использовании ядерных оценок.

Кроме того, обзор существующих наиболее популярных программных средств показал, что в их функционале представлены только классические методы для оценивания параметров регрессии. На сегодняшний день разработчики пытаются расширить такой набор методов для оценивания неизвестных параметров. Однако эти попытки приводят к целому ряду затруднений, связанных с тем, что на практике в большинстве случаев данные достаточно специфичны, могут иметь самые разнообразные распределения и содержать большую долю выбросов. Комплексного решения для оценивания неизвестных параметров регрессионных моделей при работе с такими данными пока не предложено. Для восстановления функции плотности распределения в большинстве известных пакетов используют гистограммный подход или метод максимального

правдоподобия. Встроенные инструменты для работы с преобразованием Фурье и вейвлет-анализом либо отсутствуют, либо ориентированы на спектральный анализ.

В связи с вышеизложенным возникает предположение о необходимости разработки и применения новых подходов для оценивания неизвестных параметров регрессионных зависимостей. Целью данного исследования является разработка математического и алгоритмического обеспечения для восстановления функции плотности с использованием вейвлет-анализа и преобразования Фурье и адаптивного оценивания параметров линейно-параметризованных регрессионных моделей. Для достижения этой цели требуется решить целый ряд задач.

Во-первых, для восстановления функции плотности распределения необходимо рассмотреть применение преобразования Фурье и вейвлет-анализа, широко используемых при аппроксимации различных функций. Это даст возможность в дальнейшем оценить распределение случайной ошибки в случае, когда его нельзя считать нормальным. Использование различных материнских вейвлетов позволит получить различные алгоритмы восстановления функции плотности и использовать их в рамках метода максимального правдоподобия.

Во-вторых, необходимо вычислить нормы для ортогональных материнских вейвлетов LITTLEWOOD & PALEY, Морле и неортогональных вейвлетов DOG и «Мексиканская шляпа». Построить на основе преобразования Фурье и вейвлет-анализа нормированные системы базисных функций и использовать их для оценивания функции плотности распределения. Кроме того, необходимо провести исследование предложенных оценок функции плотности, а также сформулировать рекомендации относительно их использования.

В-третьих, необходимо построить ряд новых адаптивных алгоритмов оценивания неизвестных параметров регрессионных зависимостей, основанных на использовании преобразования Фурье и вейвлет-анализа. Также необходимо провести исследование предложенных алгоритмов оценивания параметров регрессионной зависимости с помощью методов статистического моделирования. Кроме того, проанализировать точность оценивания неизвестных параметров при различной степени отклонения ошибки от нормального распределения.

В-четвертых, как показал анализ имеющихся программных средств, реализованные в них методы оценивания неизвестных параметров регрессионных зависимостей ограничиваются, как правило, МНК и ММП на основе нормального распределения, поэтому разработка собственного программного комплекса, реализующего предложенные алгоритмы, является актуальной. Функционал такого программного комплекса должен позволять моделировать случайные величины, распределенные по нормальному, экспоненциальному и другим законам распределения, выполнять оценивание функции плотности распределения с помощью преобразования Фурье или вейвлет-преобразования, а также оценивать параметры регрессионной модели с помощью МНК и адаптивных алгоритмов на основе преобразования Фурье или вейвлет-преобразования.

Поставленные выше задачи рассматривались автором исследования в качестве основных. Их решение предложено в настоящей диссертационной работе.

Глава 2 Построение и исследование оценок функции плотности распределения с помощью преобразования Фурье и вейвлет-анализа

В данной главе рассматривается задача, суть которой состоит в построении разложения неизвестной функции в ряд по некоторому конечному набору ортонормированных базисных функций. Сформулированы и доказаны утверждения, определяющие нормы материнских вейвлетов. Определены оценки функции плотности распределения, где система базисных функций конструируется с помощью тригонометрического базиса или на основе различных материнских вейвлетов. С помощью вычислительных экспериментов установлено влияние на качество восстановления функции плотности распределения таких параметров, как объем выборки, число членов ряда, значение нормировочного коэффициента. Выполнен сравнительный анализ вейвлет-оценок плотности распределения и оценок на основе преобразования Фурье.

2.1 Оценка функции плотности распределения на основе различных ортонормированных базисов

Сформулируем следующую задачу: пусть задана выборка $\{x_j | j = \overline{1, n}\}$ независимых значений непрерывной случайной величины ξ на произвольном отрезке $[c, d]$, где $\begin{cases} c = \min x_j \\ d = \max x_j \end{cases}$. Закон, а также функция плотности распределения $f(t)$ случайной величины ξ являются неизвестными. Требуется построить оценку $\hat{f}_n(t)$ плотности распределения $f(t)$ с использованием выборочных данных.

Согласно [79, 80, 94, 100, 104] оценка функции плотности $\hat{f}_n(t)$ случайной величины выражается разложением в ряд по некоторым ортонормированным на отрезке $[a, b]$ базисным функциям $\psi_i(t)$

$$\hat{f}_n(t) = \sum_{i=1}^N \hat{c}_i \psi_i(t), \quad (2.1)$$

где N – число членов ряда (параметр сглаживания), \hat{c}_i – оценки коэффициентов разложения по данному базису, которые определяются по имеющимся статистическим данным

$$\hat{c}_i = \frac{1}{n} \sum_{j=1}^n \psi_i(x_j). \quad (2.2)$$

С учетом (2.2) оценку функции плотности можно записать в следующем виде:

$$\hat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n W_N(t, x_j), \quad (2.3)$$

где $W_N(t, x_j) = \sum_{i=1}^N \psi_i(x_j) \psi_i(t)$.

Отметим, что рассмотренная выше ортонормированная система функций $\psi_i(t)$ задана на отрезке $[a, b]$, который может не совпадать с фактической областью значений ξ . Поэтому для вычисления $\hat{f}_n(t)$ требуется перейти к ортонормированной системе функций $\tilde{\psi}_i(t)$, определенной на том же отрезке $[c, d]$, что и значения случайной величины. Чтобы выполнить такой переход, воспользуемся соотношением [4, 70]

$$\tilde{\psi}_i(t) = \lambda_1 \psi_i(\lambda_2 t + \lambda_3), \quad (2.4)$$

где $\lambda_1, \lambda_2, \lambda_3 \in R$. Из условий (1.18) и (1.19) и выражения (2.4) построим систему уравнений для нахождения коэффициентов $\lambda_1, \lambda_2, \lambda_3$:

$$\begin{cases} \lambda_2 c + \lambda_3 = a \\ \lambda_2 d + \lambda_3 = b \\ \lambda_1 = \sqrt{\lambda_2} \end{cases} \quad (2.5)$$

Решим систему (2.5) и получим

$$\lambda_1 = \sqrt{\frac{b-a}{d-c}}, \quad \lambda_2 = \frac{b-a}{d-c}, \quad \lambda_3 = \frac{ad-bc}{d-c}. \quad (2.6)$$

Подставляя (2.6) в (2.4), получим преобразование базисных функций при переходе от одной системы функций к другой:

$$\tilde{\Psi}_i(t) = \sqrt{\frac{b-a}{d-c}} \Psi_i\left(\frac{(b-a)t + ad - bc}{d-c}\right). \quad (2.7)$$

Итоговая оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ выражается разложением (2.1), где вместо базисных функций $\Psi_i(t)$ используются $\tilde{\Psi}_i(t)$.

2.2 Оценивание плотности распределения с помощью преобразования Фурье

Для оценки функции плотности распределения с помощью преобразования Фурье используют тригонометрическую систему функций [20, 45, 70]:

$$\{1, \sin x, \cos 2x, \sin 2x, \dots\}. \quad (2.8)$$

Система функций (2.8) ортогональна на отрезке $[-\pi, \pi]$, но не является ортонормированной [20, 34, 70]. В результате перехода от системы (2.8) к ортонормированной системе тригонометрических функций, определенных на отрезке $[-\pi, \pi]$, получаем [20, 34, 70]

$$\Psi_0 = \frac{1}{\sqrt{2\pi}}; \Psi_{2k-1} = \frac{\cos kx}{\sqrt{\pi}}; \Psi_{2k} = \frac{\sin kx}{\sqrt{\pi}}, \text{ где } k \geq 1. \quad (2.9)$$

Отметим, что рассмотренная ортонормированная система функций (2.9) задана на отрезке $[-\pi, \pi]$, который может не совпадать с фактической областью значений. Поэтому для вычисления $\hat{f}_n(t)$ требуется перейти к ортонормированной системе функций $\tilde{\Psi}_i(t)$, определенной на том же отрезке $[c, d]$, что и значения случайной величины. Переход от системы функций (2.9) к ортонормированной системе $\tilde{\Psi}_i(t)$ на произвольном отрезке с учетом (2.7) дает результат

$$\tilde{\Psi}_i(t) = \begin{cases} \frac{1}{\sqrt{d-c}}, & i=0, \\ \sqrt{\frac{2}{d-c}} \sin\left(j\pi\left(\frac{2t}{d-c} - \frac{d+c}{d-c}\right)\right), & i=2j-1, \\ \sqrt{\frac{2}{d-c}} \cos\left(j\pi\left(\frac{2t}{d-c} - \frac{d+c}{d-c}\right)\right), & i=2j. \end{cases} \quad (2.10)$$

Итоговая оценка функции плотности случайной величины, построенная с помощью дискретного преобразования Фурье (ДПФ) на произвольном отрезке $[c, d]$ выражается разложением (2.1) по ортонормированным базисным функциям (2.10).

Оценка функции плотности распределения, построенная с помощью быстрого преобразования Фурье (БПФ), выражается разложением (2.1) по ортонормированным базисным функциям (2.10), где коэффициенты \hat{c}_i имеют вид [20, 34, 70]

$$\hat{c}_i = \frac{2}{n} \sum_{j=1}^{\frac{n-1}{2}} \Psi_i(x_j). \quad (2.11)$$

Выражение (2.11) следует из периодичности коэффициентов (2.2) и свойства симметрии дискретного преобразования.

2.3 Оценивание функции плотности распределения с помощью вейвлетов

Рассмотрим в качестве ортонормированной системы $\Psi_i(t)$, согласно [67, 70], набор ортонормированных на $[0, 1]$ базисных функций:

$$\Psi_i(t) = 2^{\frac{k}{2}} \Psi(2^k t - (j-1)), \quad (2.12)$$

где $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$, $\Psi(t)$ – материнский вейвлет или вейвлет-функция. Выражение (2.12) справедливо для $i > 1$, в случае $i = 1$ функция $\Psi_1(t) = 1$ для всех $t \in [0, 1]$ и $\Psi_1(t) = 0$ в противном случае. Согласно [20] соотношение (2.1) можно использовать, если система функций (2.12) не ортогональна.

Ортонормированная система функций (2.12) задана на отрезке $[0,1]$, а случайная величина ξ – на отрезке $[c,d]$, поэтому, учитывая соотношение (2.7), получаем базисные функции, определенные на произвольном отрезке:

$$\tilde{\Psi}_i(t) = \frac{1}{\sqrt{d-c}} \Psi_i\left(\frac{t-c}{d-c}\right). \quad (2.13)$$

Оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c,d]$, выраженная разложением (2.1) по базисным функциям (2.13), называется вейвлет-оценкой функции плотности распределения.

2.3.1 Оценивание функции плотности распределения с помощью ортогональных вейвлетов

Рассмотрим ортонормированный вейвлет Хаара [20, 81, 91, 95, 101], определяемый соотношением (1.9), для которого ортонормированная система функций (2.12) на отрезке $[0,1]$ (система Хаара) принимает вид

$$\Psi_i(t) = \begin{cases} 2^{\frac{k}{2}}, & t \in [2^{-k}(j-1), 2^{-(k+1)}(2j-1)), \\ -2^{\frac{k}{2}}, & t \in [2^{-(k+1)}(2j-1), 2^{-k}j), \\ 0, & t \notin (2^{-k}(j-1), 2^{-k}j], \end{cases} \quad (2.14)$$

где i, k, j такие же, как в (2.12).

Переход от системы функций (2.14) к ортонормированной системе $\tilde{\Psi}_i(t)$ на произвольном отрезке $[c,d]$, с учетом (2.7) дает результат:

$$\tilde{\Psi}_i(t) = \frac{1}{\sqrt{d-c}} \begin{cases} 2^{\frac{k}{2}}, & t \in [2^{-k}(j-1)(d-c)+c, 2^{-(k+1)}(2j-1)(d-c)+c), \\ -2^{\frac{k}{2}}, & t \in [2^{-(k+1)}(2j-1)(d-c)+c, 2^{-k}j(d-c)+c), \\ 0, & t \notin (2^{-k}(j-1)(d-c)+c, 2^{-k}j(d-c)+c], \end{cases} \quad (2.15)$$

где i, k, j такие же, как в (2.12).

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием материнского вейвлета Хаара выражается разложением (2.1) по ортонормированным базисным функциям (2.15).

Аналитическая запись LITTLEWOOD & PALEY [10, 11, 12, 20, 69, 70] представлена в разделе 1.1.2 соотношением (1.15), тогда система функций (2.12) на отрезке $[0, 1]$ принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} \left(\frac{\sin 2\pi(2^k t - (j-1)) - \sin \pi(2^k t - (j-1))}{\pi(2^k t - (j-1))} \right), \quad (2.16)$$

где i, k, j такие же, как в (2.12). Проверим, нормирована ли система функции (2.16), для этого было получено значение нормы вейвлета LITTLEWOOD & PALEY.

Утверждение 2.1. Пусть $\psi(t) = \frac{\sin 2\pi t - \sin \pi t}{\pi t}$ – материнский вейвлет LITTLEWOOD & PALEY, порождающий систему функций $\psi_i(t) \in L^2(\mathbb{R})$, которая принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} \left(\frac{\sin 2\pi(2^k t - (j-1)) - \sin \pi(2^k t - (j-1))}{\pi(2^k t - (j-1))} \right),$$

где $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j $\|\psi_i\| = 1$.

Доказательство. Аналитическая запись материнского вейвлета LITTLEWOOD & PALEY выражается соотношением

$$\psi(t) = \frac{\sin 2\pi t - \sin \pi t}{\pi t}.$$

Тогда по определению нормы в $L^2(\mathbb{R})$ получаем

$$\|\psi(t)\|^2 = \int_{-\infty}^{+\infty} \frac{(\sin 2\pi t - \sin \pi t)^2}{(\pi t)^2} dt.$$

Представим полученный интеграл в виде суммы трех интегралов:

$$\|\psi(t)\|^2 = \int_{-\infty}^{+\infty} \frac{\sin^2 2\pi t}{\pi^2 t^2} dt - \int_{-\infty}^{+\infty} \frac{2 \sin 2\pi t \cdot \sin \pi t}{\pi^2 t^2} dt + \int_{-\infty}^{+\infty} \frac{\sin^2 \pi t}{\pi^2 t^2} dt.$$

Вычислим каждый интеграл по отдельности:

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{\sin^2 2\pi t}{\pi^2 t^2} dt &= \int_{-\infty}^{+\infty} \frac{1 - \cos 4\pi t}{2\pi^2 t^2} dt = 2 \int_1^{+\infty} \frac{1}{2\pi^2 t^2} dt - 2 \int_1^{+\infty} \frac{\cos 4\pi t}{2\pi^2 t^2} dt = \\ &= 2 \lim_{b \rightarrow +\infty} \left(-\frac{1}{2\pi^2 b} + \frac{1}{2\pi^2} \right) + 2 \lim_{b \rightarrow +\infty} \left(-\frac{\cos(4\pi b)}{2\pi^2 b} - \frac{1}{2\pi^2} \right) + \\ &= \frac{2}{\pi} \int_{-\infty}^{+\infty} \frac{\sin(4\pi t)}{4\pi t} d(4\pi t) = \frac{2\pi}{\pi} = 2; \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{2 \sin 2\pi t \cdot \sin \pi t}{\pi^2 t^2} dt &= 2 \int_1^{+\infty} \frac{\cos \pi t}{\pi^2 t^2} dt - 2 \int_1^{+\infty} \frac{\cos \pi t}{\pi^2 t^2} dt = \\ &= -2 \lim_{b \rightarrow +\infty} \left(\frac{\cos \pi b}{\pi^2 b} - \frac{1}{\pi^2} \right) - \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin \pi t}{\pi t} d(\pi t) + \\ &+ 2 \lim_{b \rightarrow +\infty} \left(\frac{\cos \pi b}{\pi^2 b} - \frac{1}{\pi^2} \right) + \frac{3}{\pi} \int_{-\infty}^{+\infty} \frac{\sin 3\pi t}{3\pi t} d(3\pi t) = -\frac{\pi}{\pi} + \frac{3\pi}{\pi} = 2; \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{\sin^2 \pi t}{\pi^2 t^2} dt &= \int_{-\infty}^{+\infty} \frac{1 - \cos 2\pi t}{2\pi^2 t^2} dt = 2 \int_1^{+\infty} \frac{1}{2\pi^2 t^2} dt - 2 \int_1^{+\infty} \frac{\cos 2\pi t}{2\pi^2 t^2} dt = \\ &= 2 \lim_{b \rightarrow +\infty} \left(-\frac{1}{2\pi^2 b} + \frac{1}{2\pi^2} \right) + 2 \lim_{b \rightarrow +\infty} \left(-\frac{\cos 2\pi b}{2\pi^2 b} - \frac{1}{2\pi^2} \right) + \\ &= \frac{2}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin 2\pi t}{2\pi t} d(2\pi t) = \frac{\pi}{\pi} = 1. \end{aligned}$$

В результате получаем, что $\|\psi(t)\|^2 = 2 - 2 + 1 = 1$.

Поскольку для любых $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$, выполняется равенство (2.12), то норма $\|\psi_i\| = 1$. Утверждение доказано.

Переход от области определения $[0,1]$ к области определения наблюдаемой случайной величины с учетом (2.13) дает результат:

$$\tilde{\Psi}_i(t) = \frac{1}{\sqrt{d-c}} 2^{\frac{k}{2}} \left(\frac{\sin 2\pi(2^k \tau - (j-1)) - \sin \pi(2^k \tau - (j-1))}{\pi(2^k \tau - (j-1))} \right), \quad (2.17)$$

где $\tau = \frac{t-c}{d-c}$, а значения i, k, j такие, как в (2.12).

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием материнского вейвлета LITTLEWOOD & PALEY выражается разложением (2.1) по нормированным базисным функциям (2.17).

Рассмотрим вейвлет Морле [10, 11, 12, 20, 70, 71], определяемый соотношением (1.14), для которого система функций (2.12) на отрезке $[0, 1]$ принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} e^{\frac{-k}{2} \frac{-(2^k t - (j-1))^2}{2}} \cos 5(2^k t - (j-1)), \quad (2.18)$$

где i, k, j такие же, как в (2.12). Система функции (2.18) не нормирована, что показано в следующем утверждении.

Утверждение 2.2. Пусть $\psi(t) = e^{-t^2/2} \cos 5t$ – материнский вейвлет Морле, порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} e^{\frac{-k}{2} \frac{-(2^k t - (j-1))^2}{2}} \cos 5(2^k t - (j-1)),$$

где $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j

$$\|\psi_i\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} (1 + e^{-25})^{1/2}.$$

Доказательство. Аналитическая запись материнского вейвлета Морле выражается соотношением

$$\psi(t) = e^{-t^2/2} \cos 5t.$$

Тогда по определению нормы в $L^2(R)$ получаем

$$\|\psi(t)\|^2 = \int_{-\infty}^{+\infty} \left(e^{-\frac{t^2}{2}} \cos 5t \right)^2 dt = \int_{-\infty}^{+\infty} e^{-t^2} \cos^2 5t dt = \frac{1}{2} \int_{-\infty}^{+\infty} e^{-t^2} (1 + \cos 10t) dt =$$

$$= \frac{1}{2} \int_{-\infty}^{+\infty} e^{-t^2} dt + \frac{1}{2} 2 \int_0^{+\infty} e^{-t^2} \cos 10t dt = \frac{\sqrt{\pi}}{2} + \frac{\sqrt{\pi}}{2} e^{-25} = \frac{\sqrt{\pi}}{2} (1 + e^{-25}).$$

Поскольку для любых $k \geq 0$, $1 \leq j \leq 2^k$, такие что $i = 2^k + j$, выполняется равенство (2.12), то норма $\|\psi_i\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} (1 + e^{-25})^{\frac{1}{2}}$. Утверждение доказано.

Отсюда следует, что базис (2.18) на основе материнского вейвлета Морле не нормирован: $\|\psi_i\| = \frac{\sqrt[4]{\pi}}{\sqrt{2}} (1 + e^{-25})^{\frac{1}{2}} \neq 1$. В соответствии с [20, 69] такая система базисных функций может быть использована для построения (2.1), если ввести нормировочный множитель $z = \frac{1}{\|\psi_i(t)\|}$. Тогда система функций (2.18) становится

ортонормированной и принимает вид

$$\psi_i(t) = z 2^{\frac{k}{2}} e^{\frac{-(2^k t - (j-1))^2}{2}} \cos 5(2^k t - (j-1)), \quad (2.19)$$

где i, k, j аналогичны (2.12), $z = \frac{\sqrt{2}}{\sqrt[4]{\pi}} (1 + e^{-25})^{-\frac{1}{2}}$.

Переход от системы функций (2.19) к нормированной системе $\tilde{\psi}_i(t)$, определенной на отрезке $[c, d]$, который соответствует области определения наблюдаемой случайной величины, с учетом (2.7) дает результат

$$\tilde{\psi}_i(t) = z \frac{1}{\sqrt{d-c}} 2^{\frac{k}{2}} e^{\frac{-(2^k \tau - (j-1))^2}{2}} \cos 5(2^k \tau - (j-1)), \quad (2.20)$$

где i, k, j аналогичны (2.5), $\tau = \frac{t-c}{d-c}$, $z = \frac{\sqrt{2}}{\sqrt[4]{\pi}} (1 + e^{-25})^{-\frac{1}{2}}$.

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием материнского вейвлета Морле выражается разложением (2.1) по нормированным базисным функциям (2.20).

2.3.2 Оценивание функции плотности распределения с помощью неортогональных вейвлетов

Аналитическая запись вейвлета DOG представлена в разделе 1.1.2 [10 – 12, 20, 70] соотношением (1.13), при переходе к системе функций (2.12) на отрезке $[0,1]$ получаем следующее соотношение:

$$\psi_i(t) = 2^{\frac{k}{2}} \left(e^{-\frac{1}{2}(2^k t - (j-1))^2} - \frac{1}{2} e^{-\frac{1}{8}(2^k t - (j-1))^2} \right), \quad (2.21)$$

где i, k, j такие, как в (2.12). Найдем значение нормы полученной системы функций (2.21), для этого докажем следующее утверждение.

Утверждение 2.3. Пусть $\psi(t) = e^{-\frac{t^2}{2}} - \frac{1}{2} e^{-\frac{t^2}{8}}$ – материнский вейвлет DOG, порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} \left(e^{-\frac{1}{2}(2^k t - (j-1))^2} - \frac{1}{2} e^{-\frac{1}{8}(2^k t - (j-1))^2} \right),$$

где $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j

$$\|\psi_i\| = \left(\frac{3}{2} \sqrt{\pi} - \frac{2}{5} \sqrt{10\pi} \right)^{\frac{1}{2}}.$$

Доказательство. Аналитическая запись материнского вейвлета DOG выражается соотношением

$$\psi(t) = e^{-\frac{t^2}{2}} - \frac{1}{2} e^{-\frac{t^2}{8}}.$$

Тогда по определению нормы в $L^2(R)$ получаем

$$\begin{aligned} \|\psi(t)\|^2 &= \int_{-\infty}^{+\infty} \left(e^{-\frac{t^2}{2}} - \frac{1}{2} e^{-\frac{t^2}{8}} \right)^2 dt = 2 \int_0^{+\infty} \left(e^{-t^2} - e^{-\frac{t^2}{2} \cdot \frac{t^2}{8}} + \frac{1}{4} e^{-\frac{t^2}{4}} \right) dt = \\ &= 2 \int_0^{+\infty} e^{-t^2} dt - 2 \int_0^{+\infty} e^{-\frac{5t^2}{8}} dt + \frac{1}{2} \int_0^{+\infty} e^{-\frac{t^2}{4}} dt = \sqrt{\pi} - \sqrt{\frac{8\pi}{5}} + \frac{1}{4} \sqrt{4\pi} = \frac{3}{2} \sqrt{\pi} - \frac{2}{5} \sqrt{10\pi}. \end{aligned}$$

В силу того, что для любых $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$, выполняется равенство (2.12), норма $\|\psi_i\| = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{\frac{1}{2}}$. Утверждение доказано.

Отсюда следует, что базис (2.21) на основе материнского вейвлета DOG не нормирован: $\|\psi_i\| = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{\frac{1}{2}} \neq 1$. В соответствии с [20, 70] такая система базисных функций может быть использована для построения (2.1), если ввести нормировочный множитель $z = \frac{1}{\|\psi_i(t)\|}$. Тогда система функций (2.21) становится

нормированной и принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} z \left(e^{-\frac{1}{2}(2^k t - (j-1))^2} - \frac{1}{2} e^{-\frac{1}{8}(2^k t - (j-1))^2} \right), \quad (2.22)$$

где i, k, j такие, как в (2.12), $z = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{-\frac{1}{2}}$.

Переход от системы функций (2.22) к нормированной системе $\tilde{\psi}_i(t)$ на произвольном интервале $[c, d]$ с учетом (2.7) дает результат

$$\tilde{\psi}_i(t) = \frac{z}{\sqrt{d-c}} 2^{\frac{k}{2}} \left(e^{-\frac{\tau^2}{2}} - \frac{1}{2} e^{-\frac{\tau^2}{8}} \right), \quad (2.23)$$

где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, а значения i, k, j такие же, как в (2.12),

$$z = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{-\frac{1}{2}}.$$

Итоговая вейвлет оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием DOG вейвлета выражается разложением (2.1) по нормированным базисным функциям (2.23).

Материнский вейвлет «Мексиканская шляпа» [10, 21, 69, 70] определяется соотношением (1.12). В этом случае система функций (2.12) на отрезке $[0,1]$ принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} \left(1 - (2^k t - (j-1))^2\right) e^{\frac{-(2^k t - (j-1))^2}{2}}, \quad (2.24)$$

где i, k, j такие же, как в (2.12).

Вычислим значение нормы полученной системы функций (2.24), для этого докажем следующее утверждение.

Утверждение 2.4. Пусть $\psi(t) = (1-t^2)e^{-\frac{t^2}{2}}$ – материнский вейвлет «Мексиканская шляпа», порождающий систему функций $\psi_i(t) \in L^2(\mathbb{R})$, которая принимает вид

$$\psi_i(t) = 2^{\frac{k}{2}} \left(1 - (2^k t - (j-1))^2\right) e^{\frac{-(2^k t - (j-1))^2}{2}},$$

где $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$. Тогда для любых i, k, j $\|\psi_i(t)\| = \frac{\sqrt{3}}{2} \pi^{\frac{1}{4}}$.

Доказательство. Аналитическая запись материнского вейвлета «Мексиканская шляпа» выражается соотношением

$$\psi(t) = (1-t^2)e^{-\frac{t^2}{2}}.$$

Тогда по определению нормы в $L^2(\mathbb{R})$ получаем

$$\begin{aligned} \|\psi(t)\|^2 &= \int_{-\infty}^{+\infty} \left(e^{-\frac{t^2}{2}} (1-t^2) \right)^2 dt = \int_{-\infty}^{+\infty} e^{-t^2} (1-t^2)^2 dt = \int_{-\infty}^{+\infty} e^{-t^2} dt - 2 \int_{-\infty}^{+\infty} e^{-t^2} t^2 dt + 2 \int_0^{+\infty} e^{-t^2} t^4 dt = \\ &= \sqrt{\pi} - 2 \frac{1}{2} \sqrt{\pi} + \frac{2 \cdot 3}{2^3} \sqrt{\pi} = \frac{3}{4} \sqrt{\pi}. \end{aligned}$$

Поскольку для любых $k \geq 0, 1 \leq j \leq 2^k$, такие что $i = 2^k + j$, выполняется равенство (2.12), то норма $\|\psi_i(t)\| = \frac{\sqrt{3}}{2} \pi^{\frac{1}{4}}$. Утверждение доказано.

Отсюда следует, что система функций (2.24) на основе вейвлета «Мексиканская шляпа» не нормирована: $\|\psi_i(t)\| = \frac{\sqrt{3}}{2} \pi^{\frac{1}{4}} \neq 1$. Действуя по аналогии с ранее

рассматриваемым вейвлетом DOG, введем нормировочный множитель $z = \frac{1}{\|\psi_i(t)\|}$.

Тогда с учетом этого множителя система функций (2.24) становится нормированной и принимает вид:

$$\psi_i(t) = 2^{\frac{k}{2}} z (1 - (2^k t - (j-1))^2) e^{-(2^k t - (j-1))^2}, \quad (2.25)$$

где i, k, j такие же, как в (2.12), $z = \frac{2}{\sqrt{3}} \pi^{\frac{1}{4}}$.

Переход от системы функций (2.25) к нормированной системе $\tilde{\psi}_i(t)$ на произвольном интервале $[c, d]$ с учетом (2.7) дает результат

$$\tilde{\psi}_i(t) = \frac{z}{\sqrt{d-c}} 2^{\frac{k}{2}} (1 - \tau^2) e^{-\frac{\tau^2}{2}}, \quad (2.26)$$

где $\tau = \frac{2^k}{d-c} (t-c) - (j-1)$, а значения i, k, j такие же, как в (2.12), $z = \frac{2}{\sqrt{3}} \pi^{\frac{1}{4}}$.

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием материнского вейвлета «Мексиканская шляпа» выражается разложением (2.1) по нормированным базисным функциям (2.26).

Отметим, что материнские вейвлеты DOG и «Мексиканская шляпа» не являются ортогональными [10-12, 20, 70]. Согласно [70] такие вейвлеты также могут быть использованы для построения системы базисных функций (2.12). Однако отсутствие ортогональности системы базисных функций может негативно сказаться на качестве восстановления функции плотности. Для решения этой проблемы предлагается ввести нормировочный множитель \tilde{z} , являющейся более общей поправкой. В этом случае системы функций на произвольном отрезке $[c, d]$ выражаются соотношениями (2.23) и (2.26) с нормировочным множителем \tilde{z}

вместо множителя z . Для выбора значения параметра \tilde{z} можно рассмотреть несколько способов.

Первый способ состоит в использовании одного из основных свойств функции плотности

$$\int_{-\infty}^{+\infty} \hat{f}_n(t) dt = 1. \quad (2.27)$$

Согласно (2.3) оценка функции плотности представляется разложением в ряд по нормированным базисным функциям $\tilde{\psi}_i(t) = \tilde{z}\psi_i(t)$ в виде

$$\hat{f}_n(t) = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N \tilde{z}\psi_i(t)\tilde{z}\psi_i(x_j). \quad (2.28)$$

Тогда с учетом (2.28) соотношение (2.27) можно записать

$$\int_{-\infty}^{+\infty} \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N \tilde{z}\psi_i(t)\tilde{z}\psi_i(x_j) dt = 1,$$

а затем преобразовать к виду

$$\frac{\tilde{z}^2}{n} \sum_{j=1}^n \sum_{i=1}^N \psi_i(x_j) \int_{-\infty}^{+\infty} \psi_i(t) dt = 1.$$

Значение нормировочного множителя выражается следующим образом:

$$\tilde{z} = \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N \psi_i(x_j) \int_{-\infty}^{+\infty} \psi_i(t) dt \right)^{\frac{1}{2}}. \quad (2.29)$$

Вычислим нормировочный множитель для системы базисных функций на основе вейвлета «Мексиканская шляпа», для этого докажем следующее утверждение.

Утверждение 2.5. Пусть $\psi(t) = (1-t^2)e^{-\frac{t^2}{2}}$ – материнский вейвлет «Мексиканская шляпа», порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид

$$\psi_i(\tau) = \tilde{z}^2 (1-\tau^2) e^{-\frac{\tau^2}{2}},$$

где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, $k \geq 0$, $1 \leq j \leq 2^k$, $i = 2^k + j$. Тогда для любых i , k , j

нормировочный множитель $\tilde{z} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{\frac{1}{2}}$.

Доказательство. Аналитическая запись материнского вейвлета «Мексиканская шляпа» выражается соотношением $\psi(t) = (1-t^2)e^{-\frac{t^2}{2}}$, а система функций на его основе принимает вид $\psi_i(\tau) = \tilde{z} 2^{\frac{k}{2}} (1-\tau^2)e^{-\frac{\tau^2}{2}}$, где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, $k \geq 0$, $1 \leq j \leq 2^k$, $i = 2^k + j$.

Тогда с учетом (2.27) получаем

$$\begin{aligned} \tilde{z} &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \int_{-\infty}^{+\infty} (1-\tau^2) e^{-\frac{\tau^2}{2}} d\tau \right)^{\frac{1}{2}} = \\ &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \left(\int_{-\infty}^{+\infty} e^{-\frac{\tau^2}{2}} d\tau - \int_{-\infty}^{+\infty} \tau^2 e^{-\frac{\tau^2}{2}} d\tau \right) \right)^{\frac{1}{2}} = \\ &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) (2\sqrt{\pi} - \sqrt{2\pi}) \right)^{\frac{1}{2}} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{\frac{1}{2}}. \end{aligned}$$

Получаем значение $\tilde{z} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{\frac{1}{2}}$. Утверждение

доказано.

Аналогично определим нормировочный множитель для системы базисных функций на основе DOG вейвлета, для этого докажем следующее утверждение.

Утверждение 2.6. Пусть $\psi(t) = e^{-\frac{t^2}{2}} - \frac{1}{2} e^{-\frac{t^2}{8}}$ – материнский вейвлет DOG,

порождающий систему функций $\psi_i(t) \in L^2(R)$, которая принимает вид

$$\psi_i(\tau) = 2^{\frac{k}{2}} \tilde{z} \left(e^{-\frac{\tau^2}{2}} - \frac{1}{2} e^{-\frac{\tau^2}{8}} \right),$$

где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, $k \geq 0$, $1 \leq j \leq 2^k$, $i = 2^k + j$. Тогда для любых i , k , j

нормировочный множитель $\tilde{z} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{-\frac{1}{2}}$.

Доказательство. Аналитическая запись материнского вейвлета DOG выражается соотношением $\psi(t) = e^{-\frac{t^2}{2}} - \frac{1}{2} e^{-\frac{t^2}{8}}$, а система функций на его основе

принимает вид $\psi_i(\tau) = 2^{\frac{k}{2}} \tilde{z} \left(e^{-\frac{\tau^2}{2}} - \frac{1}{2} e^{-\frac{\tau^2}{8}} \right)$, где $\tau = \frac{2^k}{d-c}(t-c) - (j-1)$, $k \geq 0$,

$1 \leq j \leq 2^k$, $i = 2^k + j$.

Тогда согласно (2.27) получаем

$$\begin{aligned} \tilde{z} &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \int_{-\infty}^{+\infty} e^{-\frac{\tau^2}{2}} - \frac{1}{2} e^{-\frac{\tau^2}{8}} d\tau \right)^{-\frac{1}{2}} = \\ &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \left(\int_{-\infty}^{+\infty} e^{-\frac{\tau^2}{2}} d\tau - \frac{1}{2} \int_{-\infty}^{+\infty} e^{-\frac{\tau^2}{8}} d\tau \right) \right)^{-\frac{1}{2}} = \\ &= \left(\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) (2\sqrt{\pi} - \sqrt{2\pi}) \right)^{-\frac{1}{2}} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{-\frac{1}{2}}. \end{aligned}$$

Получаем значение $\tilde{z} = \left(\frac{\sqrt{\pi}}{n} (2 - \sqrt{2}) \sum_{j=1}^n \sum_{i=1}^N 2^{-\frac{k}{2}} \psi_i(x_j) \right)^{-\frac{1}{2}}$. Утверждение

доказано.

Второй способ заключается в анализе степени близости $\hat{f}_n(t)$ и $f(t)$ путем минимизации значения статистики χ^2 [2, 14, 25]:

$$\chi^2 = \sum_{r=1}^T \frac{(O_r - np_r)^2}{np_r}, \quad (2.30)$$

где T – количество интервалов разбиения выборки, полученное по формуле Стерджеса [2, 25]; t_r – середина r -го интервала; O_r – фактическое число

элементов выборки попавших в r -й интервал; p_r – вероятность r -го исхода в n испытаниях. Справедливо следующее утверждение $p_r = \int_{x_r}^{x_{r+1}} f(x) dx$ из которого следует, что $p_r = f(t_r)(x_{r+1} - x_r)$. Тогда (2.30) можно преобразовать:

$$\chi^2 = n \sum_{r=1}^T (x_{r+1} - x_r) \frac{\left(\frac{O_r}{n} - f(t_r) \right)^2}{f(t_r)}. \quad (2.31)$$

С учетом соотношения (2.28) выражение (2.31) принимает вид

$$\chi^2 = \sum_{r=1}^T (x_{r+1} - x_r) \frac{\left(O_r - \sum_{j=1}^n \sum_{i=1}^N \tilde{z}^2 \psi_i(t_r) \psi_i(x_j) \right)^2}{\sum_{j=1}^n \sum_{i=1}^N \tilde{z}^2 \psi_i(t_r) \psi_i(x_j)}.$$

Тогда искомое значение \tilde{z}^* будет получено как решение следующей оптимизационной задачи:

$$\tilde{z}^* = \arg \min_{\tilde{z}} \chi^2. \quad (2.32)$$

Процесс минимизации можно выполнить с помощью любого метода одномерной оптимизации, предварительно локализовав интервал значений \tilde{z} наиболее удобным способом [50, 54, 55]. В этом случае исходная функция плотности распределения должна быть известна априорно, что не всегда выполнимо на практике. Поэтому $f(t)$ можно заменить на ее эмпирическую оценку, полученную по выборке. Это фактически можно считать третьим способом определения \tilde{z} .

В результате вейвлет-оценка функции плотности $\hat{f}_n(t)$ случайной величины на произвольном отрезке $[c, d]$ с использованием материнского вейвлета «Мексиканская шляпа» или DOG-вейвлета выражается разложением (2.1) по нормированным базисным функциям (2.23) или (2.26) соответственно.

Отметим, что системы функций (2.21) и (2.24) можно ортонормировать путем процедуры ортогонализации Грама – Шмидта [33]. Особенностью указанного метода является необходимость коррекции результата на каждом шаге выполнения

процедуры ортогонализации, что связано с сопутствующими ей искажениями соответствующих параметров [46]. Поэтому предложенные методы определения нормировочного множителя \tilde{z} следует считать удобным инструментом повышения качества оценок.

2.4 Исследование качества оценок функции плотности на основе преобразования Фурье

Для определения качества восстановления функции плотности с помощью анализа Фурье был проведен ряд вычислительных экспериментов. Исследования проводились для стандартного нормального распределения при различных значениях параметра сглаживания. Прежде всего было исследовано качество восстановления функции плотности в зависимости от количества членов ряда (2.1) и определено его наилучшее значение. Используя найденное наилучшее значение числа членов ряда, было исследовано влияние объема выборки на качество оценивания функции плотности.

Для решения первой задачи были смоделированы 1000 выборок, каждая из которых имеет объем $n = 200$. Выборки состояли из независимых значений случайной величины, распределенной по нормальному закону с параметрами $(0,1)$. Для этих выборок были построены при каждом значении N , начиная с 5 - го и до 50 - го, оценки плотности распределения (2.1) по ортонормированным базисным функциям (2.10). Результаты восстановления $f(t)$ представлены на рисунках 2.1, *a–в* для $N = 5$, $N = 10$ и $N = 20$ соответственно. Отметим, что вместе с полученными оценками на рисунке 2.1 представлен график истинной плотности распределения стандартного нормального закона $f(t)$. Результаты восстановления $f(t)$ для $N = 5$ показывают, что имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности. Близкие результаты оценивания функции плотности к $N = 5$ наблюдаются и при числе членов ряда $N = 6$, $N = 7$ и $N = 8$. В случае, когда значение параметра сглаживания $N = 9$, $N = 10$, $N = 11$ и $N = 12$ оценки

степеней свободы $\nu = 8$ для $\alpha \leq 0,05$ и $\alpha \leq 0,01$ равны $\chi^2_{кр} = 16.919$ и $\chi^2_{кр} = 21.666$ соответственно. Усредненные значения статистики χ^2 по 1000 смоделированным выборкам объемом 200 представлены в таблице 2.1.

Таблица 2.1 – Значения статистики χ^2 в зависимости от параметра сглаживания для преобразования Фурье

Статистика χ^2	$N = 5$	$N = 10$	$N = 20$
		17.951	9.353

Из таблицы 2.1 видно, что наименьшее значение статистики χ^2 наблюдается при $N = 10$, что соответствует результатам представленным на рисунке 2.1. Таким образом, можно сделать вывод, что существует наилучшее значение параметра сглаживания N для оценок функции плотности (2.1) по ортонормированным базисным функциям (2.10) – это $N = 10$. Также были проведены исследования зависимости качества получаемой оценки от величины параметра сглаживания N с объемами выборки $n = 100$ и $n = 500$. Полученные результаты качественно совпали с приведенными выше.

Результаты статистического моделирования, показывающие зависимость качества оценки функции плотности от объема выборки, для выборок объемами $n = 100$ и $n = 500$ при $N = 10$ представлены на рисунках 2.2, *а* и *б* соответственно.

Количественная характеристика зависимости качества $\hat{f}_n(t)$ от объема выборки выполнена по критерию χ^2 . В таблице 2.2 представлены усредненные значения статистики χ^2 по 1000 выборкам. Также в таблице приведены критические значения критерия $\chi^2_{кр}$ Пирсона при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$. Результаты расчета согласия из таблицы 2.2 позволяют сделать вывод, что объем исходной выборки влияет на качество оценки и величину отклонения от истинной функции плотности распределения. С увеличением объема

Таблица 2.3 – Время работы алгоритма для оценки дискретного и быстрого преобразования Фурье

N	БПФ, с	ДПФ, с
5	0.107	0.228
10	0.208	0.401
15	0.358	0.722
20	0.107	0.228

Согласно результатам таблицы 2.3 время работы быстрого преобразования почти в 2 раза меньше по сравнению с временем работы дискретного преобразования.

2.5 Исследование качества оценок функции плотности на основе ортогональных вейвлетов

Для сравнения качества восстановления функции плотности на основе ортогональных вейвлетов, таких как вейвлет Хаара, LITTLEWOOD & PALEY, Морле, также был проведён ряд вычислительных экспериментов. Исследования проходили при разных условиях формирования выборок и различных значениях параметра сглаживания. Во-первых, было выполнено исследование качества восстановления функции плотности в зависимости от количества членов ряда. Опираясь на результаты статистического моделирования, было определено наилучшее значение параметра сглаживания. Во-вторых, используя найденное наилучшее значение числа членов ряда, было исследовано влияние объема выборки на качество оценивания функции плотности.

2.5.1 Вейвлет Хаара

Прежде всего был проведен поиск наилучшего значения параметра сглаживания. Для этого были смоделированы выборки, каждая из которых имеет

Оценка функции плотности, построенная с помощью вейвлета Хаара, имеет ступенчатый вид, число ступеней меняется в зависимости от параметра сглаживания. Для $N = 5$ получаем достаточно грубую оценку функции плотности всего 4 ступени, при увеличении числа членов ряда количество ступеней увеличивается (в частности при $N = 25$ их становится 20). Наилучший результат наблюдается при $N = 16$. Увеличение числа ступеней ведет к потере качества оценивания функции плотности, в виде наличия дополнительных экстремумов.

На рисунке 2.4 показаны результаты моделирования для случайной величины, распределенной по нормальному закону с параметрами $(6, 1.5)$.

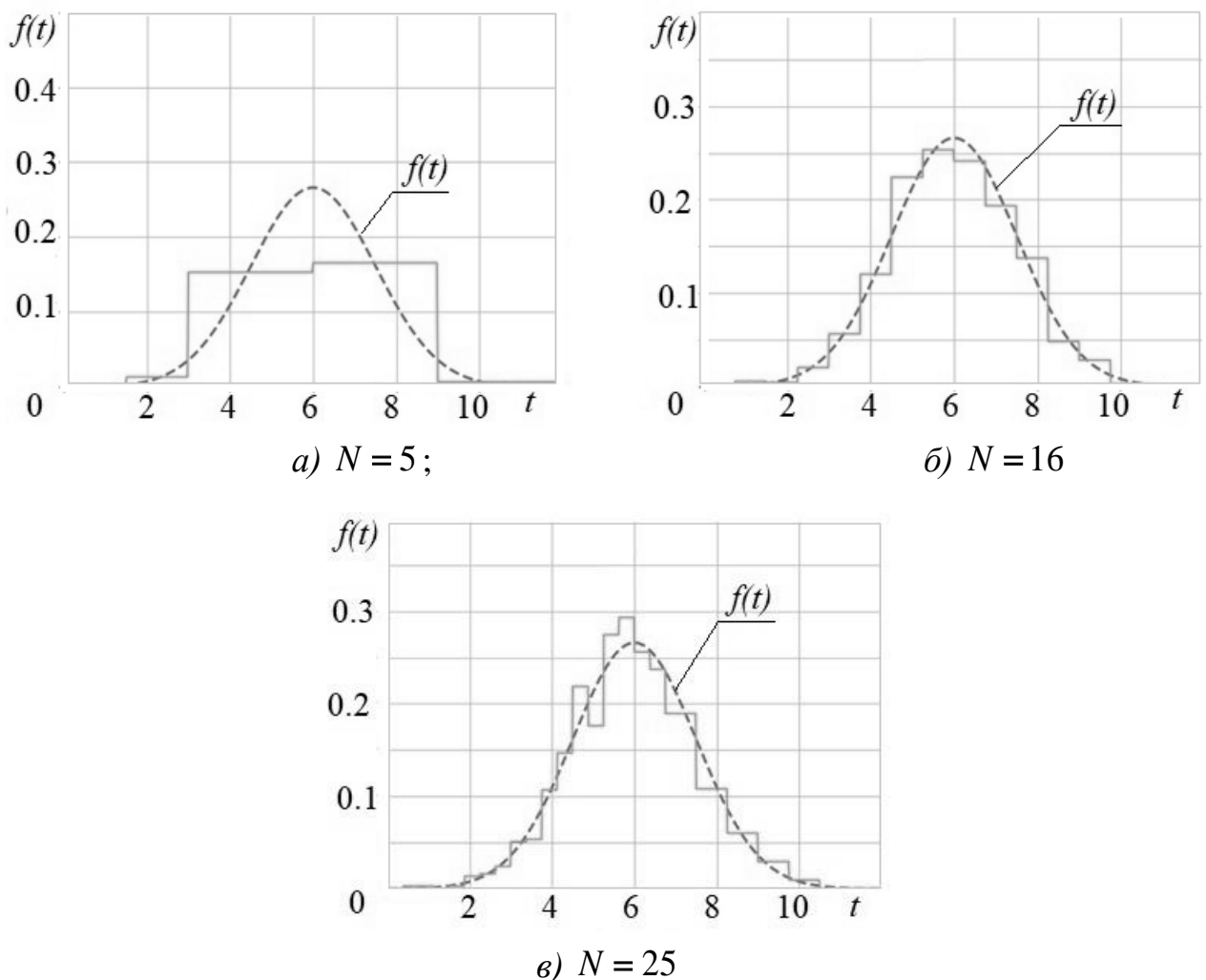


Рисунок 2.4 – Оценка плотности распределения на основе вейвлета Хаара для $N(6, 1.5)$: а) $N = 5$; б) $N = 16$; в) $N = 25$

Аналогично результатам, представленным выше, при $N = 5$ оценка функции плотности распределения содержит 4 ступени. С увеличением значения параметра сглаживания количество ступеней увеличивается. При $N = 16$ отмечается наилучшее качество оценивания функции плотности. При $N = 25$ получаем 20 ступеней и наличие точек минимума, которые отсутствуют у истинной функции плотности. В ходе вычислительных экспериментов было установлено, что для нормального распределения наилучшее значение параметра сглаживания не меняется.

Результаты восстановления $f(t)$ для экспоненциального распределения для $N = 5$, $N = 16$ и $N = 25$ представлены на рисунках 2.5, а–в соответственно.

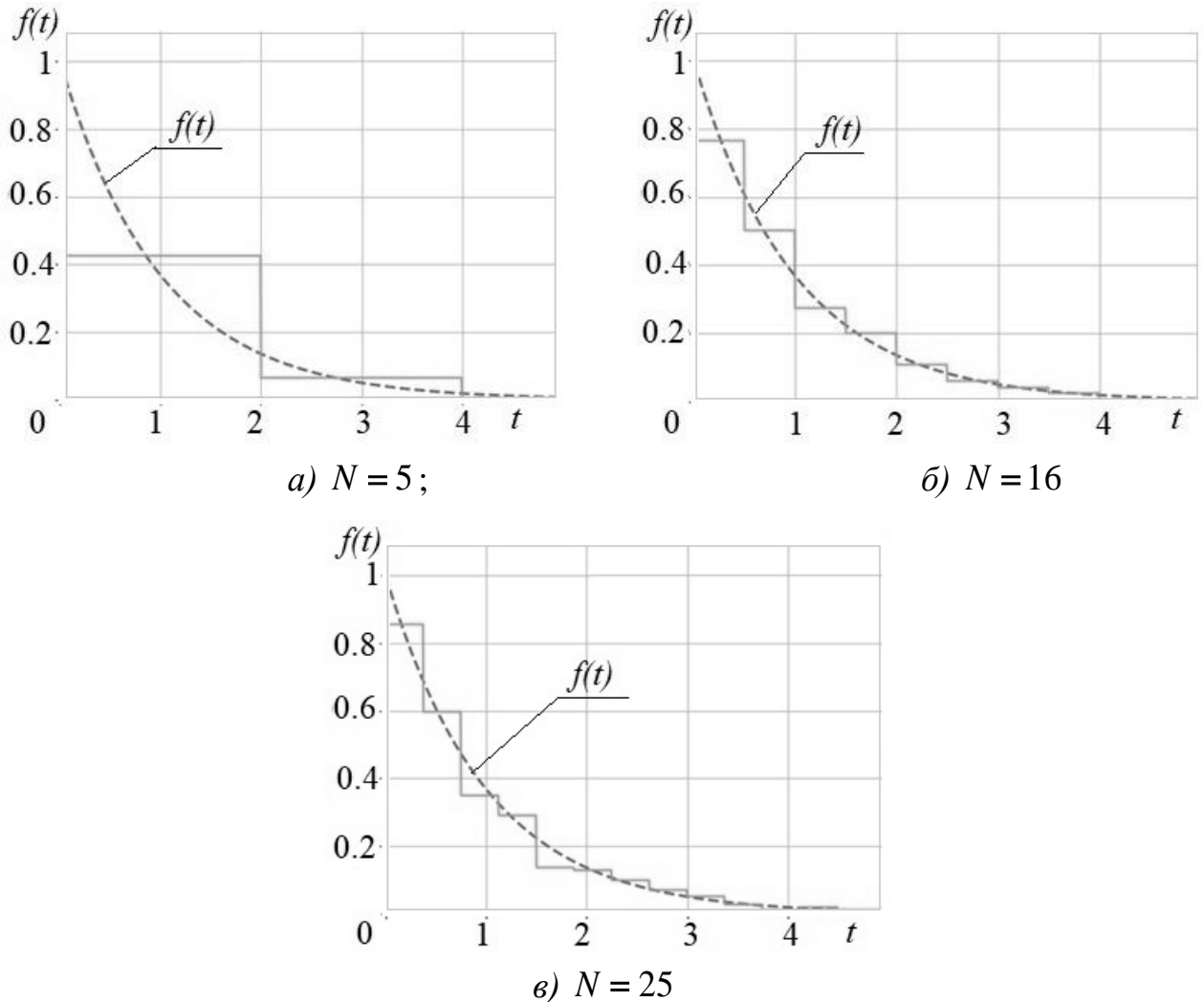


Рисунок 2.5 – Оценка плотности распределения на основе вейвлета Хаара для экспоненциального распределения: а) $N = 5$; б) $N = 16$; в) $N = 25$

Согласно результатам статистического моделирования при значении $N = 5$ оценка функции плотности неудовлетворительного качества и содержит всего 3 ступени. С увеличением числа членов ряда качество вейвлет-оценки на основе материнского вейвлета Хаара улучшается. При $N = 16$ оценка функции плотности состоит из 9 ступеней и оказывается наилучшей. Дальнейшее увеличение значения параметра сглаживания приводит к росту числа ступеней. При $N = 25$ оценка функции плотности содержит дополнительные экстремумы, которые ухудшают её качество.

Количественная оценка степени близости $\hat{f}_n(t)$ и $f(t)$ была проведена с помощью проверки согласия по критерию χ^2 . Критические значения критерия χ^2 при числе степеней свободы $\nu = 8$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены разделе 2.4. Усредненные результаты значения статистики χ^2 по 1000 выборкам объемом 200 представлены в таблице 2.4, из которой видно, что значения статистики χ^2 при $N = 16$ наименьшие для всех рассмотренных распределений, а значит, при таком значении параметра сглаживания оценка $\hat{f}_n(t)$ и теоретическая функция максимально близки. По результатам исследований можно отметить, что для вейвлет оценки на основе вейвлета Хаара наилучшим значением параметра сглаживания является $N = 16$.

Таблица 2.4 – Значения статистики χ^2 в зависимости от параметра сглаживания для вейвлета Хаара

Вид распределения	Статистика χ^2		
	$N = 5$	$N = 16$	$N = 25$
$N(0,1)$	19.456	13.147	15.09
$N(6,1.5)$	20.537	15.494	19.418
Экспоненциальное ($\lambda = 1$)	26.709	16.589	22.382

Далее были проведены исследования зависимости качества получаемой оценки от величины параметра сглаживания N с объемами выборки $n=100$ и $n=500$. Полученные результаты качественно совпали с приведенными выше. Результаты статистического моделирования при $N=16$ для выборок объемом $n=100$ и $n=500$, состоящих из независимых значений случайной величины ξ , распределенной стандартному нормальному закону приведены на рисунках 2.6, *a* и *б* соответственно.

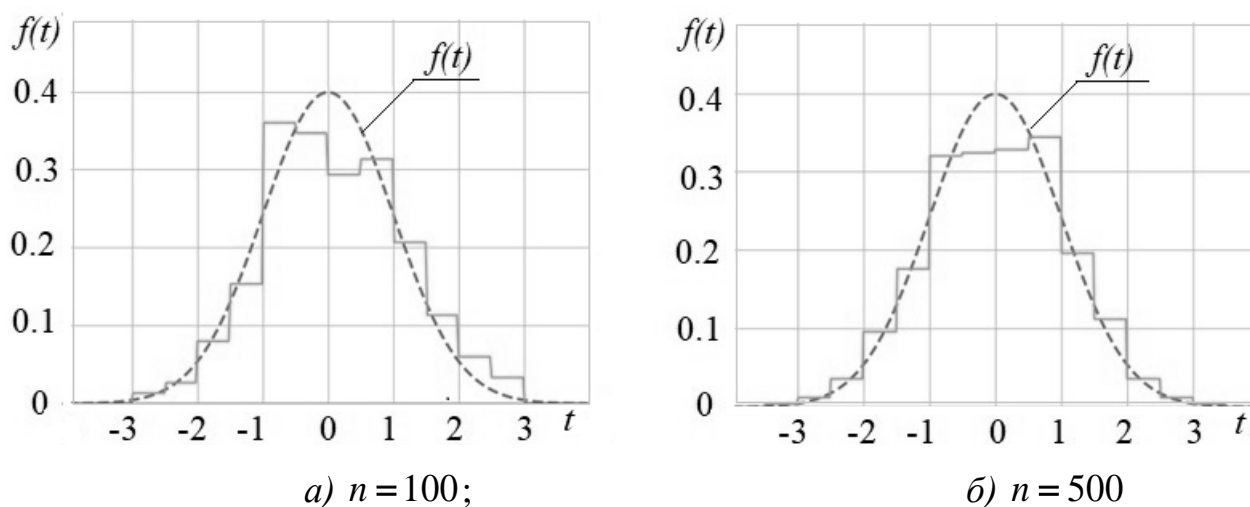


Рисунок 2.6 – Оценка плотности распределения на основе вейвлета Хаара:

a) n = 100; б) n = 500

Количественная характеристика зависимости качества оценивания функции плотности от объема выборки была проведена с помощью проверки согласия по критерию χ^2 . Усредненные результаты значения статистики χ^2 по 1000 выборкам объемом 200 представлены в таблице 2.5. Также в таблице 2.5 приведены критические значения критерия $\chi^2_{кр}$ при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$. Из таблицы 2.5 следует, что с увеличением объема выборки, значение χ^2 уменьшается, а качество восстановления функции плотности становится лучше. Объем выборки не значительно влияет на качество восстановления функции плотности.

Таблица 2.5 – Значения статистики χ^2 в зависимости от объема выборки для вейвлета Хаара

n	T	$\chi^2_{кр}$		Статистика χ^2
		$\alpha \leq 0.05$	$\alpha \leq 0.01$	
100	8	15.507	20.09	13.584
200	9	16.919	21.666	13.147
500	10	18.307	23.209	12.909

Согласно [3], если доопределить набор ортонормированных базисных функций (2.12) для $(N+1) \leq i \leq 2N$, то предполагаемое значение параметра сглаживания может быть найдено с помощью соотношение

$$N_{\min} = \arg \min_{1 < N < n} \tau_n, \quad (2.32)$$

$$\text{где } \tau_n = \sum_{j=N+1}^{2N} \left(\frac{1}{n} \sum_{i=1}^n \tilde{\psi}_j(x_i) \right)^2.$$

Результаты оценивания значения параметра N согласно (2.32) приведены в таблице 2.6. Рассматривались смоделированные выше выборки объемом $n = 1000$, состоящие из независимых значений случайной величины, распределенной по нормальному закону с параметрами $(0, 1)$. Из таблицы 2.6 видно, что значения параметра сглаживания не могут быть использованы для построения оценок функции плотности (2.1) по ортонормированным базисным функциям (2.15).

Таблица 2.6 – Оценка параметра сглаживания

n	100	500	1000
N_{\min}	0.00245	0.00045	0.00002
\tilde{N}_{\min}	15.037	15.361	16.022

Проводя численные эксперименты, удалось установить, что наилучшее значение N должно быть не меньше значения, вычисленного следующим образом

$$\tilde{N}_{\min} = \arg \min_{1 < N < 35} \left[1000 \sum_{j=N+1}^{2N} \left(\frac{1}{T} \sum_{i=1}^T \tilde{\psi}_j(t_i) \right)^2 \right], \quad (2.33)$$

где T – количество интервалов разбиения выборки, полученное по формуле Стерджеса [2], t_i – середина i – го интервала. Результаты оценки параметра сглаживания с использованием соотношения (2.33) также приведены в таблице 2.6 из которой видно, что значения \tilde{N}_{\min} близки к наилучшим, полученным при моделировании.

2.5.2 Вейвлет LITTLEWOOD & PALEY

По аналогии с исследованием вейвлет-оценки на основе вейвлета Хаара было выполнено исследование качества восстановления функции плотности на основе вейвлета LITTLEWOOD & PALEY. Рассматривались выборки объемом $n = 200$, смоделированные в разделе 2.5.1. В первую очередь, было исследовано качество восстановления $\hat{f}_n(t)$ в зависимости от количества членов ряда N . Для этих выборок при каждом значении N , начиная с 5-го и до 50-го, были построены вейвлет оценки функции плотности по ортонормированным базисным функциям (2.17). Результаты восстановления $f(t)$ при $N = 5$, $N = 10$ и $N = 20$ для нормального закона распределения с параметрами $(0, 1)$ представлены на рисунках 2.9, а–в соответственно. Вместе с полученными оценками на рисунке 2.7 представлен график истинной плотности распределения $f(t)$. Из рисунков 2.7, а–в видно, что значение параметра сглаживания N влияет на величину отклонения вейвлет-оценки от истинной плотности распределения. При малом значении $N = 5$ оценка $\hat{f}_n(t)$ и теоретическая функция очень близки. При $N = 10$ и $N = 20$ вейвлет оценки функции плотности на основе LITTLEWOOD & PALEY существенно отклоняются от $f(t)$. Такое поведение $\hat{f}_n(t)$ оказалось характерным для $N \geq 10$, что было установлено в результате серии вычислительных экспериментов.

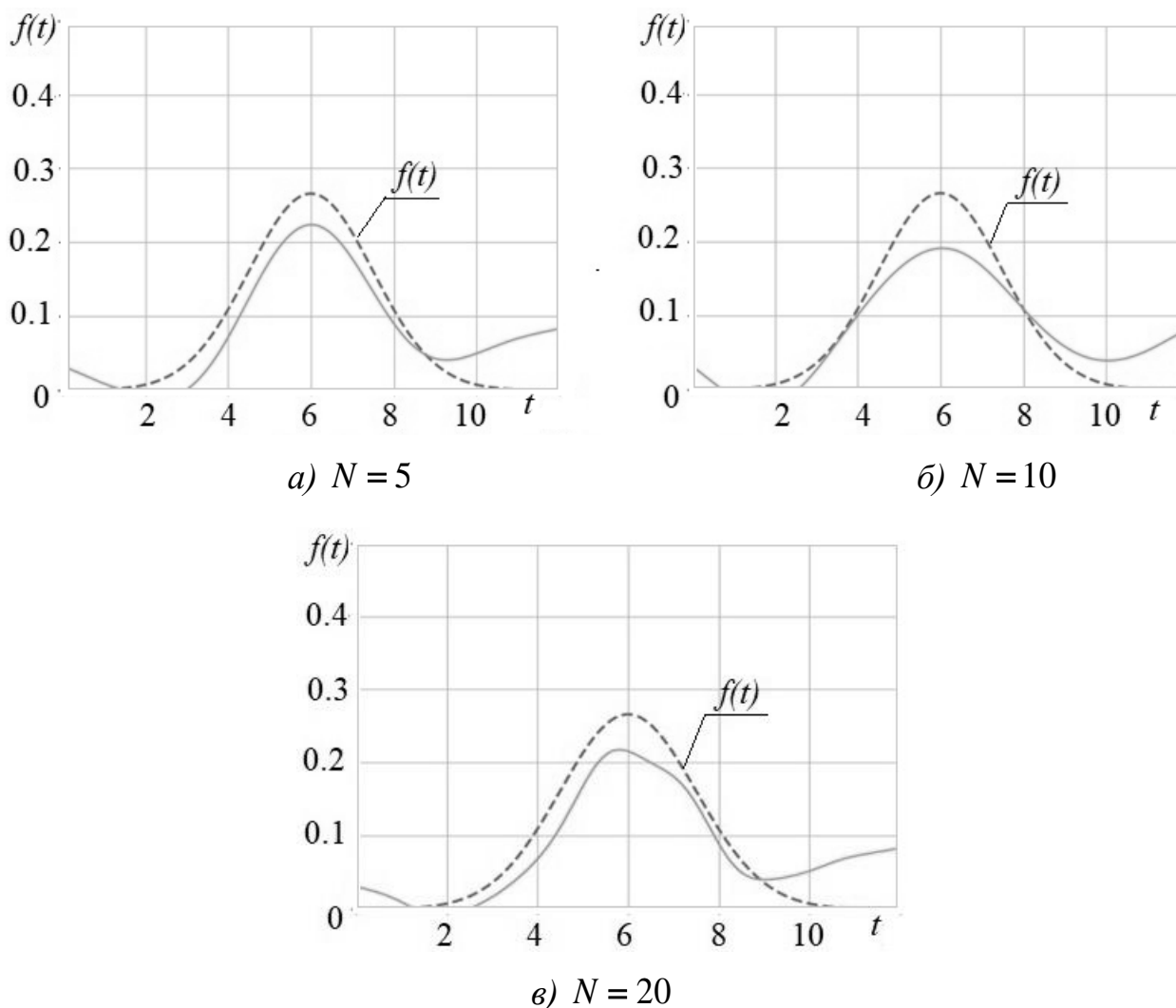


Рисунок 2.8 – Оценка плотности распределения на основе вейвлета LITTLEWOOD & PALEY для $N(6,1.5)$: а) $N = 5$; б) $N = 10$; в) $N = 20$

На рисунке 2.9, а–в представлены результаты восстановления функции плотности для экспоненциального распределения при $N = 5$, $N = 10$ и $N = 20$ соответственно. Вместе с полученными оценками на рисунках 2.9, а–в представлен график истинной плотности распределения $f(t)$. Говорить о качественном восстановлении функции плотности не приходится. Наилучшее значение параметра сглаживания соответствует $N = 5$. При $N = 10$ и $N = 20$ у вейвлет-оценки на основе вейвлета LITTLEWOOD & PALEY появляются дополнительные экстремумы, что приводит к ложному представлению о поведении истинной функции плотности распределения. Дальнейшее увеличение значения параметра

сглаживания приводит к увеличению числа экстремумов у оценки функции плотности и, следовательно, к ухудшению качества восстановления $f(t)$.

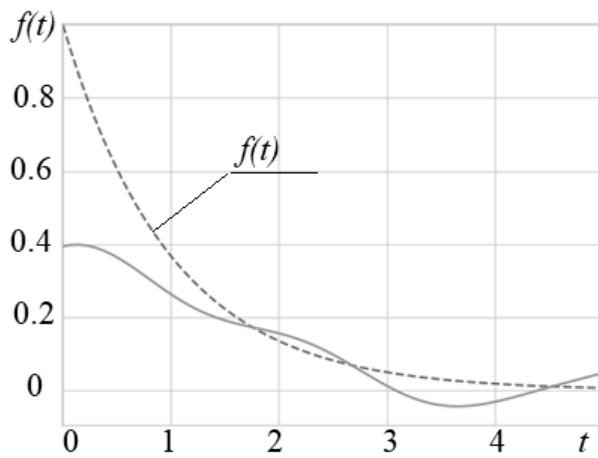
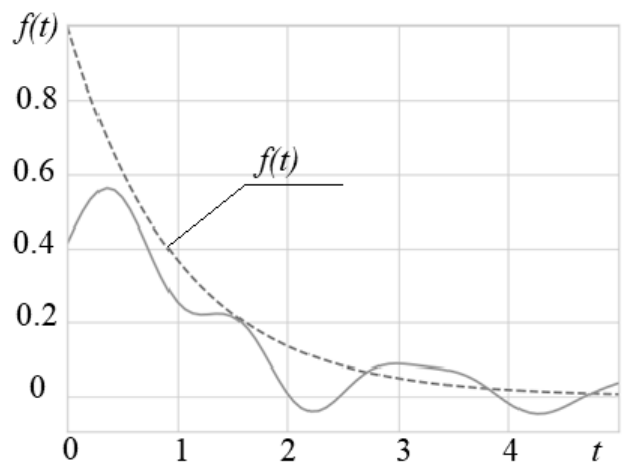
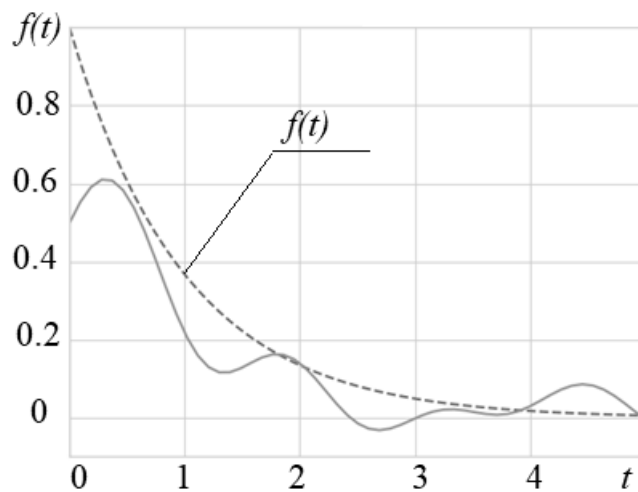
а) $N = 5$ б) $N = 10$ в) $N = 20$

Рисунок 2.9 – Оценка плотности распределения на основе вейвлета

LITTLEWOOD & PALEY: а) $N = 5$; б) $N = 10$; в) $N = 20$

Для количественной оценки степени близости $\hat{f}_n(t)$ и $f(t)$ была проведена проверка согласия по критерию χ^2 . Критические значения критерия χ^2 при числе степеней свободы $\nu = 8$ для $\alpha \leq 0,05$ и $\alpha \leq 0,01$ приведены в разделе 2.4. Усредненные результаты значения статистики χ^2 по 1000 выборкам, каждая из которых объемом 200, представлены в таблице 2.7. Данные таблицы 2.7

показывают, что для рассматриваемых распределений значение статистики χ^2 при $N = 5$ наименьшее, а значит оценка $\hat{f}_n(t)$ обладает наилучшим качеством восстановления функции плотности.

Таблица 2.7 – Значения статистики χ^2 в зависимости от параметра сглаживания для вейвлета LITTLEWOOD & PALEY

Вид распределения	Статистика χ^2		
	$N = 5$	$N = 10$	$N = 20$
$N(0,1)$	11.405	15,782	24.081
$N(6,1.5)$	15.923	20.234	28.561
Экспоненциальное ($\lambda = 1$)	24.023	62.940	84.412

Также были проведены исследования зависимости качества получаемой оценки $\hat{f}_n(t)$ от величины параметра сглаживания N с объемами выборки $n = 100$ и $n = 500$. Результаты качественно совпали с приведенными выше и позволили сделать вывод, что наилучшее значение для данного вейвлета будет $N = 5$.

Примеры восстановления функции плотности при $N = 5$ для выборок объемом $n = 100$ и $n = 500$, состоящих из независимых значений случайной величины ξ , распределенной стандартному нормальному закону приведены на рисунках 2.10, *a* и *б* соответственно. Рисунок 2.10 показывает, что объем исходной выборки влияет на качество вейвлет оценки и величину отклонения от истинной функции плотности распределения. С увеличением объема исходной выборки качество вейвлет-оценки улучшается незначительно.

Количественная характеристика зависимости качества $\hat{f}_n(t)$ от объема выборки выполнена по критерию χ^2 . Усредненные результаты значения статистики χ^2 по 1000 выборкам представлены в таблице 2.8. Также в таблице

приведены критические значения критерия $\chi^2_{кр}$ Пирсона при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$.

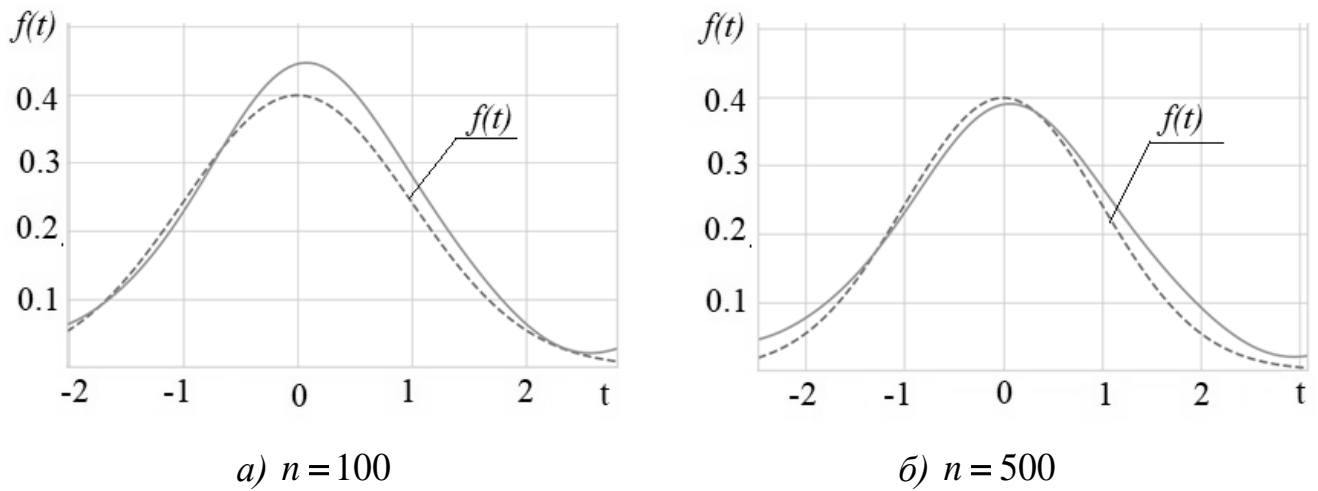


Рисунок 2.10 – Оценка плотности распределения на основе вейвлета LITTLEWOOD & PALEY: а) $n = 100$; б) $n = 500$

Таблица 2.8 – Значения статистики χ^2 в зависимости от объема выборки для вейвлета LITTLEWOOD & PALEY

n	T	$\chi^2_{кр}$		Статистика χ^2
		$\alpha \leq 0.05$	$\alpha \leq 0.01$	
100	8	15.507	20.09	11.272
200	9	16.919	21.666	10.905
500	10	18.307	23.209	10.621

Отметим, что согласно таблице 2.8 объем выборки оказывает не значительное влияние на качество оценки $\hat{f}_n(t)$, что согласуется с рисунком 2.12.

2.5.3 Вейвлет Морле

Перейдем к исследованию восстановления функции плотности распределения вейвлет-оценки на основе вейвлета Морле. Рассматривались выборки,

смоделированные в разделе 2.5.1. Порядок исследования совпадал с представленным для вейвлета Хаара и LITTLEWOOD & PALEY. Оценки функции плотности (2.1) строились по нормированным базисным функциям (2.20). Результаты восстановления $f(t)$ при $N = 5$, $N = 10$ и $N = 20$ для стандартного нормального распределения представлены на рисунках 2.11, а–в соответственно.

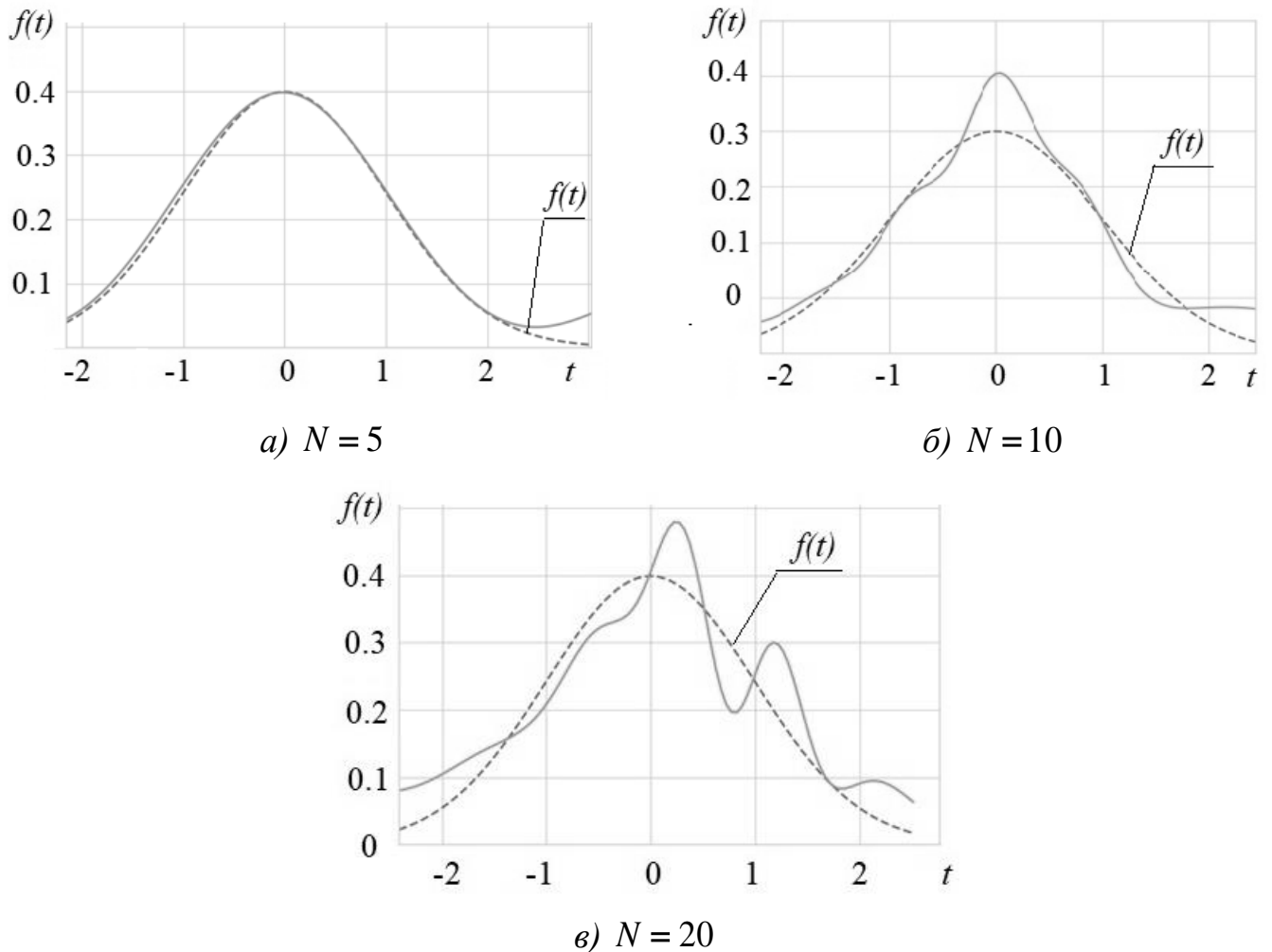


Рисунок 2.11 – Оценка плотности распределения на основе вейвлета Морле для $N(0, 1)$: а) $N = 5$; б) $N = 10$; в) $N = 20$

При числе членов ряда $N = 5$ качество восстановления функции плотности наилучшее. При $N = 10$ и $N = 20$ качество $\hat{f}_n(t)$ ухудшается. Отметим, что дальнейшее увеличение числа членов ряда приводит к наибольшему отклонению

вейвлет-оценки, полученной на основе вейвлета Морле, от истинной функции плотности $f(t)$.

Далее на рисунках 2.12, *а–в* приведены результаты для выборки, смоделированной по нормальному закону распределения с параметрами (6, 1.5) при $N = 5$, $N = 10$ и $N = 20$ соответственно.

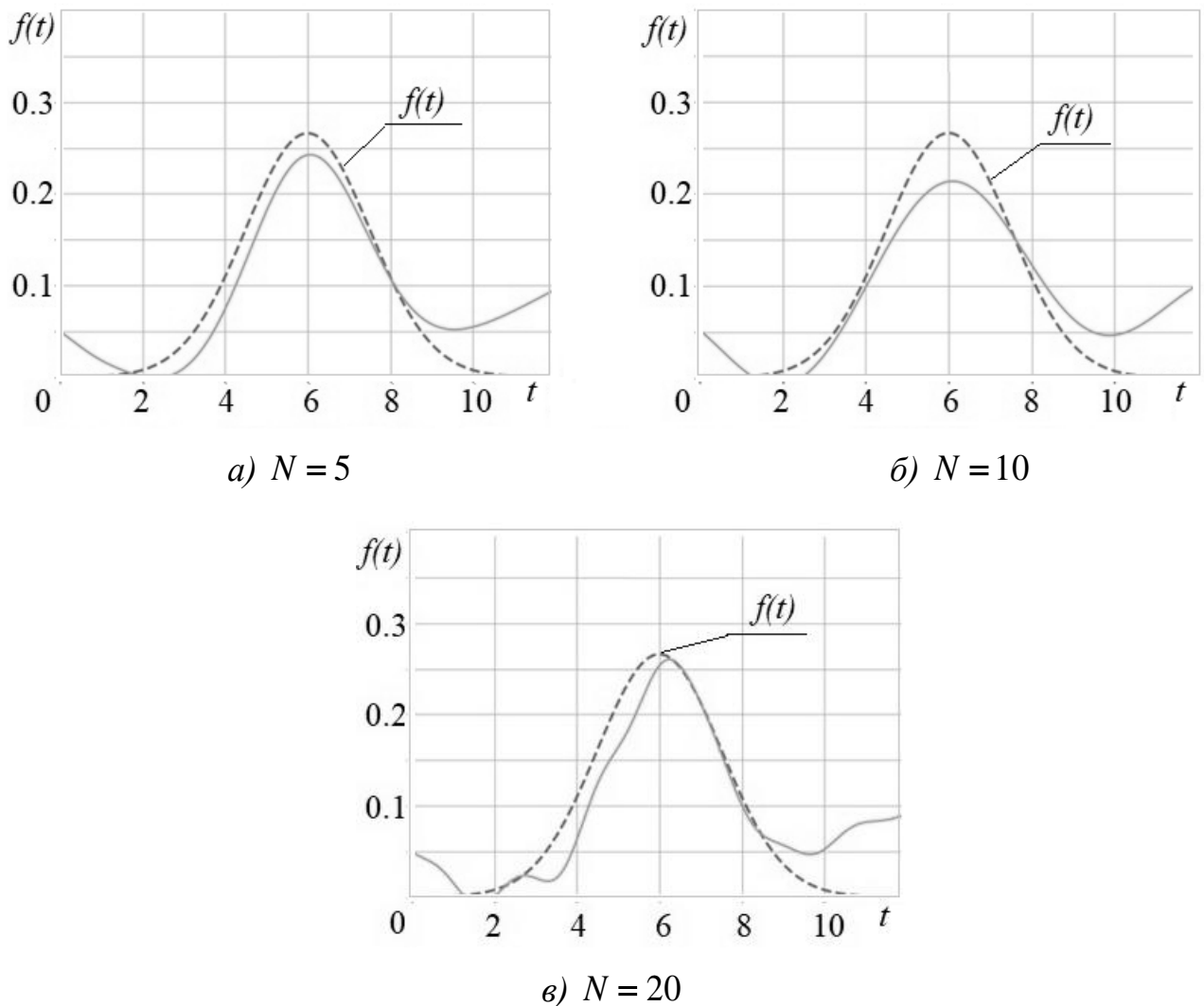


Рисунок 2.12 – Оценка плотности распределения на основе вейвлета Морле для $N(6, 1.5)$: а) $N = 5$; б) $N = 10$; в) $N = 20$

Наилучшее качество восстановления $f(t)$ отмечается при $N = 5$. При $N = 10$ и при $N = 20$ наблюдается существенное отклонение вейвлет-оценки от истинной плотности распределения. Как показали вычислительные эксперименты, дальнейшее увеличение значения параметра сглаживания приводит к еще более

значительному отклонению вейвлет-оценки $\hat{f}_n(t)$ от истинной плотности распределения, а значит, к потере качества восстановления $f(t)$.

На рисунке 2.13, а–в приведены результаты моделирования для экспоненциального закона распределения с параметром $\lambda = 1$. При числе членов ряда $N = 5$ качество восстановления функции наилучшее. При $N = 10$ и $N = 20$ качество оценки функции плотности ухудшается за счет появления дополнительных экстремумов. С увеличением значения параметра сглаживания число экстремумов оценки $\hat{f}_n(t)$ увеличивается.

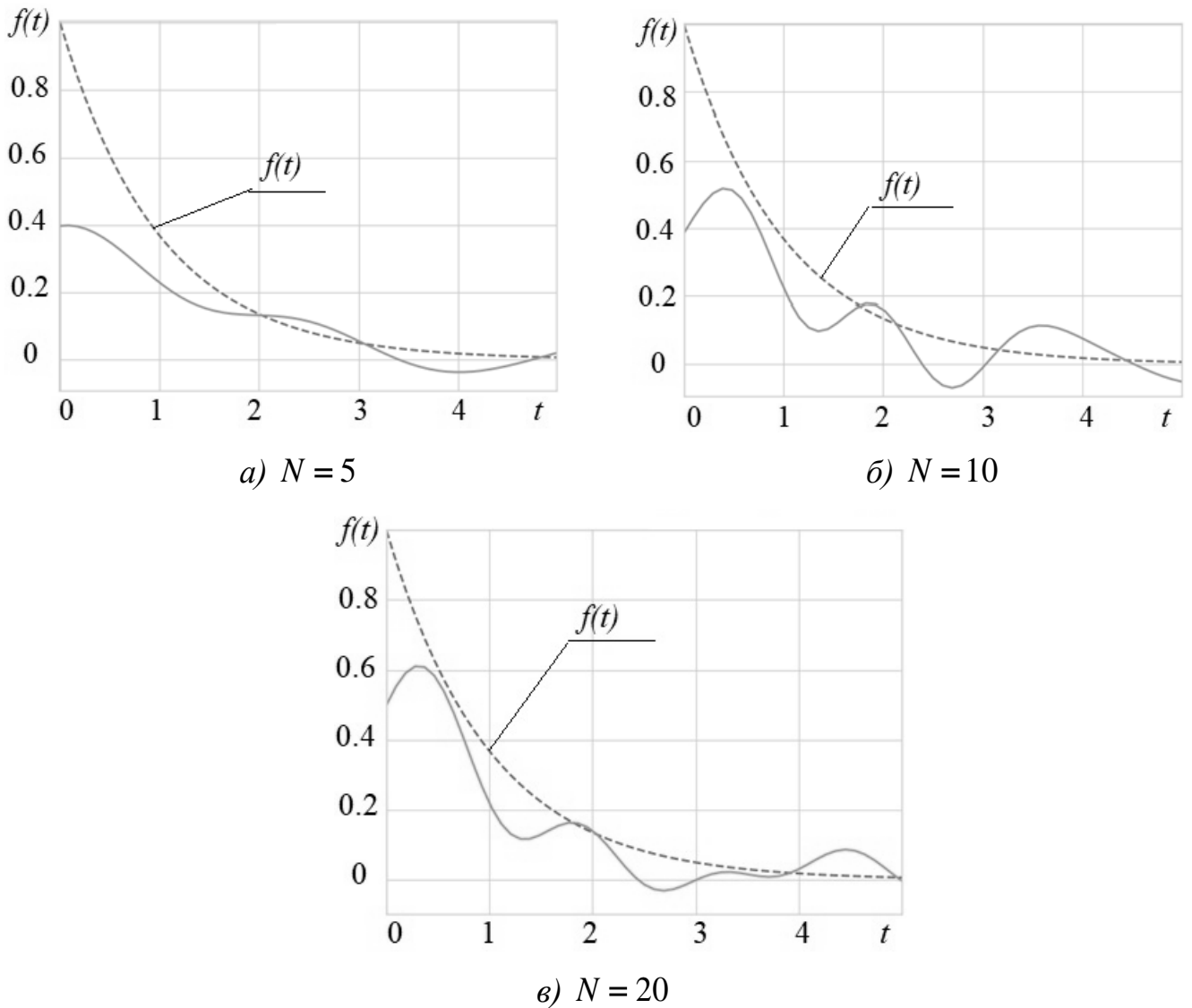


Рисунок 2.13 – Оценка плотности распределения на основе вейвлета Морле для экспоненциального закона распределения: а) $N = 5$; б) $N = 10$; в) $N = 20$

Усредненные результаты значения статистики при $N = 5$, $N = 10$ и $N = 20$ по 1000 вычислительным экспериментам представлены в таблице 2.9, объем выборок -200. Критические значения критерия χ^2 при числе степеней свободы $\nu = 8$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены в разделе 2.4.

Таблица 2.9 – Значения статистики χ^2 в зависимости от показателя сглаживания для вейвлета Морле

Вид распределения	Статистика χ^2		
	$N = 5$	$N = 10$	$N = 20$
$N(0, 1)$	10.527	14.162	23.909
$N(6, 1.5)$	15.561	26.528	32.665
Экспоненциальное ($\lambda = 1$)	23.937	49.469	92.047

Из таблицы 2.9 следует, что для рассматриваемых распределений значение статистики χ^2 при $N = 5$ наименьшее, а значит оценка $\hat{f}_n(t)$ наиболее качественно восстанавливает $f(t)$.

Исследования зависимости качества получаемой оценки от величины N с объемами выборки $n = 100$ и $n = 500$ позволили сделать вывод, что наилучшее значение параметра сглаживания для вейвлета Морле – $N = 5$.

Результаты для выборок объемом $n = 100$ и $n = 500$ из независимых значений случайной величины ξ , распределенной по нормальному закону с параметрам $(0, 1)$ при наилучшем значении $N = 5$ представлены на рисунках 2.14, *a* и *б* соответственно. Результаты, представленные на рисунке 2.14, позволяют сделать вывод, что объеме исходной выборки не значительно влияет на величину отклонения $\hat{f}_n(t)$ от истинной функции плотности распределения. Вместе с полученными оценками на рисунках 2.16, *a* и *б* представлен график истинной плотности распределения $f(t)$.

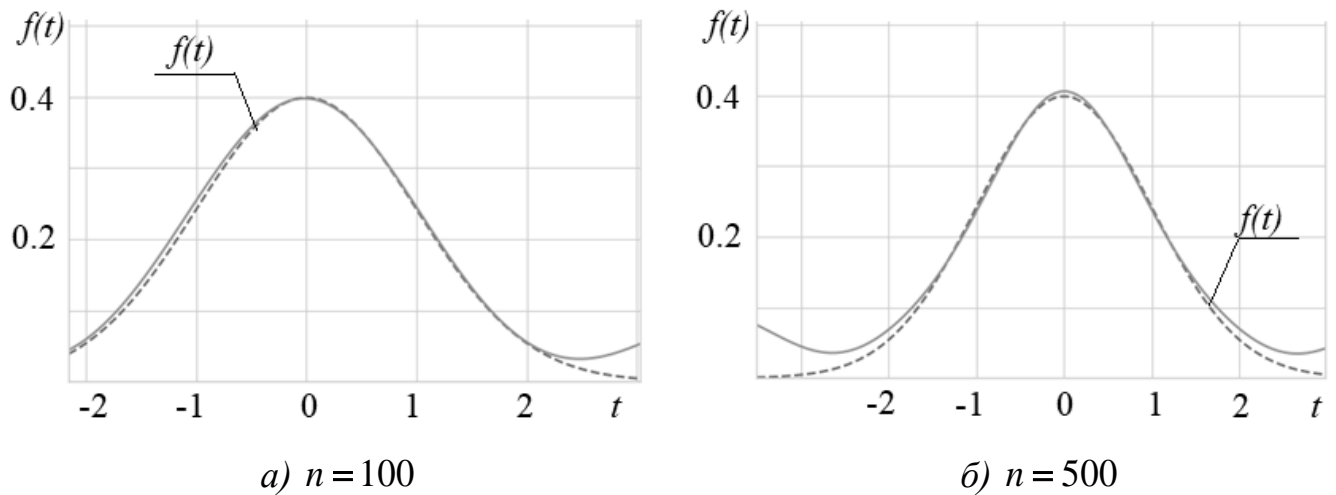


Рисунок 2.14 – Оценка плотности распределения на основе вейвлета Морле:

а) $n = 100$; б) $n = 500$

Усредненные значения статистики χ^2 по 1000 вычислительным экспериментам для различных объемов выборки представлены в таблице 2.10. Критические значения критерия $\chi^2_{кр}$ при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ также приведены в таблице 2.5.

Таблица 2.10 – Значения статистики χ^2 в зависимости от объема выборки для вейвлета Морле

n	T	$\chi^2_{кр}$		Статистика χ^2
		$\alpha \leq 0.05$	$\alpha \leq 0.01$	
100	8	15.507	20.09	10.724
200	9	16.919	21.666	10.527
500	10	18.307	23.209	10.021

Из таблицы 2.10 следует, что объем исходной выборки не значительно влияет на качество вейвлет оценки и, следовательно, величину отклонения от истинной функции плотности распределения.

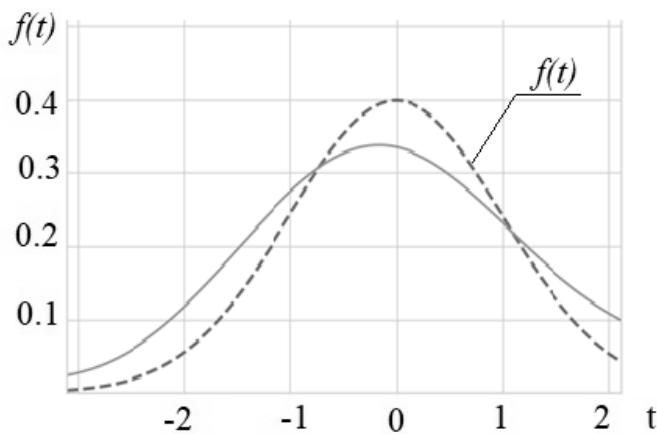
2.6 Исследование качества оценок функции плотности на основе неортогональных вейвлетов

Для сравнения качества восстановления функции плотности на основе неортогональных вейвлетов был выполнен ряд вычислительных экспериментов по аналогии с представленными выше для ортогональных вейвлетов.

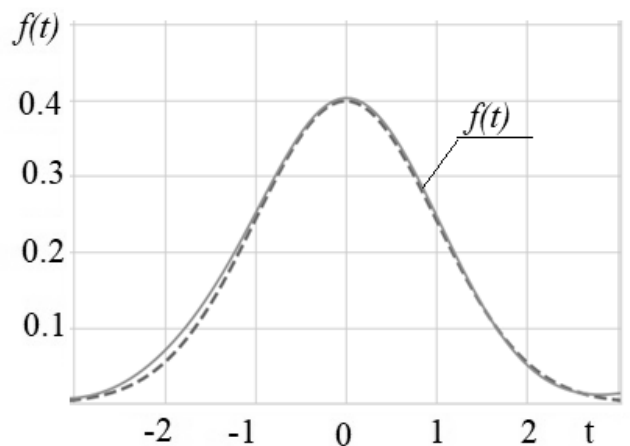
2.6.1 Вейвлет DOG

Прежде всего было проведено исследование качества восстановления функции плотности в зависимости от количества членов ряда N . Рассматривались выборки, смоделированные в разделе 2.5.1. Для этих выборок при каждом значении N , начиная с 5 и до 50, были построены вейвлет оценки функции плотности $\hat{f}_n(t)$ (2.1) по нормированным базисным функциям (2.23).

Результаты восстановления $f(t)$ при $N = 15$, $N = 34$ и $N = 45$ для стандартного нормального распределения представлены на рисунках 2.15, а–в соответственно. Вместе с полученными оценками на рисунке 2.15 представлен график истинной плотности распределения $f(t)$.



а) $N = 15$



б) $N = 34$

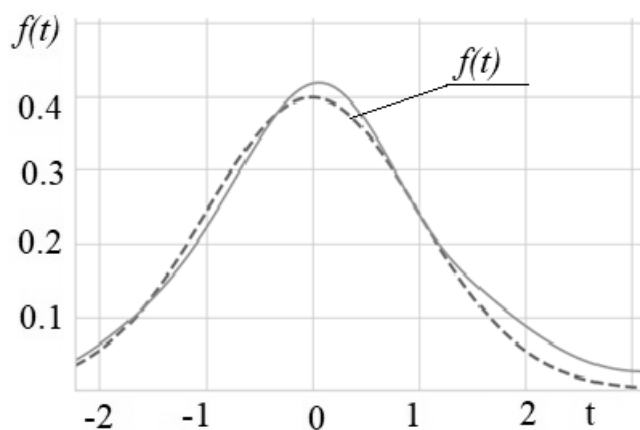
в) $N = 45$

Рисунок 2.15 – Оценка плотности распределения на основе вейвлета DOG для $N(0,1)$: а) $N = 15$; б) $N = 34$; в) $N = 45$

Отметим, что для числа членов ряда $N = 15$ имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности (см. рисунок 2.15, а). В случае $N = 34$ качество восстановления функции плотности наилучшее (см. рисунок 2.15, б). С ростом значения параметра сглаживания качество вейвлет-оценки функции плотности на основе вейвлета DOG ухудшается ($N = 45$) за счет увеличения величины отклонения от истинной функции плотности распределения (см. рисунок 2.15, в).

Результаты восстановления функции плотности для выборки, смоделированной по нормальному закону распределения с параметрами (6, 1.5), при $N = 15$, $N = 34$ и $N = 45$ приведены на рисунке 2.16, а–в соответственно. Вместе с полученными оценками на рисунках представлен график истинной плотности распределения $f(t)$. Для $N = 15$ имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности, аналогично результатам на рисунке 2.16, а. В случае $N = 34$ качество восстановления функции наилучшее, что согласуется с рисунком 2.18, б. С ростом значения параметра сглаживания величина отклонения $\hat{f}_n(t)$ от истинной функции плотности распределения увеличивается, а следовательно, качество вейвлет-оценки ухудшается (см. рисунок 2.16, в).

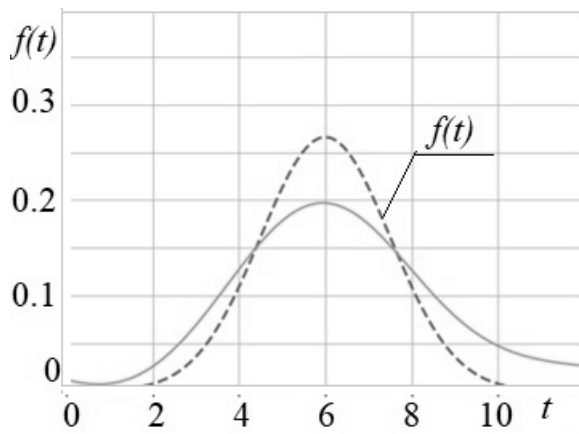
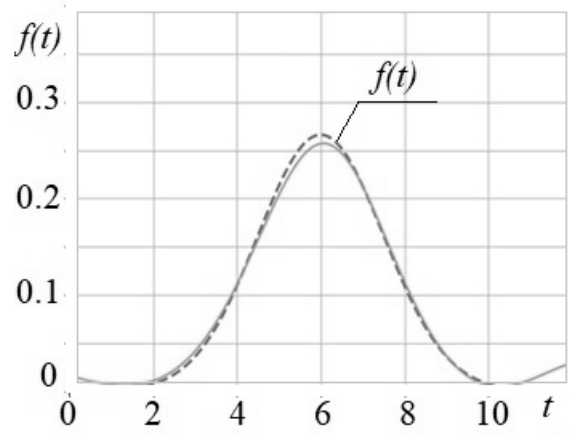
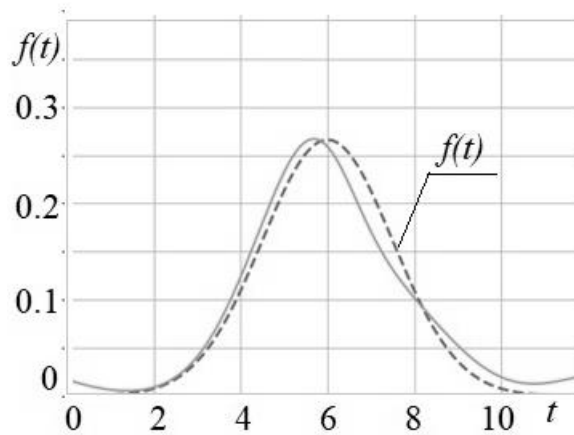
а) $N = 15$ б) $N = 34$ в) $N = 45$

Рисунок 2.16 – Оценка плотности распределения на основе вейвлета DOG

для $N(6, 1.5)$: а) $N = 15$; б) $N = 25$; в) $N = 34$

В ходе вычислительных экспериментов было показано, что наилучшее значение числа членов ряда для нормального распределения с параметрами $(0, 1)$ и с параметрами $(6, 1.5)$ не меняется.

На рисунке 2.17 а–в приведены результаты моделирования для экспоненциального закона распределения с параметром $\lambda = 1$. Отметим, что результаты, представленные на рисунке 2.19, качественно совпали с приведенными выше. Для $N = 15$ имеет место существенное отклонение оценки $\hat{f}_n(t)$ от истинной плотности (см. рисунок 2.17, а). В случае $N = 34$ качество восстановления функции наилучшее (см. рисунок 2.17, б). С увеличением значения параметра сглаживания N качество вейвлет-оценки ухудшается (см. рисунок 2.17, в). Появляются

дополнительные экстремумы, которые дают ложное представление о поведении истинной функции распределения.

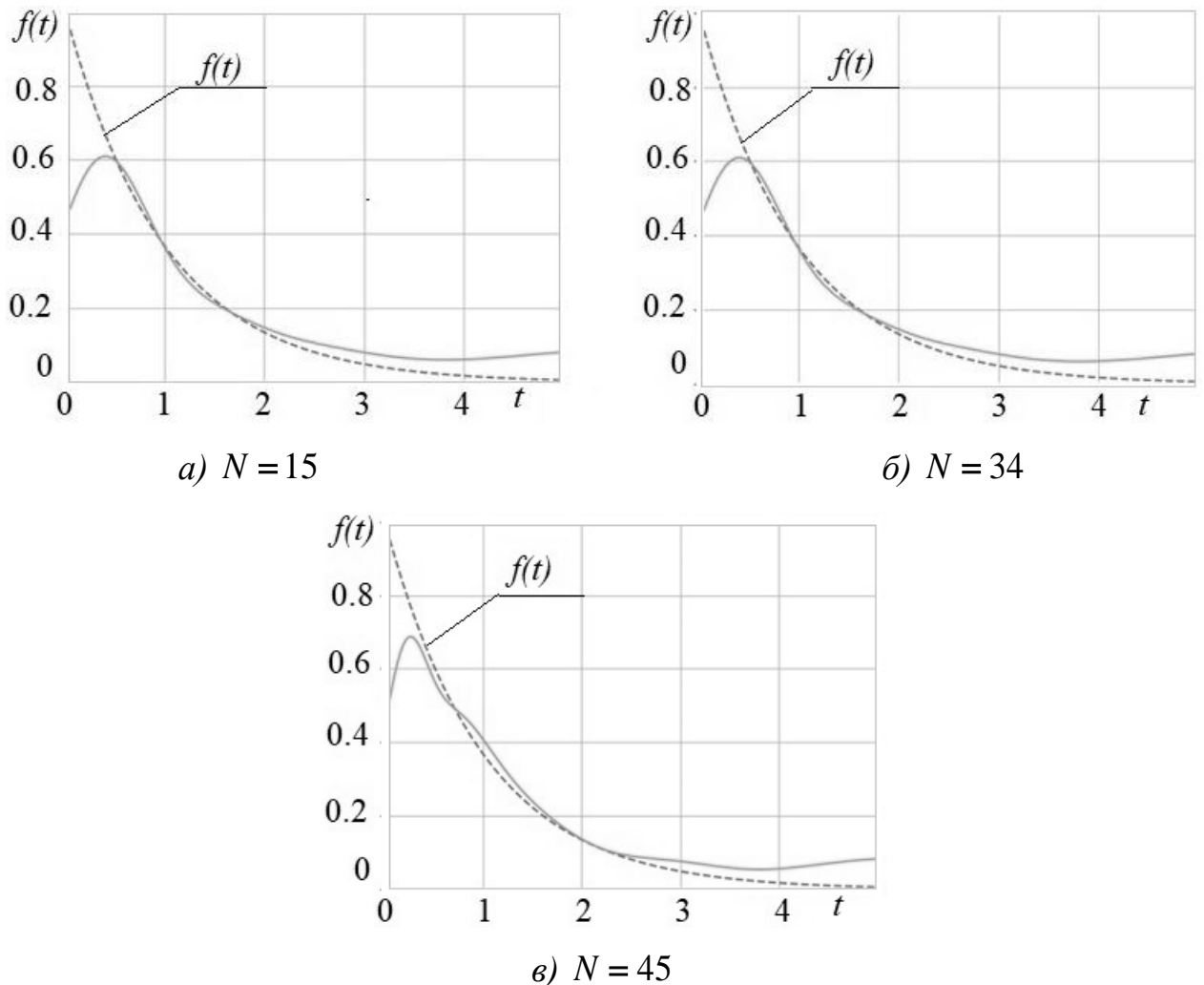


Рисунок 2.17 – Оценка плотности распределения на основе вейвлета DOG для экспоненциального закона распределения: а) $N = 15$; б) $N = 34$; в) $N = 45$

Для количественной оценки степени близости $\hat{f}_n(t)$ и $f(t)$ была проведена проверка согласия по критерию χ^2 . Критические значения критерия χ^2 при числе степеней свободы $\nu = 8$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены в разделе 2.4. Усредненные результаты значений статистики χ^2 по 1000 выборкам объема 200 представленные в таблице 2.11, показывают, что при $N = 34$ качество восстановления функции плотности наилучшее. Из таблицы 2.11 можно сделать

вывод, что наилучшим значением параметра сглаживания для вейвлет оценки, построенной на основе вейвлета DOG, является $N = 34$.

Таблица 2.11 – Значения статистики χ^2 в зависимости от показателя сглаживания для вейвлета DOG

Вид распределения	Статистика χ^2		
	$N = 15$	$N = 34$	$N = 45$
$N(0,1)$	16.995	8.779	25.44
$N(6,1.5)$	19.213	12.213	27.193
Экспоненциальное ($\lambda = 1$)	23.712	18.586	34.019

Аналогичные исследования с выборками объема $n = 100$ и $n = 500$ подтвердили это предположение. Наилучшее значение для неортогонального вейвлета DOG почти в 5 раз больше наилучших значений для ортогональных вейвлетов LITTLEWOOD & PALEY и Морле.

Было выполнено исследование качества восстановления функции плотности от объема выборки. Результаты статистического моделирования при $n = 100$ и $n = 500$ для выборок из независимых значений случайной величины ξ , распределенной по стандартному нормальному закону при наилучшем значении, $N = 34$ показаны на рисунках 2.18, *a* и *б* соответственно. Из рисунка 2.18 следует, что на качество вейвлет-оценки функции плотности влияет не только выбор значения параметра сглаживания N , но и бъем исходной выборки, так с увеличением объёма выборки качество оценки функции плотности распределения становится лучше.

Количественная характеристика зависимости качества вейвлет-оценки от объема выборки согласно критерию χ^2 приведены в таблица 2.12. Усреднение проводилось по 1000 вычислительным экспериментам для каждого из объемов

установлено, что наилучшее значение $\tilde{\zeta}$ для вейвлета DOG следует искать на интервале $[0.5, 1]$, поскольку случаев, когда наилучшее значение $\tilde{\zeta}$ находилось вне этого интервала, не наблюдалось. Для смоделированных выше выборок были построены вейвлет-оценки функции плотности на основе вейвлета DOG, где значение параметра $\tilde{\zeta}$ определялось на интервале $[0.5, 1]$ предложенными в разделе 2.3.2 способами. Результаты восстановления $f(t)$ для выборки объемом $n = 500$, состоящей из независимых значений случайной величины ξ , распределенной по нормальному закону с параметрами $(0, 1)$ с помощью вейвлета DOG при $N = 34$, представлены на рисунке 2.19.

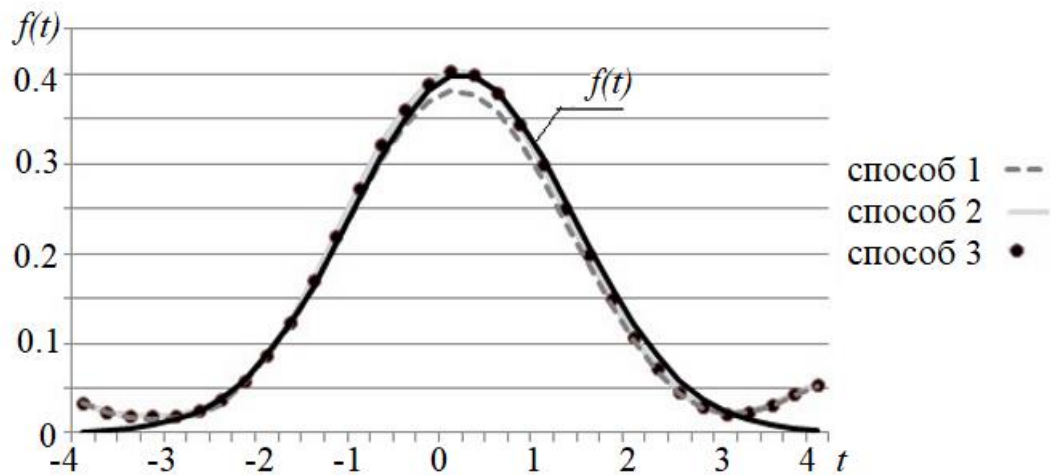


Рисунок 2.19 – Оценка плотности распределения на основе вейвлета DOG

Можно сделать вывод, что результаты восстановления функции плотности способом 1 и способом 2 визуально не отличаются друг от друга и практически совпадают с функцией плотности стандартного нормального распределения. Рисунок 2.19 показывает, что восстановленная на основе вейвлета DOG функция плотности с нормировочным коэффициентом $\tilde{\zeta}$, полученным способом 3, является наилучшей. Оценка количественной близости $\hat{f}_n(t)$ и $f(t)$ в зависимости от способа вычисления $\tilde{\zeta}$ для 1000 выборок объемом $n = 100$, $n = 200$ и $n = 500$ по критерию χ^2 , где критические значения $\chi^2_{кр}$ Пирсона при числе степеней свободы

$\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$, приведены в таблице 2.13. Результаты, представленные в таблице 2.13, показывают, что значения величины χ^2 намного меньше $\chi_{кр}^2$ для пятипроцентного уровня значимости. Следовательно, вейвлет оценка $\hat{f}_n(t)$ и теоретическая функция плотности $f(t)$ очень близки. Значение параметра \tilde{z} , полученное при восстановлении $\hat{f}_n(t)$ на основе вейвлета DOG, не совпало с $z = \left(\frac{3}{2}\sqrt{\pi} - \frac{2}{5}\sqrt{10\pi}\right)^{-\frac{1}{2}}$, полученным согласно утверждению 2.3.

Таблица 2.13 – Результаты расчета статистики χ^2 в зависимости от способа вычисления нормирующего множителя для вейвлета DOG

n	T	$\chi_{кр}^2$		Способ вычисления \tilde{z}	DOG	
		$\alpha \leq 0.05$	$\alpha \leq 0.01$		\tilde{z}	χ^2
100	8	14.067	18.475	1	0.7	9.023
				2	0.723	9.023
				3	0.722	9.023
200	9	16.919	21.666	1	0.7	8.779
				2	0.72	8.779
				3	0.72	8.779
500	10	18.307	23.209	1	0.711	8.516
				2	0.724	8.516
				3	0.723	8.516

Таким образом, проведенные вычислительные эксперименты подтверждают предположение о том, что свойство ортогональности материнского вейвлета влияет на качество восстановления функции плотности и поправки, которая учитывает только значение нормы недостаточно. Следует заметить, что при оценивании $\hat{f}_n(t)$ на основе вейвлета DOG наилучшее значение параметра \tilde{z} принадлежит

отрезку $[0.7, 0.73]$. Результаты моделирования для выборок объемом $n = 100$ и $n = 500$ качественно совпали с приведенными выше.

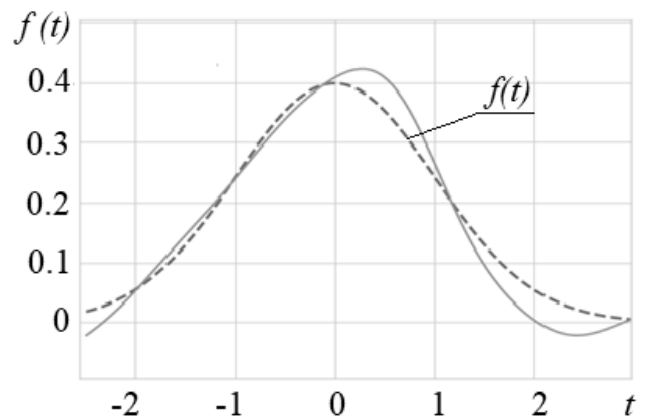
2.6.2 Вейвлет «Мексиканская шляпа»

По аналогии с исследованием, представленным в разделе 2.6.1, было выполнено исследование качества восстановления функции плотности на основе вейвлета «Мексиканская шляпа». Рассматривались выборки, смоделированные в разделе 2.5.1. Порядок исследования совпадал с представленным для вейвлета DOG. Оценки функции плотности (2.1) строились по нормированным базисным функциям (2.26).

Результаты восстановления $f(t)$ при $N = 8$, $N = 15$ и $N = 20$ для стандартного нормального распределения представлены на рисунках 2.20, *a–в* соответственно. Вместе с полученными оценкам на рисунке 2.20 представлен график истинной плотности распределения $f(t)$. Отметим, что для числа членов ряда $N = 5$, $N = 6, \dots, N = 9$ качество восстановления функции наилучшее (см. рисунке 2.20, *a*). С увеличением числа N ($N \geq 8$) качество оценки функции плотности ухудшается. Таким образом, из рисунка 2.20, *a–в* следует, что наилучшее значение параметра сглаживания будет $N = 5$.



a) $N = 8$



б) $N = 15$

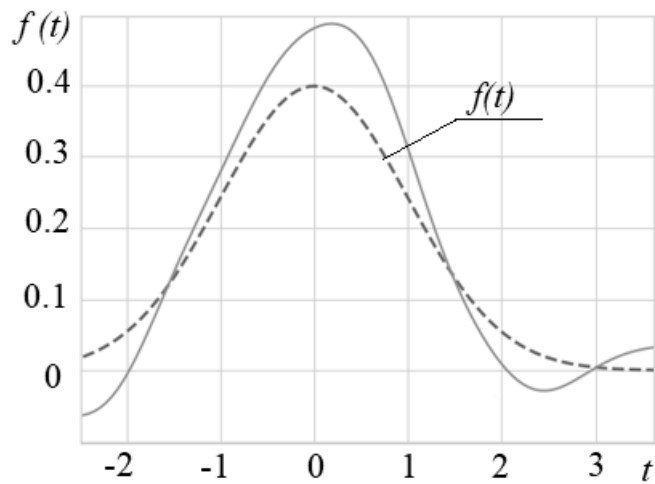
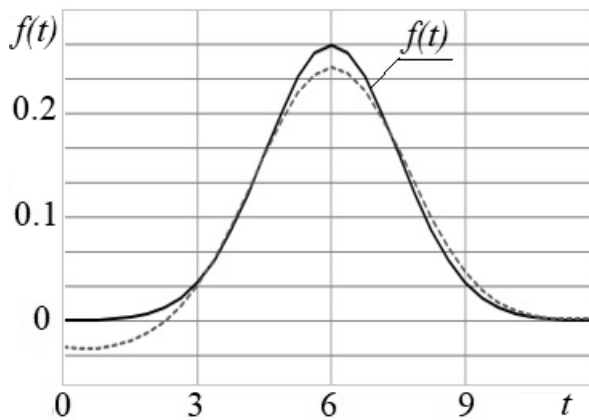
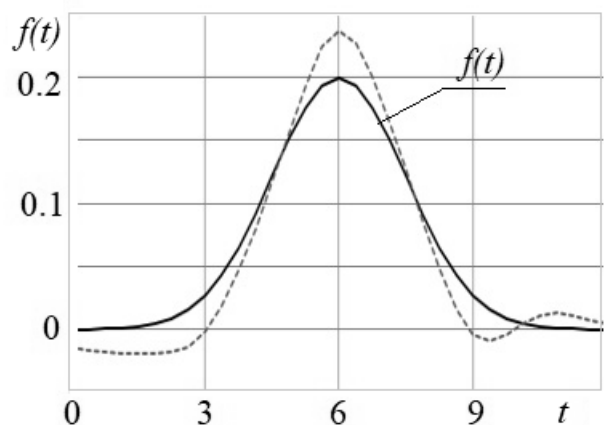
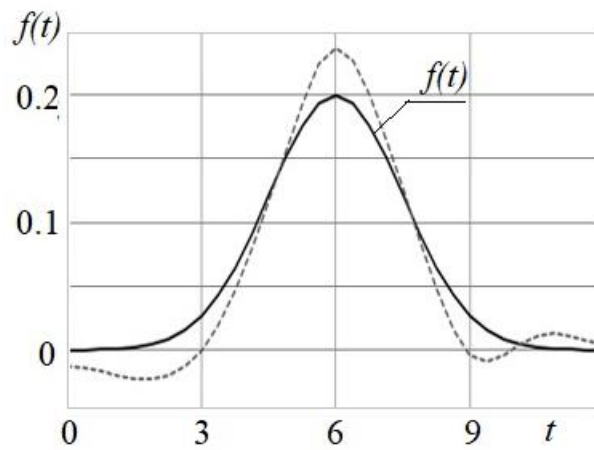
в) $N = 20$

Рисунок 2.20 - Оценка плотности распределения на основе вейвлета «Мексиканская шляпа» для $N(0, 1)$: а) $N = 8$; б) $N = 15$; в) $N = 20$

Результаты восстановления функции плотности для выборки, смоделированной по нормальному закону распределения с параметрами $(6, 1.5)$ приведены на рисунках 2.21 а–в соответственно. Для числа членов ряда $N = 8$ качество восстановления функции наилучшее (см. рисунок 2.21, а). При $N = 15$ и $N = 20$ качество вейвлет-оценки $f(t)$ на основе вейвлета «Мексиканская шляпа» становится хуже, за счет увеличения величины отклонения $\hat{f}_n(t)$ от истинной функции распределения (см. рисунок 2.21, б и в).

а) $N = 8$ б) $N = 15$

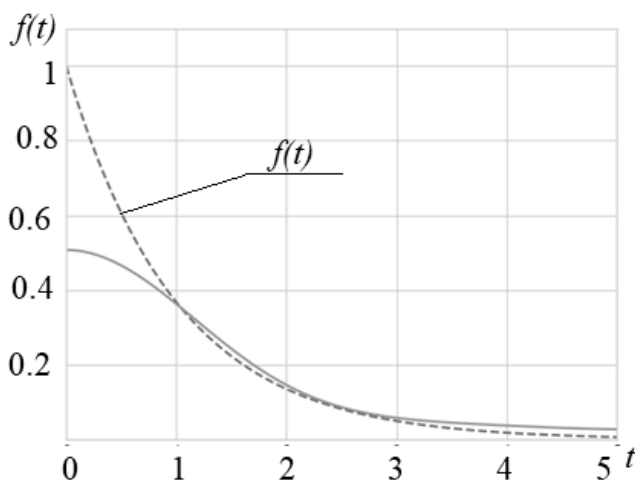


в) $N = 20$

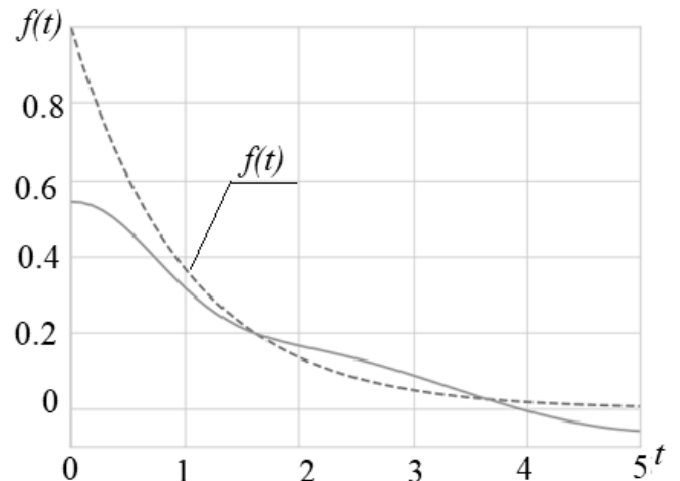
Рисунок 2.21 - Оценка плотности распределения на основе вейвлета «Мексиканская шляпа» для $N(6, 1.5)$: а) $N = 8$; б) $N = 15$; в) $N = 20$

Наилучшее значение числа членов, полученное в результате статистического моделирования, для нормального распределения с параметрами $(0, 1)$ и для нормального распределения с параметрами $(6, 1.5)$ совпало и равно $N = 8$. Отметим, что в редких случаях оценки $\hat{f}_n(t)$, построенные с помощью вейвлета «Мексиканская шляпа», на «хвостах» распределения принимают отрицательные значения.

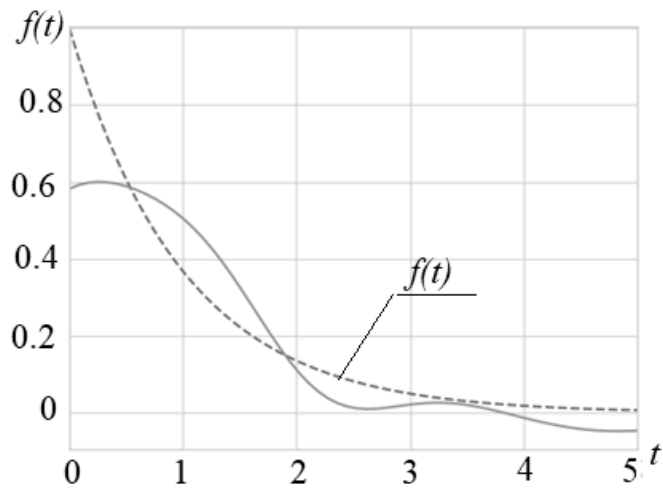
На рисунке 2.22, а–в приведены результаты моделирования для экспоненциального закона распределения с параметром $\lambda = 1$.



а) $N = 8$



б) $N = 15$



в) $N = 20$

Рисунок 2.22 – Оценка плотности распределения на основе вейвлета «Мексиканская шляпа» для экспоненциального закона распределения:

а) $N = 8$; б) $N = 15$; в) $N = 20$

Результаты, представленные на рисунке 2.22, качественно совпали с приведенными выше. Для числа членов ряда $N = 8$ вейвлет оценка функции плотности на основе вейвлета «Мексиканская шляпа» оказалась наилучшей (см. рисунок 2.22, а). С увеличением значения параметра сглаживания качество оценки функции плотности ухудшается, за счет роста величины отклонения $\hat{f}_n(t)$ от истинной функции плотности распределения (см. рисунок 2.24, б и в).

Из рисунков 2.20-2.22 видно, что значение параметра сглаживания N влияет на величину отклонения вейвлет оценки от плотности распределения. Наилучший результат восстановления функции плотности распределения соответствует значению $N = 8$. С увеличением параметра сглаживания качество вейвлет оценки ухудшается. Также были проведены исследования зависимости качества получаемой оценки от величины параметра сглаживания N с объемами выборки $n = 100$ и $n = 500$. Результаты качественно совпали с приведенными выше.

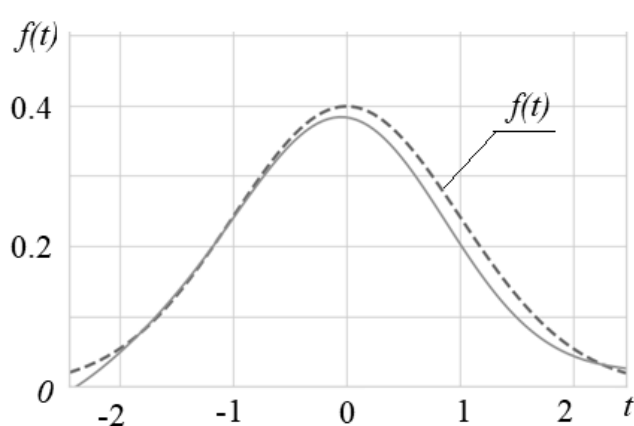
Усредненные значения статистики χ^2 по 1000 вычислительным экспериментам для выборок объемом 200 приведены в таблица 2.14. Критические значения критерия χ^2 при $\nu = 8$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены в разделе 2.4.

Таблица 2.14 – Значения статистики χ^2 в зависимости от показателя сглаживания для вейвлета «Мексиканская шляпа»

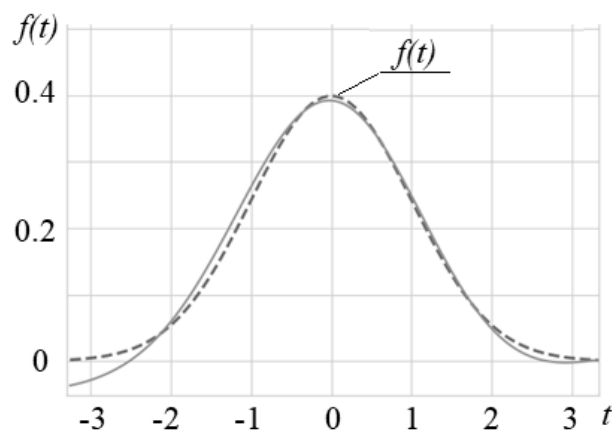
Вид распределения	Статистика χ^2		
	$N = 8$	$N = 15$	$N = 20$
$N(0, 1)$	9.229	13.852	20.249
$N(6, 1.5)$	14.934	17.404	27.024
Экспоненциальное ($\lambda = 1$)	17.694	24.112	29.443

Значения статистики χ^2 , представленные в таблице 2.14, показывают, что при $N = 8$ качество восстановления функции плотности наилучшее.

Также было выполнено исследование качества восстановления функции плотности от объема выборки. На рисунке 2.23 а), б) показаны оценки функции плотности при $N = 8$ для выборок объемом $n = 100$ и $n = 500$, состоящих из независимых значений случайной величины ξ , распределенной по стандартному нормальному закону.



а) $n = 100$



б) $n = 500$

Рисунок 2.23 – Оценка плотности распределения на основе вейвлета «Мексиканская шляпа»: а) $n = 100$; б) $n = 500$

Количественная характеристика зависимости качества вейвлет оценки от объема выборки согласно критерию χ^2 приведены в таблице 2.15. Усреднение проводилось по 1000 вычислительным экспериментам для каждого из объемов выборки. Критические значения критерия $\chi^2_{кр}$ при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены в таблице 2.15.

Таблица 2.15 – Значение статистики χ^2 в зависимости от объема выборки для вейвлета «Мексиканская шляпа»

n	T	$\chi^2_{кр}$		Статистика χ^2
		$\alpha \leq 0.05$	$\alpha \leq 0.01$	
100	8	15.507	20.09	10.071
200	9	16.919	21.666	9.829
500	10	18.307	23.209	9.416

Представленные результаты на рисунке 2.25 и значения статистики χ^2 из таблица 2.15, показывают, что объем исходной выборки не значительно влияет на качество вейвлет-оценки, построенной на основе вейвлета «Мексиканская шляпа».

Также было исследовано, влияние способа вычисления значения параметра \tilde{z} на качество аппроксимации функции плотности. На первом этапе было проведено дополнительное исследование для локализации интервала значений \tilde{z} , состоящее почти из 1000 вычислительных экспериментов [52, 53]. В результате было установлено, что наилучшее значение \tilde{z} для вейвлета «Мексиканская шляпа» следует искать на интервале $[0.5, 1]$, поскольку случаев, когда наилучшее значение \tilde{z} находилось вне этого интервала, не наблюдалось. Для смоделированных выше выборок были построены вейвлет-оценки функции плотности на основе вейвлета «Мексиканская шляпа», где значение параметра \tilde{z} определялось на интервале $[0.5, 1]$ предложенными выше способами. Результаты восстановления $f(t)$ для выборки объемом $n = 500$, состоящей из независимых значений случайной величины ξ , распределенной по нормальному закону с параметрам $(0, 1)$ с

помощью вейвлета «Мексиканская шляпа», представлены на рисунке 2.24. Из рисунка 2.24 можно сделать вывод, что результаты восстановления функции плотности способом 2 и способом 3 визуально не отличаются друг от друга и практически совпадают с функцией плотности стандартного нормального распределения.

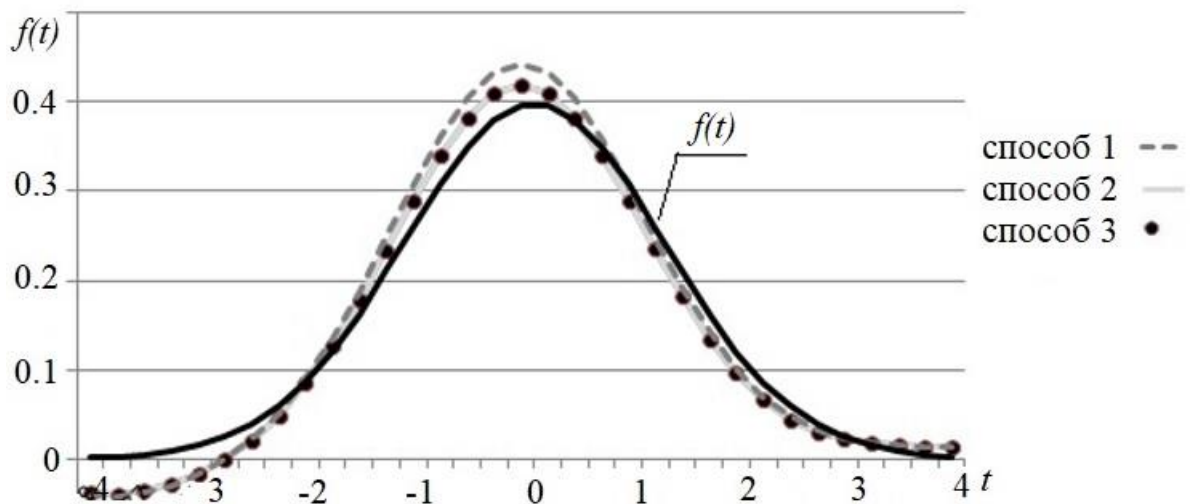


Рисунок 2.24 – Оценка плотности распределения на основе вейвлета DOG

Оценка количественной близости $\hat{f}_n(t)$ и $f(t)$ в зависимости от способа вычисления \tilde{z} для 1000 выборок объемом $n=100$, $n=200$, $n=500$ по критерию χ^2 , где критические значения $\chi^2_{кр}$ Пирсона при числе степеней свободы $\nu = T - 1$ для $\alpha \leq 0.05$ и $\alpha \leq 0.01$ приведены в таблице 2.16.

Результаты, представленные в табл. 2.16, показывают, что значения величины χ^2 намного меньше $\chi^2_{кр}$ для пятипроцентного уровня значимости. Следовательно, вейвлет оценка $\hat{f}_n(t)$, построенная на основе вейвлета «мексиканская шляпа» и теоретическая функция плотности $f(t)$ очень близки. Значение параметра \tilde{z} , полученное при восстановлении $\hat{f}_n(t)$ не совпало с параметром $z = \frac{4}{3\sqrt{\pi}}$, полученным согласно утверждению 2.4.

Таблица 2.16 – Результаты расчета статистики χ^2 в зависимости от способа вычисления нормирующего множителя для вейвлета «Мексиканская шляпа»

n	T	$\chi^2_{кр}$		Способ вычисления \tilde{z}	«Мексиканская шляпа»	
		$\alpha \leq 0.05$	$\alpha \leq 0.01$		\tilde{z}	χ^2
100	8	15.507	20.09	1	0.79	10.071
				2	0.799	10.072
				3	0.798	10.07
200	9	16.919	21.666	1	0.77	9.829
				2	0.774	9.828
				3	0.772	9.828
500	10	18.307	23.209	1	0.77	9.416
				2	0.779	9.416
				3	0.776	9.415

Таким образом, проведенные вычислительные эксперименты подтверждают предположение о том, что свойство ортогональности материнского вейвлета влияет на качество восстановления функции плотности и поправки, которая учитывает только значение нормы, недостаточно. Следует заметить, что при оценивании $\hat{f}_n(t)$ на основе вейвлета «Мексиканская шляпа» наилучшее значение параметра \tilde{z} принадлежит отрезку $[0.77, 0.8]$. Результаты моделирования для выборок объемом $n = 100$ и $n = 500$ качественно совпали с приведенными выше.

2.7 Сравнительный анализ вейвлет оценок плотности распределения

В разделе 2.5-2.6 для выборок объемом $n = 100$, $n = 200$ и $n = 500$ состоящих из множества независимых значений случайной величины ξ , распределенных по таким законам как стандартный нормальный закон, нормальный закон с параметрами (6,1.5) и показательный законом с параметром $\lambda = 1$ были построены

вейвлет оценки функции плотности с использованием вейвлетов Хаара, Морле, LITTLEWOOD & PALEY, DOG и «Мексиканская шляпа» при наилучших значениях параметра N . Вычислено значения статистики χ^2 с помощью соотношения (2.30), результаты приведены в таблицах 2.6, 2.8, 2.10, 2.12, 2.15. Критические значения критерия χ^2 Пирсона при числе степеней свободы $k = T - 1$ для уровней значимости $\alpha \leq 0.05$ и $\alpha \leq 0,01$ взяты из источника [25] и представлены в разделе 2.4.

Сравнивая, данные в таблицах 2.6, 2.8, 2.10, 2.12, 2.15 с критическими значениями, можно сделать вывод, что значения величины χ^2 меньше $\chi^2_{кр}$ как для пятипроцентного уровня значимости, так и для однопроцентного уровня значимости. Следовательно, вейвлет-оценка $\hat{f}_n(t)$ и теоретическая функция плотности $f(t)$ очень близки. Наилучшее восстановление функции плотности в различных условиях обеспечивается с помощью вейвлет оценки на основе материнского вейвлета DOG и вейвлета Хаара. Однако у вейвлета Хаара есть существенный недостаток: вейвлет оценка, построенная с его помощью, имеет ступенчатый вид, что делает невозможным применение вейвлета Хаара для оценивания неизвестных параметров регрессионных зависимостей. Подобные исследование были проведены также для выборок меньшего объема и значения χ^2 для исследуемых в работе вейвлетов, уместенно совпали с представленными в таблице.

Выводы по главе 2

Основные результаты, полученные в данной главе:

1. Построены оценки функции плотности распределения на основе преобразования Фурье.

2. Построены оценки функции плотности распределения на основе ортогональных вейвлетов, таких как вейвлет Хаара, LITTLEWOOD & PALEY Морле, и неортогональных вейвлетов, таких как DOG, «Мексиканская шляпа».

Кроме того, доказан ряд утверждений, позволяющих значительно улучшить качество восстановления функции плотности распределения.

3. Выполнено исследование качества оценок функции плотности в зависимости от параметра сглаживания и объема выборки при различных условиях формирования выборок. Даны рекомендации, позволяющие улучшить качество восстановления функции плотности. Также предложено соотношение, позволяющее вычислить значение параметра сглаживания.

4. Выполнен сравнительный анализ вейвлет-оценок плотности распределения и оценок на основе преобразования Фурье с помощью значения статистики χ^2 , на основании которого установлена наилучшая оценка плотности распределения.

Глава 3 Алгоритмы адаптивного оценивания параметров регрессионных моделей с использованием вейвлетов и преобразования Фурье

В этой главе предложен итерационный алгоритм оценивания параметров регрессионных моделей на основе преобразования Фурье и вейвлет-анализа. Представлены логарифмические функции правдоподобия на основе преобразования Фурье и ортогональных вейвлетов, таких как вейвлет LITTLEWOOD & PALEY и вейвлет Морле, а также неортогональных вейвлетов, таких как вейвлеты DOG и «Мексиканская шляпа». Приведены производные полученных логарифмических функций.

3.1 Алгоритм оценивания параметров регрессионных моделей

Рассмотрим задачу оценивания параметров регрессионного уравнения (1.28), для решения которой воспользуемся методом максимального правдоподобия [2, 22, 51, 52]. Будем считать, что уравнение регрессии (1.28) является истинным. Значения остатков $e_i = y_i - X_i \hat{\theta}$ являются независимыми случайными величинами с плотностью распределения $h(e_i, \theta)$, что следует из предположения о независимости случайных ошибок. Тогда логарифмическая функция правдоподобия принимает вид

$$\ln L(e_1, \dots, e_n, \hat{\theta}) = \ln \left(\prod_{i=1}^n h(e_i, \hat{\theta}) \right) = \sum_{i=1}^n \ln \left(h(e_i, \hat{\theta}) \right).$$

Для решения данной задачи была проведена модификация алгоритма адаптивного оценивания параметров на основе обобщённого лямбда-распределения [19, 63] из раздела 1.5.3.

Шаг 1. Определить начальное значение вектора неизвестных параметров $\hat{\theta}^0 = (\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_q^0)^T$ уравнения (1.28), при $l = 0$, где l – номер итерации. В данном

алгоритме предполагается использование метода наименьших квадратов [2, 16, 26, 51, 65] для определения начального приближения.

Шаг 2. Вычислить значения остатков $e_i = y_i - X_i \hat{\theta}^l$ регрессионной зависимости.

Шаг 3. Получить оценку функции плотности распределения с помощью соотношением (2.1) по различным нормированным базисным функциям.

Шаг 4. Определить значение логарифмической функции правдоподобия (1.24).

Шаг 5. Найти значение оценки вектора неизвестных параметров $\hat{\theta}^{l+1} = \arg \max_{\theta} L(e_1, \dots, e_n, \hat{\theta}^l)$.

Шаг 6. Итерационный процесс завершается, если $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| < \delta$, где δ – заданная погрешность вычисления. Если же $\|\hat{\theta}^{l+1} - \hat{\theta}^l\| > \delta$, то выполняется переход на шаг 2 при условии $l = l + 1$.

На шаге 3 алгоритма могут использоваться различные варианты материнских вейвлетов, что позволяет говорить о ряде новых алгоритмов оценивания неизвестных параметров регрессионных зависимостей. Построенные оценки функции плотности используются для определения логарифмической функции правдоподобия в рамках четвертого шага.

Отметим, что результаты второй главы показали в редких случаях возможное появление отрицательных значений на «хвостах» оценок функции плотности. Это ведет к тому, что вычисление логарифмической функции правдоподобия невозможно. Косвенным решением данной проблемы является использование качественного начального приближения, использование, которого не будет приводить к необходимости вычисления значения функции плотности на хвостах распределения.

Перейдем к рассмотрению конкретных вариантов оценивания параметров регрессионных моделей.

3.2 Оценивание параметров регрессионных моделей с использованием преобразования Фурье

Функция правдоподобия, определенная с помощью дискретного преобразования Фурье, принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}), \quad (3.1)$$

где $\tilde{\psi}_i(y_s - X_s \hat{\theta}), \tilde{\psi}_i(y_j - X_j \hat{\theta})$ определяются соотношением (2.10).

Логарифмическое представление (3.1) выражается соотношением:

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}) \right), \quad (3.2)$$

Тогда оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ можно найти при максимизации по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.2). Для максимизации функции (3.2) применялся градиентный метод [50]. Процесс решения задачи состоит в том, что, начиная с некоторого начального значения вектора неизвестных параметров $\hat{\theta}^0 = (\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_q^0)^T$, осуществляется последовательный переход с заданным шагом в направлении вектора градиента, состоящего из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\frac{\partial \ln(L(e_1, \dots, e_n, \hat{\theta}))}{\partial \hat{\theta}_r} = \sum_{s=1}^n \frac{1}{\sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta})} \times \left(\sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}'_{i_r}(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}) + \sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}'_{i_r}(y_j - X_j \hat{\theta}) \right) \quad (3.3)$$

Приведем схему решения градиентного метода:

Шаг 1. Определение $\hat{\theta}^0 = (\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_q^0)^T$ и $\text{grad}(\ln(L(e_1, \dots, e_n, \hat{\theta}^0)))$, $s = 0$,

где s – номер итерации.

Шаг 2. Выбор шага ρ .

Шаг 3. Определение следующей точки $\hat{\theta}^{s+1} = \hat{\theta}^s + \rho \cdot \text{grad}(\ln(L(e_1, \dots, e_n, \hat{\theta}^s)))$.

Шаг 4. Определение $\text{grad}(\ln(L(e_1, \dots, e_n, \hat{\theta}^{s+1})))$.

Шаг 5. Решение найдено, если $|\ln(L(e_1, \dots, e_n, \hat{\theta}^{s+1})) - \ln(L(e_1, \dots, e_n, \hat{\theta}^s))| < \varepsilon$, где ε – заданная погрешность вычисления. Если $|\ln(L(e_1, \dots, e_n, \hat{\theta}^{s+1})) - \ln(L(e_1, \dots, e_n, \hat{\theta}^s))| > \varepsilon$, то выполняется переход на шаг 2 при условии $s = s + 1$.

При использовании быстрого преобразования Фурье функция правдоподобия принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{2}{n^n} \prod_{s=1}^n \sum_{j=1}^{\frac{n-1}{2}} \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}), \quad (3.4)$$

где $\tilde{\psi}_i(y_s - X_s \hat{\theta}), \tilde{\psi}_i(y_j - X_j \hat{\theta})$ такие же, как в (3.1).

Логарифмическое представление (3.4) выражается соотношением:

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -2n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^{\frac{n-1}{2}} \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}) \right). \quad (3.5)$$

Оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ определяются путем максимизации по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.5), градиент которой состоит из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\frac{\partial \ln(L(e_1, \dots, e_n, \hat{\theta}))}{\partial \hat{\theta}_r} = \sum_{s=1}^m \frac{1}{\sum_{j=1}^{\frac{n-1}{2}} \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta})} \times \left(\sum_{j=1}^{\frac{n-1}{2}} \sum_{i=1}^N \tilde{\psi}'_{\hat{\theta}_r}(y_s - X_s \hat{\theta}) \tilde{\psi}_i(y_j - X_j \hat{\theta}) + \sum_{j=1}^n \sum_{i=1}^N \tilde{\psi}_i(y_s - X_s \hat{\theta}) \tilde{\psi}'_{\hat{\theta}_r}(y_j - X_j \hat{\theta}) \right) \quad (3.6)$$

Максимизация функции (3.5) выполнялась градиентным методом [50].

3.3 Оценивание параметров регрессионных моделей с использованием ортогональных вейвлетов

При решении задачи оценивания параметров регрессионного уравнения (1.28) на основе материнского вейвлета LITTLEWOOD & PALEY функция правдоподобия принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \cdot \frac{\sin 2\pi\tau_s - \sin \pi\tau_s}{\pi\tau_s} \frac{\sin 2\pi\tau_j - \sin \pi\tau_j}{\pi\tau_j}, \quad (3.7)$$

где $\tau_s = \frac{2^k}{d-c}(y_s - X_s \hat{\theta} - c) - (j-1)$. $\tau_j = \frac{2^k}{d-c}(y_j - X_j \hat{\theta} - c) - (j-1)$. а i, k, j такие же, как в (2.12).

Логарифмическое представление (3.7) выражается соотношением:

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \frac{\sin 2\pi\tau_s - \sin \pi\tau_s}{\pi\tau_s} \frac{\sin 2\pi\tau_j - \sin \pi\tau_j}{\pi\tau_j} \quad (3.8)$$

Оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ определяются путем максимизации по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.5), градиент которой состоит из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\begin{aligned} \frac{\partial \ln(L(e_1, \dots, e_n, \hat{\theta}))}{\partial \hat{\theta}_r} &= \sum_{s=1}^n \frac{1}{\sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \cdot \frac{\sin 2\pi\tau_s - \sin \pi\tau_s}{\pi\tau_s} \frac{\sin 2\pi\tau_j - \sin \pi\tau_j}{\pi\tau_j}} \cdot \sum_{j=1}^n \sum_{i=1}^N \frac{2^k}{d-c} \times \\ &\times \left[\left(\frac{(-2\pi\varphi_r(x_{sr}) \cos 2\pi\tau_s + \pi\varphi_r(x_{sr}) \cos \pi\tau_s) \pi\tau_s + \pi\varphi_r(x_{sr}) (\sin 2\pi\tau_s - \sin \pi\tau_s)}{(\pi\tau_s)^2} \right) \times \right. \\ &\quad \times \frac{\sin 2\pi\tau_j - \sin \pi\tau_j}{\pi\tau_j} + \frac{\sin 2\pi\tau_s - \sin \pi\tau_s}{\pi\tau_s} \times \\ &\quad \left. \times \left(\frac{(-2\pi\varphi_r(x_{jr}) \cos 2\pi\tau_j + \pi\varphi_r(x_{jr}) \cos \pi\tau_j) \pi\tau_j + \pi\varphi_r(x_{jr}) (\sin 2\pi\tau_j - \sin \pi\tau_j)}{(\pi\tau_j)^2} \right) \right]. \end{aligned} \quad (3.9)$$

Максимизация функции (3.8) выполнялась градиентным методом [49], описанным в разделе 3.2.

Для материнского вейвлета Морле функция правдоподобия принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} e^{-\frac{\tau_s^2}{2}} e^{-\frac{\tau_j^2}{2}} \cos 5\tau_s \cos 5\tau_j, \quad (3.10)$$

где τ_s, τ_j такие же, как в (3.7).

Логарифмическое представление (3.10) выражается соотношением

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} e^{-\frac{\tau_s^2}{2}} e^{-\frac{\tau_j^2}{2}} \cos 5\tau_s \cos 5\tau_j \right). \quad (3.11)$$

Оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ определяются максимизацией по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.11), градиент, которой состоит из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\begin{aligned} \frac{\partial \ln(L(e_1, \dots, e_n, \hat{\theta}))}{\partial \hat{\theta}_r} &= \sum_{s=1}^n \frac{1}{\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} e^{-\frac{\tau_s^2}{2}} e^{-\frac{\tau_j^2}{2}} \cos 5\tau_s \cos 5\tau_j} \times \\ &\times \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left[\left(\frac{\tau_s \varphi_r(x_{sr})}{2} e^{-\frac{\tau_s^2}{2}} \cos 5\tau_s + 5\varphi_r(x_{sr}) e^{-\frac{\tau_s^2}{2}} \sin 5\tau_s \right) e^{-\frac{\tau_j^2}{2}} \cos 5\tau_j + \right. \\ &\left. + \left(\frac{\tau_j \varphi_r(x_{jr})}{2} e^{-\frac{\tau_j^2}{2}} \cos 5\tau_j + 5\varphi_r(x_{jr}) e^{-\frac{\tau_j^2}{2}} \sin 5\tau_j \right) e^{-\frac{\tau_s^2}{2}} \cos 5\tau_s \right]. \end{aligned} \quad (3.12)$$

Максимизация функции (3.11) выполнялась градиентным методом [50], описанном выше.

3.4 Оценивание параметров регрессионных моделей с использованием неортогональных вейвлетов

При решении задачи оценивания параметров регрессионного уравнения (1.28) на основе вейвлета DOG функция правдоподобия принимает вид

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right). \quad (3.13)$$

где τ_s, τ_j такие же, как в (3.7).

Логарифмическое представление (3.13) выражается соотношением

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right) \right). \quad (3.14)$$

Оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ определяются путем максимизации по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.14), градиент которой состоит из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\begin{aligned} \frac{\partial \ln(L(e_1, \dots, e_n, \hat{\theta}))}{\partial \hat{\theta}_r} &= \sum_{s=1}^n \frac{1}{\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right)} \times \\ &\times \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left[\left(\tau_s \Phi_r(x_{sr}) \Phi(x_s) e^{\frac{-\tau_s^2}{2}} - \frac{1}{8} \tau_s \Phi_r(x_{sr}) e^{\frac{-\tau_s^2}{8}} \right) \left(e^{\frac{-\tau_j^2}{2}} - \frac{1}{2} e^{\frac{-\tau_j^2}{8}} \right) + \right. \\ &\left. + \left(\tau_j \Phi_r(x_{jr}) e^{\frac{-\tau_j^2}{2}} - \frac{1}{8} \tau_j \Phi_r(x_{jr}) e^{\frac{-\tau_j^2}{8}} \right) \left(e^{\frac{-\tau_s^2}{2}} - \frac{1}{2} e^{\frac{-\tau_s^2}{8}} \right) \right]. \quad (3.15) \end{aligned}$$

Максимизация функции (3.14) выполнялась градиентным методом [49], представленным в разделе 3.2.

Для материнского вейвлета «Мексиканская шляпа» функция правдоподобия принимает вид:

$$L(e_1, \dots, e_n, \hat{\theta}) = \frac{1}{n^n} \prod_{s=1}^n \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} (1 - \tau_s^2) e^{\frac{-\tau_s^2}{2}} (1 - \tau_j^2) e^{\frac{-\tau_j^2}{2}}, \quad (3.16)$$

где τ_s, τ_j такие же, как в (3.7).

Логарифмическое представление (3.16) выражается соотношением:

$$\ln(L(e_1, \dots, e_n, \hat{\theta})) = -n \ln n + \sum_{s=1}^n \ln \left(\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} (1-\tau_s^2) e^{-\frac{\tau_s^2}{2}} (1-\tau_j^2) e^{-\frac{\tau_j^2}{2}} \right). \quad (3.17)$$

Оценки неизвестных параметров $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ определяются путем максимизации по $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ логарифмической функции правдоподобия (3.16), градиент которой состоит из частных производных по каждой переменной $\theta_1, \theta_2, \dots, \theta_q$:

$$\begin{aligned} \frac{\partial \ln(L(y, \theta))}{\partial \theta_r} &= \sum_{s=1}^n \frac{1}{\sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} (1-\tau_s^2) e^{-\frac{\tau_s^2}{2}} (1-\tau_j^2) e^{-\frac{\tau_j^2}{2}}} \times \\ &\times \sum_{j=1}^n \sum_{i=1}^N \frac{z^2 2^k}{d-c} \left[\left(2\tau_s \varphi_r(x_{sr}) e^{-\frac{\tau_s^2}{2}} + \tau_s \varphi_r(x_{sr}) (1-\tau_s^2) e^{-\frac{\tau_s^2}{8}} \right) (1-\tau_j^2) e^{-\frac{\tau_j^2}{2}} + \right. \\ &\left. + \left(2\tau_j \varphi_r(x_{jr}) e^{-\frac{\tau_j^2}{2}} + \tau_j \varphi_r(x_{jr}) (1-\tau_j^2) e^{-\frac{\tau_j^2}{2}} \right) (1-\tau_s^2) e^{-\frac{\tau_s^2}{2}} \right]. \end{aligned} \quad (3.18)$$

Максимизация функции (3.17) выполнялась градиентным методом [50].

3.5 Исследование алгоритмов адаптивного оценивания параметров регрессионных моделей с использованием вейвлетов и преобразования Фурье

Исследуем предложенные выше алгоритмы оценивания неизвестных параметров θ регрессионной зависимости (1.28) на основе преобразования Фурье и различных материнских вейвлетов с помощью методов статистического моделирования. Рассмотрим регрессионную зависимость

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \varepsilon. \quad (3.19)$$

Отметим, что значения входного фактора x регрессионного уравнения (3.19) определялись из отрезка $[-4, 4]$, количество неизвестных параметров равно трем, а

их истинные значения $\theta_1 = 4$, $\theta_2 = -9$, $\theta_3 = 2$. Элементы вектора ошибок наблюдений ε_i моделировались независимыми с функцией распределения вида

$$F(x) = (1 - \mu)F_1(x, m_1, \sigma_1) + \mu F_2(x, m_2, \sigma_2),$$

где $F_i(x, m_i, \sigma_i)$ – функция нормального распределения с математическим ожиданием m_i и дисперсией σ_i^2 ; $i = 1, 2$. $\mu \in [0, 1]$ – параметр смеси.

При выполнении вычислительных экспериментов будем считать $m_1 = m_2 = 0$. Отметим, что моделирование ошибки выполняется при различной степени отклонения от нормального распределения. Доли наблюдений с дисперсиями σ_1^2 и σ_2^2 в выборке определяются с помощью параметра μ . где при $\mu = 0$ и $\mu = 1$ распределение ошибки будет нормальным. В процессе моделирования предполагалось, что $\sigma_2^2 \geq \sigma_1^2$, а значения дисперсий σ_1^2 и σ_2^2 задавались через значения уровня шума [24], который определяется соотношением

$$\rho = \frac{\delta}{\delta_c} \cdot 100\%,$$

где δ – дисперсия ошибки, $\delta_c^2 = (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$ – интенсивность сигнала.

Определение точности оценивания параметров выполнялось с помощью абсолютных отклонений оценок неизвестных параметров от истинных значений

$$v_1 = \frac{1}{R} \sum_{i=1}^R \sum_{j=1}^3 \frac{|\theta_j - \hat{\theta}_{ij}|}{|\theta_j|}, \quad (3.20)$$

где R – число вычислительных экспериментов, $\hat{\theta}_{ij}$ – оценка j -го параметра регрессионной зависимости (3.19) в i -ом вычислительном эксперименте.

Количество вычислительных экспериментов для различных комбинаций μ и ρ варьировалось от 100 до 500. Варьирование значение параметра смеси μ в диапазоне от 0 до 0.5 с шагом 0.05 позволило исследовать точность оценивания неизвестных параметров уравнения (3.19) при разной степени отклонения распределения случайной ошибки от нормального распределения. Моделирование

исходных данных выполнялось в соответствии с регрессионной зависимостью (3.19), а оценивание неизвестных параметров этой модели в соответствии с разработанным алгоритмом на основе таких материнских вейвлетов, как LITTLEWOOD&PALEY Морле, DOG, «Мексиканская шляпа». Кроме того, вычисление оценок выполнялось методом наименьших квадратов. Также для сравнения был использован адаптивный алгоритм оценивания параметров с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения [19]. Точность оценивания определялась значением показателя v_1 согласно соотношению (3.20). Результаты исследования точности оценивания неизвестных параметров уравнения (3.19) представлены на рисунке 3.1 для объема выборки 200, и на рисунке 3.2 для объема выборки 500 при фиксированном уровне шума $\rho_1=10\%$ и $\rho_2=100\%$. Вычисление нормы отклонения v_1 проводилось с усреднением по 100 вычислительным экспериментам.

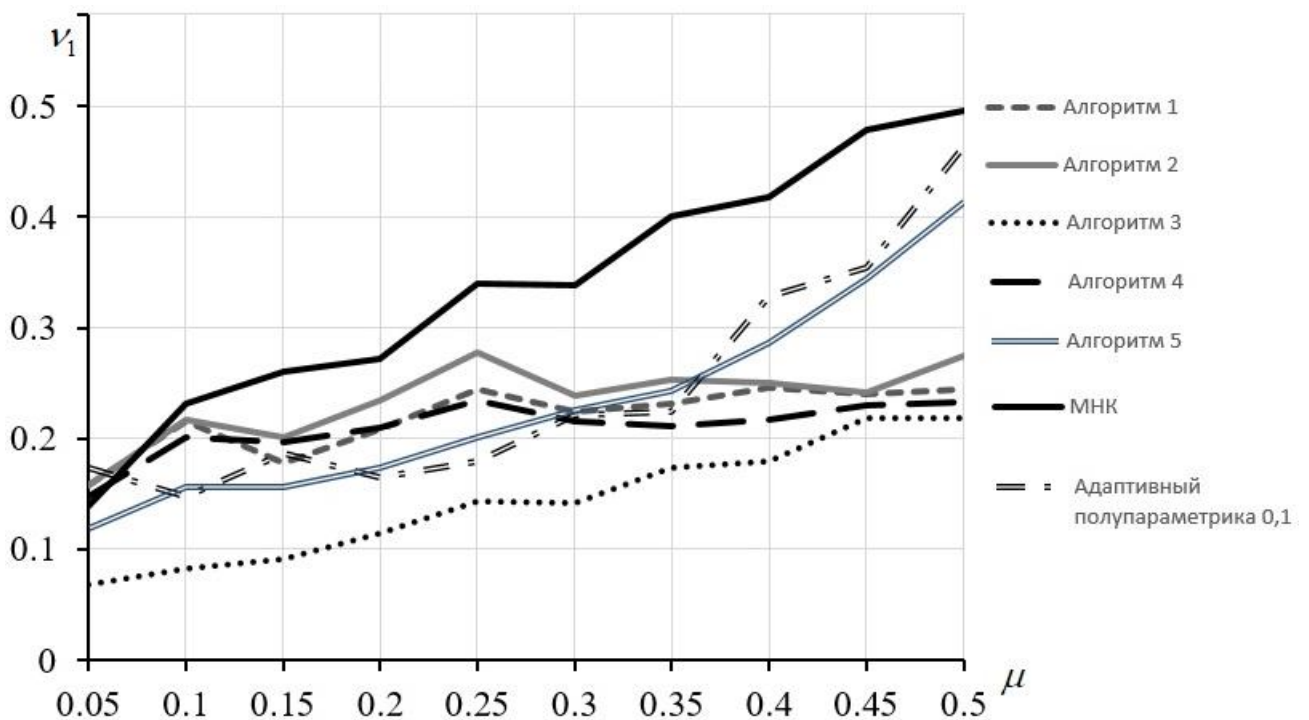


Рисунок 3.1 – Значение нормы отклонений в зависимости от μ ,
объем выборки – 200

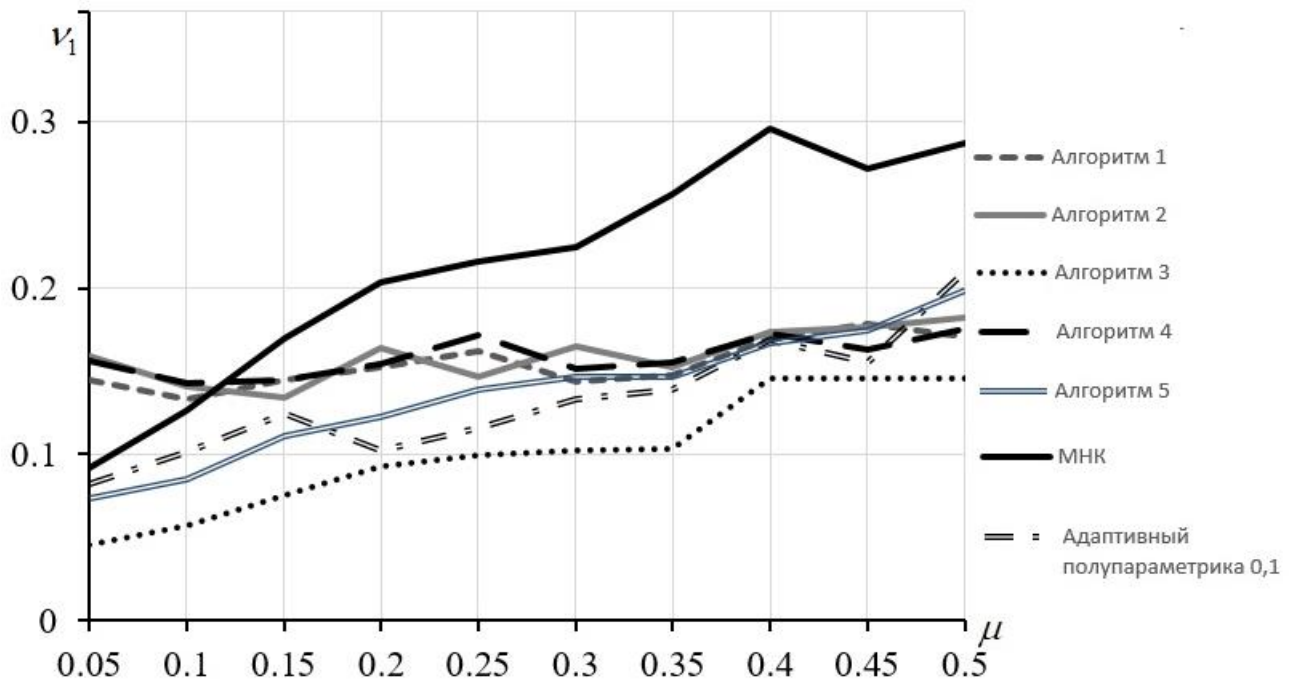


Рисунок 3.2 – Значение нормы отклонений в зависимости от μ ,
объем выборки – 500

Для удобства был обозначен адаптивный алгоритм на основе вейвлета LITTLEWOOD&PALEY - как алгоритм 1, в котором наилучшее значение параметра сглаживания установлено в разделе 2.5.2 и равно $N = 5$; алгоритм на основе вейвлета Морле - как алгоритм 2, в котором наилучшее значение параметра сглаживания установлено в разделе 2.5.3 и равно $N = 10$; адаптивный алгоритм на основе вейвлета DOG - как алгоритм 3, в котором наилучшее значение параметра сглаживания установлено в разделе 2.6.1 и равно $N = 34$; алгоритм на основе вейвлета «Мексиканская шляпа» - как алгоритм 4, в котором наилучшее значение параметра сглаживания установлено в разделе 2.6.2 и равно $N = 8$; адаптивный алгоритм на основе преобразования Фурье - как алгоритм 8, в котором наилучшее значение параметра сглаживания установлено в разделе 2.4 и равно $N = 8$.

Из рисунков 3.1 и 3.2 видно, что при значениях $\mu \in [0.05, 0.1]$ алгоритмы 1, 2, 4, 5 имеют не очень высокую точность, несколько уступая даже методу наименьших квадратов и методу с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.05$ и

объеме выборки 500 значение показателя v_1 , полученное в результате применения метода наименьших квадратов, на 40 % меньше значений v_1 , полученных при использовании алгоритмов 2 и 4, и на 36 % меньше значения v_1 , полученного при использовании алгоритма 1. Аналогичный результат при $\mu = 0.05$ и объеме выборки 500 наблюдается для значения v_1 , полученного при использовании метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения, 47 % и 42 % соответственно. Отметим, что для выборки объёмом 200 при $\mu = 0.05$ значения показателей v_1 , полученных в результате применения метода наименьших квадратов и метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения, меньше значений v_1 , полученных при использовании алгоритмов 1, 2 и 4 (см. рисунок 3.1). С ростом значений μ , то есть с ростом числа грубых ошибок наблюдений в исходных данных, алгоритмы на основе вейвлетов дают более точную оценку параметров регрессионной модели в сравнении с методом наименьших квадратов и с методом полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.5$ и объеме выборки 500 нормы отклонений v_1 , полученные при использовании алгоритмов 1 и 4, меньше на 43 % нормы отклонения v_1 , полученной при оценивании параметров методом наименьших квадратов, и на 24% меньше нормы отклонения v_1 , полученной при применении метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения. Значение нормы отклонения v_1 при $\mu = 0.5$, объеме выборки 500 и использовании алгоритма 2 на 39% меньше значения v_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 19% меньше нормы отклонения v_1 , полученной при применении метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения. Для выборки объёмом 200 при $\mu = 0.5$ значения показателей v_1 , полученных при использовании алгоритмов 1, 2,

4 и 5 меньше значений v_1 , полученных в результате применения метода наименьших квадратов и метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения (см. рисунок 3.1). Наилучшее качество оценивания неизвестных параметров уравнения (3.1) получается в результате использования в алгоритме вейвлета DOG (алгоритм 3) в качестве базисного вейвлета, причем как при малых значениях параметра μ , так и близким к 0.5. При $\mu = 0.05$ и объеме выборки 500 значение показателя v_1 , полученного при использовании алгоритма 3, на 44% меньше значения v_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 38% меньше нормы отклонения v_1 , полученной при применении метода с полупараметрическим восстановлением функции плотности на основе обобщенного лямбда-распределения. При $\mu = 0.5$ и объеме выборки 500 значение показателя v_1 , полученного при использовании алгоритма 3, на 46 % меньше значения v_1 , полученного при оценивании параметров регрессионной модели методом наименьших квадратов, и на 29 % меньше нормы отклонения v_1 , полученной при применении метода полупараметрического восстановления функции плотности на основе обобщенного лямбда-распределения. Таким образом, алгоритмы оценивания, основанные на таких материнских вейвлетах, как Морле, LITTLEWOOD&PALEY, DOG и «Мексиканская шляпа» и преобразовании Фурье показывают хорошую точность оценивания при разных условиях формирования выборок.

Выводы по главе 3

Основные результаты, полученные в данной главе:

1. Представлены логарифмические функции правдоподобия, полученные на основе преобразования Фурье; ортогональных вейвлетов, таких как LITTLEWOOD & PALEY и вейвлет Морле; неортогональных вейвлетов, таких как вейвлетов DOG и «Мексиканская шляпа».

2. Предложен ряд итерационных алгоритмов оценивания параметров регрессионных моделей на основе преобразования Фурье и вейвлет-анализа.

3. Определены выражения для производных, полученных логарифмических функций правдоподобия.

4. Выполнено исследование предложенных алгоритмов на основе преобразования Фурье и различных материнских вейвлетов для оценивания неизвестных параметров регрессионной зависимости.

Глава 4 Программная система адаптивного оценивания неизвестных параметров регрессионных моделей и восстановления функции плотности

В этой главе описываются общие сведения о программном комплексе оценивания параметров регрессионных моделей и восстановления функции плотности (WTiRM V1.0), разработанным автором, такие как функциональное назначение, логическая структура, технические характеристики. Также описан графический интерфейс, представлено алгоритмическое наполнение, описаны принципы работы с входными и выходными данными.

4.1 Структура программного комплекса

Программный комплекс для восстановления функции плотности и оценивания параметров регрессионных моделей на основе вейвлет-анализа (WTiRM V1.0) предназначен для решения следующих задач [52]:

- моделирование случайных величин, имеющих нормальное, засоренное нормальное и экспоненциальное распределение;
- нахождение оценок функции плотности на основе вейвлет преобразования;
- вычисление оценок функции плотности на основе преобразования Фурье;
- поиск оценок параметров регрессионных моделей методом наименьших квадратов;
- нахождение оценок параметров регрессионных моделей на основе вейвлет преобразования;
- нахождение оценок параметров регрессионных моделей на основе преобразования Фурье.

Для выполнения поставленных задач функционал программной системы был разделен на элементы, отвечающие за реализацию следующих функций:

- интерфейса программы, который обеспечивает взаимодействие с пользователем;
- оценивание функции плотности распределения и неизвестных параметров

регрессионной зависимости в соответствии с выбранным методом и выбранными параметрами;

- взаимодействие с файловой системой для обеспечения хранения результатов моделирования.

Интерфейс программной системы является интуитивно понятным и простым в освоении. Структура интерфейса программы обеспечивает возможность задействовать компоненты и модули программы при помощи минимального числа шагов, особенно для решения наиболее часто применяемых задач. Программный комплекс (WTiRM v1.0) был реализован на языке Python v 3.7.5 [86]. Этот язык обладает хорошим функционалом в области анализа данных и большим количеством статистических библиотек. Для реализации графического интерфейса был взят фреймворк разработки Qt Designer [97, 98]. Данный инструмент позволяет удобно создавать современные графические интерфейсы.

Для функционирования программы необходим персональный компьютер с операционной системой, соответствующей следующим требованиям:

- Минимальный объем оперативной памяти: 1ГБ;
- Минимальный объем свободного дискового пространства: 512МБ;
- Допускается выполнение .exe файлов.

Программный комплекс разработан с использованием модульного программирования и состоит из главного модуля; модуля оценивания функции плотности на основе вейвлет-анализа; модуля оценивания функции плотности с использованием преобразования Фурье; модуля оценивания параметров регрессии. Модули выполнены как самостоятельные части программы, обеспечивающие выполнение заданных функций обработки автономно от других программных модулей. Рассмотрим разработанные программные модули:

- модуль main является основным и формирует необходимые экранные формы интерфейса;

- модуль wavelets (см. таблица 4.1) содержит функции, получающие значения ортонормированного базиса на произвольном отрезке $[c, d]$ с

использованием вейвлет-анализа; преобразование Фурье; вычисления вейвлет оценки функции плотности.

Таблица 4.1 - Функции модуля wavelets

Функция	Описание
MHWave	Построение ортонормированного базиса на основе материнского вейвлета «Мексиканская шляпа».
MorleWave	Построение ортонормированного базиса на основе материнского вейвлета Морле.
DOGWave	Построение ортонормированного базиса на основе материнского вейвлета DOG.
LPWave	Построение ортонормированного базиса на основе материнского вейвлета LITTLEWOOD & PALEY.
Wavelet_Estimate	Вычисление вейвлет оценки функции плотности.
Golden_Section_Method	Реализация метода поиска оптимального значения коэффициента z , путем минимизации статистики χ^2 .

Модуль FourierTransform (см. таблица 4.2) – это модуль в котором реализованы алгоритмы восстановления функции плотности на основе вейвлет-анализа.

Таблица 4.2 - Функции модуля FourierTransform.

Функция	Описание
DFF	Вычисляет оценку функции плотности с использованием дискретного преобразования Фурье.
BFF	Вычисляет оценку функции плотности с использованием быстрого преобразования Фурье.

Модуль MLE (см. таблица 4.3) – это модуль, в котором реализован алгоритм адаптивного оценивания параметров регрессионных моделей.

Таблица 4.3 - Функции модуля MLE

Функция	Описание
algFT	Реализует алгоритм адаптивного оценивания параметров регрессионных моделей.
funcReg	Оценка значений отклика (1.28).

Для удобной работы в программном комплексе для восстановления функции плотности и оценивания параметров регрессионных моделей на основе различных ортонормированных базисов (WTiRM V1.0) пользователю предоставлено 3 режима работы. В каждом режиме доступен свой набор функций, который был определен в ходе разработки комплекса WTiRM V1.0. Описание функций представлено в таблице 4.4 функции.

Таблица 4.4 – Режимы программного комплекса

Режим	Описание
Оценивание функции плотности на основе вейвлет-анализа	Ввод данных для моделирования случайной выборки и запись смоделированной выборки в файл; выбор готового файла, содержащего выборку со случайными величинами; выбор метода оценивания функции плотности и ввод параметров; запуск алгоритма оценивания функции плотности.
Оценивание функции плотности на основе преобразования Фурье	Ввод данных для моделирования случайной выборки и запись смоделированной выборки в файл; выбор готового файла, содержащего выборку со случайными величинами; выбор метода оценивания функции плотности и ввод параметров; запуск алгоритма оценивания функции плотности.

Таблица 4.4 – продолжение

Оценивание параметров регрессии	Ввод данных для моделирования случайной выборки и запись смоделированной выборки в файл; выбор готового файла, содержащего выборку с регрессионной моделью; выбор методов оценивания параметров регрессионных моделей и ввод параметров; запуск алгоритма оценивания неизвестных параметров регрессионной модели.
---------------------------------	---

Модули, необходимые для функционирования интерфейса:

- `ui_main` – это модуль верстки интерфейса, разработанного в Qt Designer [96, 97], конвертированный из формата `ui`. Модуль реализован с помощью GUI фреймворков (`graphical user interface`), [97, 98]. При помощи `PySide2`, [97, 98] осуществлена привязка языка Python к инструментарию Qt [86, 97, 98];

- `ui_functions` представляет собой класс `UiFunctions`, в котором реализованы функции для реализации динамической части интерфейса;

- `ui_styles` – это модуль стилей интерфейса.

Модуль `show` - это модуль для построения графиков. Построение графиков реализовано с использованием пакета `Matplotlib`.

Рассмотрим далее более подробно каждый из режимов таблицы 4.4 с учетом особенностей работы программной системы.

4.2 Режим работы программного комплекса

При запуске программного комплекса открывается приветственное окно, в левой части которого располагается скрытое меню. В верхнем левом углу расположена кнопка, при нажатии на которую появляются доступные режимы работы. С помощью данного меню пользователь может переключаться с одного режима работы программного комплекса на другой. Выбор режима влияет на вид рабочей области окна (см. рисунок 4.1).

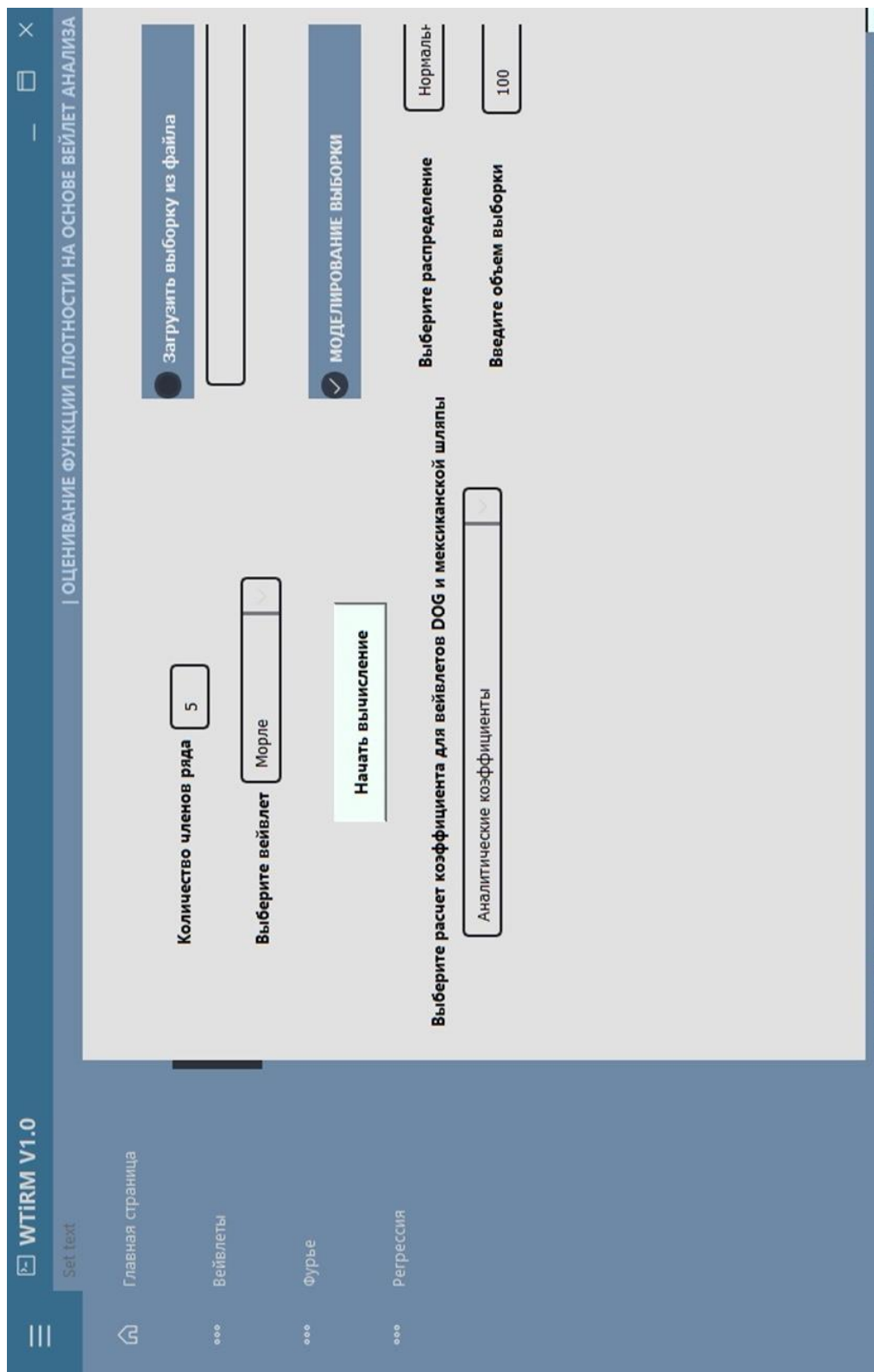


Рисунок 4.1 – Содержимое левого меню

В меню имеются следующие пункты, которые соответствуют режимам работы программного комплекса:

- пункт меню «Главная страница» содержит информация о программном комплексе, а также справочную информацию и инструкции для пользователя;
- пункт меню «Вейвлеты» включает переход на страницы оценивания функции плотности с помощью вейвлетов;
- пункт меню «Фурье» включает переход на страницы оценивание функции плотности с помощью ряда Фурье;
- пункт меню «Регрессия» включает оценивание параметров регрессионных моделей с использованием вейвлетов.

4.2.1 Восстановление функции плотности распределения

4.2.1.1 Режим «Оценивание функции плотности на основе преобразования Фурье»

При выборе режима оценивания функции плотности на основе преобразования Фурье пользователю необходимо выбрать вкладку «Фурье» в левом меню программы. В результате откроется диалоговое окно, вид которого представлен на рисунке А.1 (см. приложение А). Здесь пользователь может выбирать необходимые ему методы и данные для работы алгоритма оценивания функции плотности на основе преобразования Фурье. Для оценивания функции плотности с использованием Фурье пользователю предлагается ввести значение параметра сглаживания и выбрать, с каким преобразованием будет запущен алгоритм (быстрое преобразование или дискретное преобразование).

В случае загрузки готовых исходных данных пользователю следует отметить галочкой раздел «Загрузить выборку из файла» и выбрать кнопку «Загрузить файл». Загрузить готовую выборку необходимо в формате .txt.

У пользователя есть возможность смоделировать выборку самостоятельно. Для этого пользователю необходимо отметить галочкой «Моделирование

выборки», затем выбрать распределение из предложенных вариантов, а также ввести объем моделируемой выборки. Для того чтобы запустить алгоритм оценивания функции плотности с использованием преобразования Фурье, необходимо ввести количество членов ряда, а также загрузить или смоделировать входную выборку и только затем нажать кнопку «Начать вычисление». Один из способов представления результатов работы алгоритма показана на рисунке 4.2.

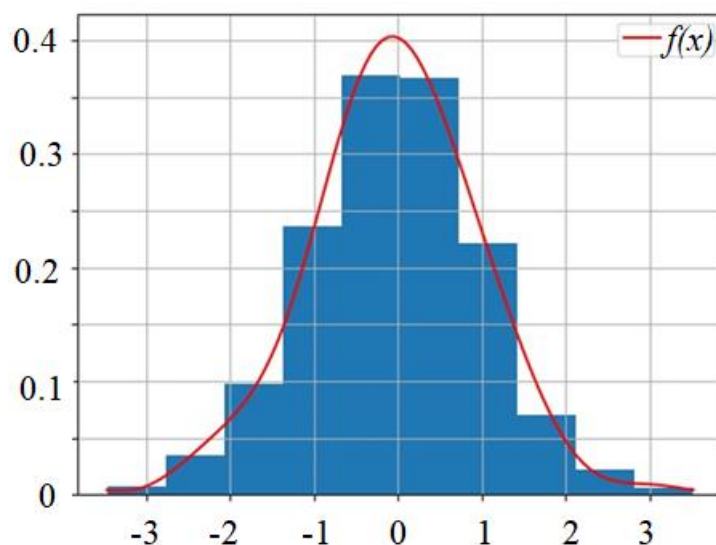


Рисунок 4.2 – Гистограмма исходной выборки и оценка функции плотности на основе преобразования Фурье

Также для анализа работы алгоритмов результаты выводятся в txt файл.

4.2.1.2 Режим «Оценивание функции плотности на основе вейвлет-анализа»

При выборе режима оценивания функции плотности на основе вейвлетов пользователю необходимо выбрать вкладку «Вейвлеты» в левом меню программы. В результате откроется диалоговое окно, вид которого показан на рисунке А.2 (см. приложение А). Здесь пользователь может выбирать необходимые ему методы оценивания и данные для корректной работы алгоритма на основе вейвлет-анализа.

Пользователь должен ввести необходимое количество членов ряда и выбрать материнский вейвлет (один или несколько), с использованием которого будет запущен алгоритм оценивания функции плотности распределения: LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа». Отметим, что пользователь может выбрать метод вычисления нормировочного коэффициента для не ортонормированных вейвлетов DOG и «Мексиканская шляпа». Также у пользователя есть возможность смоделировать выборку самостоятельно. Для моделирования выборки необходимо отметить галочкой «Моделирование выборки», затем выбрать тип распределения случайной величины из предложенных вариантов, а также ввести объем генерируемой выборки. Для того чтобы запустить алгоритм оценивания функции плотности на основе вейвлет-анализа необходимо ввести количество членов ряда и выбрать вейвлет из предложенных, при необходимости выбрать способ расчета нормировочного множителя. Загрузить или смоделировать входную выборку и только затем нажать кнопку «Начать вычисление». На рисунке 4.3 представлен один из способов представления результата работы алгоритмов.

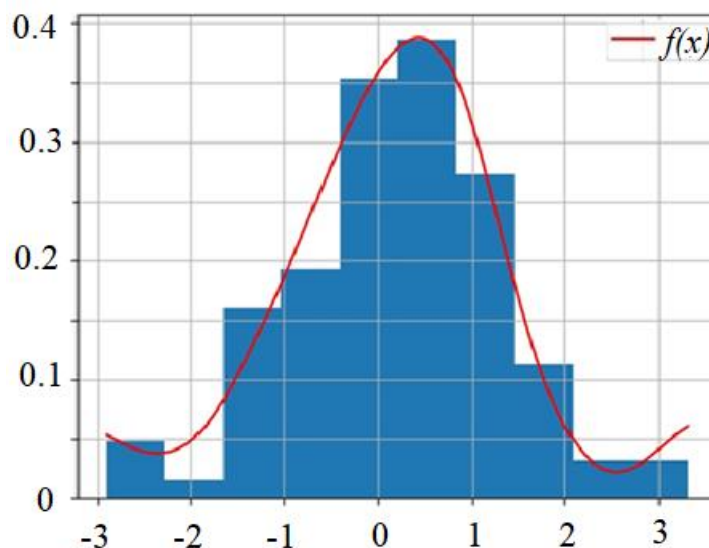


Рисунок 4.3 – Гистограмма исходной выборки и оценка функции плотности на основе вейвлет анализа

Также результаты работы алгоритмов выводятся в txt файл для дальнейшего анализа пользователем.

4.2.2 Оценивание параметров регрессионных моделей

При выборе режима адаптивного оценивания параметров регрессии пользователю необходимо выбрать вкладку «Регрессия» в меню программы расположенном слева. В результате откроется диалоговое окно, вид которого показан на рисунке А.3 (см. приложение А). Здесь пользователь может выбирать необходимые ему методы и данные для корректной работы алгоритма оценивания регрессионных моделей.

4.2.2.1 Режим моделирования

В случае, если данные необходимо смоделировать, предлагается ввод данных в разделе «Моделирование выборки». Описываемая область выделена на рисунке А.4 (см. приложение А). Здесь пользователь может сам задать процент доли выбросов (максимальное допустимое значение 50 %), процент уровня выбросов (минимальное допустимое значение 50 %), процент уровня шума (максимальное допустимое значение 50 %). объем генерируемых выборок и количество выборок. Пользователю доступно выбрать распределение случайных величин генерируемого шума.

Для генерирования значений функций и отклика регрессионной модели необходимо загрузить файл с истинными параметрами регрессии: при нажатии кнопки “Загрузить файл” откроется проводник, в котором пользователю следует выбрать нужный файл. Он должен содержать вектор параметров $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$ через пробел в одну строку: $[\theta_1 \theta_2 \dots \theta_q]$. Программа считывает содержимое файла и автоматически получает размерность вектора. Истинные значения параметров регрессионной модели необходимы, чтобы смоделировать исходные данные.

Смоделированная выборка в упомянутом выше матричном формате сохраняется в txt файл, где в названии указывается вид распределения случайной величины, уровень шума, доля выбросов и уровень выбросов. Данные генерируются в соответствии с моделью (1.28).

Перед запуском алгоритма оценивания параметров регрессионной модели пользователь должен выбрать метод (один или несколько), результаты которого хочет наблюдать, а также задать параметр сглаживания, соответствующий методу. Описываемая область выделена на рисунке А.5 (см приложение А). Методы оценивания, доступные пользователю, базируются на преобразованиях Фурье, вейвлет-анализе, где в качестве материнского вейвлета представлены вейвлеты: LITTLEWOOD & PALEY; Морле; DOG; «Мексиканская шляпа».

Для запуска алгоритмов оценивания регрессионных параметров необходимо выбрать кнопку «Начать вычисления». При корректном вводе данных откроется проводник, в котором пользователю необходимо выбрать папку для записи результирующих файлов, содержащих сгенерированную модель и файлов, содержащих результаты работы алгоритма. Далее программа запустит вычисления. В случае некорректных действий программа сообщит о возникших ошибках, указав, что именно мешает запуску.

Результатом работы режима адаптивного оценивания параметров регрессии является набор данных, хранящихся в соответствующих файлах:

- файл, содержащий восстановленные параметры регрессии;
- файлы, содержащие вектор остатков регрессионной модели для найденных параметров каждого обработанного метода;
- файл, содержащий смоделированную выборку данны в упомянутом выше матричном формате.

Продолжая работу с программой, пользователь может, изменив параметры либо не меняя настройки, снова запустить вычисления. Программный комплекс позволяет визуализировать полученные результаты в разделах оценивания функции плотности.

4.2.2.2 Режим оценивания параметров

В случае загрузки готовых исходных данных пользователю следует отметить галочкой раздел «Загрузить выборку из файла» и выбрать кнопку «Загрузить файл». Описываемая область выделена на рисунке А.6 (см. приложение А). В этом случае откроется проводник, в котором пользователю следует выбрать файл, содержащий исходные данные регрессионной модели вида (1.28). Содержимое файла для загрузки должно выглядеть так:

$$\begin{bmatrix} y_1 & \dots & y_n \\ \varphi_1(x_{11}) & \dots & \varphi_1(x_{n1}) \\ \dots & \dots & \dots \\ \varphi_q(x_{1q}) & \dots & \varphi_q(x_{nq}) \end{bmatrix}.$$

Программа считывает данные и автоматически получает размер выборки и количество неизвестных регрессионных параметров. Перед запуском алгоритма оценивания параметров регрессионной модели пользователь должен выбрать галочками метод (один или несколько), результаты которого хочет наблюдать пользователь, а также задать параметр сглаживания, соответствующий методу. Описываемая область выделена на рисунке А.7 (см. приложение А). Методы оценивания, доступные пользователю, базируются на преобразованиях Фурье, вейвлет-анализе, где в качестве материнского вейвлета представлены вейвлеты: LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа». Для запуска алгоритмов оценивания регрессионных параметров необходимо выбрать кнопку «Начать вычисления». При корректном вводе данных откроется проводник, где пользователю необходимо выбрать папку, в которой будет выполняться запись результирующих файлов, содержащих сгенерированную модель, и файлы, содержащие результаты работы алгоритма. Далее программа запустит вычисления. В случае некорректных действий программа сообщит о возникших ошибках, указав, что именно мешает запуску.

Результатом работы режима адаптивного оценивания параметров регрессии является набор данных, хранящихся в соответствующих файлах:

- файл, содержащий восстановленные параметры регрессии;
- файлы, содержащие вектор остатков регрессионной модели для найденных параметров каждого обработанного метода.

После выполнения оценивания параметров регрессионной зависимости работа в программном комплексе может продолжаться с изменением параметров и настроек или без изменения.

Выводы по главе 4

Основные результаты, полученные в данной главе:

1. Разработан программный комплекс (WTiRM v1.0), который может быть использован для определения неизвестных параметров регрессионных зависимостей и оценивания функции плотности распределения.

2. Представлено описание интерфейса и функциональных возможностей программного комплекса, рассмотрены разработанные программные модули.

3. Приводятся минимальные системные требования, показаны режимы работы программного комплекса при восстановлении функции плотности и при оценивании неизвестных параметров регрессионных зависимостей.

4. Реализованный программный комплекс был зарегистрирован в виде объекта интеллектуальной собственности как программа для ЭВМ (№ гос. рег. № 2022613412 от 14 марта 2022 г) [49].

Глава 5 Прогноз влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений

В данной главе показано применение алгоритмов адаптивного оценивания параметров регрессионных зависимостей с помощью преобразований Фурье и вейвлет-анализа при прогнозе влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений на основе показаний многозонных цифровых датчиков. Построена регрессионная модель температурных данных по результатам измерения в отдельных точках на сваях, где установлены многофазные цифровые датчики, и на основе предложенных алгоритмов получены оценки параметров регрессионной зависимости. Приведены результаты температурных изменений, полученных на основе предложенной регрессионной зависимости.

5.1. Применение методов оценивания регрессионных зависимостей при прогнозировании температурных изменений

Мерзлота занимает не менее 25 % площади всей суши земного шара. В России 65 % территории составляют районы многолетней мерзлоты. Наиболее широко распространена вечная мерзлота в Восточной Сибири и Забайкалье. Самый глубокий предел многолетней мерзлоты отмечается в верховьях реки Вилюй в Якутии. Рекордная глубина залегания многолетней мерзлоты составила 1370 м.

В связи с возрастанием темпов освоения этих регионов вопросы повышения устойчивости, надежности и долговечности сооружений, возводимых в сложных природно-климатических и инженерно-геологических условиях, приобретают большое практическое значение. Фундаменты сооружений в сезоннопромерзающих грунтах основания подвергаются воздействию процессов: промерзания, морозного пучения, оттаивания и вибродинамики. Эти процессы особенно интенсивно проявляются в период сезонного оттаивания и промерзания.

Фундаменты вертикальных стержневых элементов – это относительно легкие сооружения, которые наиболее подвержены этим негативным явлениям.

Поэтому для проектирования, строительства, реконструкции и эксплуатации сооружений, возводимых в сложных инженерно-геологических условиях, необходимо уметь делать многофакторную оценку температурных изменений, позволяющую прогнозировать будущие неблагоприятные воздействия. Последнее приобретает особое значение в связи с необходимостью разработки специальных мероприятий по охране окружающей среды и обеспечению устойчивости сооружений.

5.2. Постановка задачи

Большая часть Норильского промышленного района расположена на территории со сплошным распространением вечномёрзлых грунтов мощностью до 150 м и температурой в зоне годовых нулевых амплитуд (10 – 14 м) от -2.0 °С до -5.0 °С. Слой сезонного оттаивания на застроенных территориях достигает в среднем 2 – 3 м. Многолетнемёрзлые грунты на территории формировались несколько сотен тысяч лет. Большая часть жилых и административных объектов построена с учетом использования грунтов в процессе строительства и всего периода эксплуатации в мерзлом состоянии. Отепление грунтов приводит к неравномерным осадкам грунтовых оснований и, как следствие, к деформированию строительных конструкций. Соответственно, контроль за состоянием мерзлых грунтов в основании фундаментов должен осуществляться при помощи инструментальных пунктов наблюдений – температурных трубок, установленных в пробуренные скважины глубиной до подошвы фундамента. Температурные скважины позволяют предугадывать и предотвращать ситуации, связанные с появлением и развитием отепления грунтов, деформаций строительных конструкций. Здания и сооружения построены большей частью на сваях глубиной заложения от 8 до 40 м в зависимости от инженерно-геологических и мерзлотно-грунтовых условий. На сегодняшний день наблюдения

свидетельствуют о начале развития процесса деградации вечномёрзлых пород. В связи с этим актуальны вопросы обеспечения устойчивости зданий и сооружений. Решение задачи заключается в постоянном температурном мониторинге грунтов и построении моделей, адекватно описывающих температурные изменения.

Выберем в качестве такой модели регрессионную зависимость (1.28), которую будем считать моделью исследуемого объекта. Выходным признаком y будет температура грунта. В качестве входных факторов можно рассмотреть

- время измерения (квартал) (x_1);
- расположение скважины (x_2);
- техническую характеристику скважины (шифр сваи) (x_3);
- глубину измерения (h);
- дату измерения (год) (d).

Отметим, что из предложенных выше факторов только «глубина измерения» является количественным фактором и принимает значение в метрах, соответствующих глубине отметки i - го измерения. Следующие факторы являются качественными и имеют уровни:

1. фактор «время измерения» имеет четыре уровня:

- уровень 1: «первый квартал»,
- уровень 2: «второй квартал»,
- уровень 3: «третий квартал»,
- уровень 4: «четвертый квартал»;

2. фактор «расположение скважины» имеет 14 уровней:

- | | |
|-----------------|-------------------|
| уровень 1: «А», | уровень 8: «З»; |
| уровень 2: «Б», | уровень 9: «К», |
| уровень 3: «В», | уровень 10: «Л», |
| уровень 4: «Г», | уровень 11: «БВ», |
| уровень 5: «Д», | уровень 12: «ВГ», |
| уровень 6: «Е», | уровень 13: «ГД», |
| уровень 7: «Ж», | уровень 14: «ДЕ»; |

3. фактор «шифр сваи» имеет уровни, соответствующие номеру сваи в расположении сооружения, согласно принятой сметной документации;

Каждому уровню качественных факторов поставим в соответствие фиктивную переменную x_{ij} , принимающую только два значения: «1», если значение соответствующего фактора в i -м наблюдении равно j -му уровню, или «0», если значение фактора в i -м наблюдении равно не j -му уровню. Для фактора «время измерения» индекс j принимает значение от 1 до 4; для фактора «расположение сваи» индекс j принимает значение от 5 до $p+4$ и для фактора «шифр сваи» индекс j принимает значение $p+5$ до $p+s+4$.

Фактор «дата измерения» используется, как количественный и принимает значения, равные разности между годом последнего измерения и годом, соответствующим i -му наблюдению.

Тогда матрица значений регрессоров принимает вид

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{14} & x_{15} & \dots & x_{1(p+4)} & x_{1(p+5)} & \dots & x_{1(p+s+4)} & h_1 & h_1^2 & d_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{n4} & x_{n5} & \dots & x_{n(p+4)} & x_{n(p+5)} & \dots & x_{n(p+s+4)} & h_n & h_n^2 & d_n \end{bmatrix},$$

где p - количество уровней фактора «расположение скважины», s - количество уровней фактора «шифр сваи»; n - число измерений температуры.

Отметим, что введенная выше матрица X имеет линейно зависимые столбцы и для использования разработанных алгоритмов был выполнен переход к линейно независимому базису факторного пространства путём исключения одного или нескольких столбцов, влияющих на линейную зависимость. Эта процедура является хорошо известной редукцией модели [17].

Требуется по имеющимся значениям отклика и входных факторов, выполнить наиболее точное оценивание вектора неизвестных параметров регрессионного уравнения (1.28). Решение поставленной задачи было выполнено с помощью предложенных выше алгоритмов на основе преобразования Фурье и вейвлет анализа.

5.3. Оценки параметров уравнениям температурных изменений

Рассмотрим данные температурного мониторинга состояния грунтов для трех зданий и сооружений, расположенных в городе Норильске. Замер выполнялся на разной глубине с помощью температурных трубок, установленных в пробуренные скважины глубиной до подошвы фундамента. Построим оценку температурных изменений, позволяющую расчетным путем прогнозировать возможные неблагоприятные воздействия нарушений мерзлотно-грунтовых условий (отеплении грунтов), влияющих на устойчивость и долговечность жилых зданий и сооружений.

В таблицах Б.1 и Б.2 (см. приложение Б) представлены данные температурного мониторинга состояния грунтов сооружения № 1. Данные собраны за период с 1989 года по 2020, где объем равен 141. Поскольку среди входных факторов присутствуют как количественные, так и качественные, то модель (1.28) может быть записана в форме модели ковариационного анализа, которая является более компактной

$$y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_r + \theta_1 h_{ijk} + \theta_2 h_{ijk}^2 + \theta_3 d_{ijk} + \varepsilon_{ijk}, \quad (5.1)$$

где $\theta = (\mu, \alpha_1, \dots, \alpha_4, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_s, \theta_1, \theta_2, \theta_3)^T$ – вектор неизвестных параметров, y_{ijk} , h_{ijk} , d_{ijk} , ε_{ijk} – значения i наблюдения температуры, глубины измерения, «дата измерения» и случайной ошибки для «времени измерения» j , «расположение скважины» k и «шифра» r соответственно. Как уже было отмечено оценивание параметров данной модели проводилось после её редукции. В итоге были получены оценки базиса функций, допускающих оценку [17], часть из которых представляет собой так называемые функции парных сравнений.

Оценки функций, допускающих оценку, уравнения (5.1), вычисленные с помощью алгоритмов из главы 3, представлены в таблице 5.1 и обозначаются через $\hat{\phi}_i(\theta)$. Отметим, что при описании результатов использовались обозначения, введённые в разделе 3.5 при исследовании данных алгоритмов адаптивного оценивания параметров регрессионных зависимостей.

Таблица 5.1 – Оценки параметров регрессионной модели для сооружения № 1

$\hat{\theta}$	Алгоритм 1	Алгоритм 2	Алгоритм 3	Алгоритм 4	Алгоритм 5
$\hat{\phi}_1(\theta)$	-4.82908	-4.73388	-5.4063	-5.4063	-5.40626
$\hat{\phi}_2(\theta)$	-0.25405	-0.00168	-1.21383	-0.80444	-0.50777
$\hat{\phi}_3(\theta)$	-0.57854	-0.88112	-0.7102	-1.16544	-0.89304
$\hat{\phi}_4(\theta)$	0.714391	0.732774	0.31003	1.039593	1.612851
$\hat{\theta}_1$	0.395301	0.311688	0.73762	0.807283	0.684927
$\hat{\theta}_2$	-0.01837	-0.01257	-0.03663	-0.04233	-0.03334
$\hat{\theta}_3$	-0.03488	-0.02632	-0.04536	-0.04883	-0.03644

График оценок температурных изменений в зависимости от наблюдений представлен на рисунке 5.1.

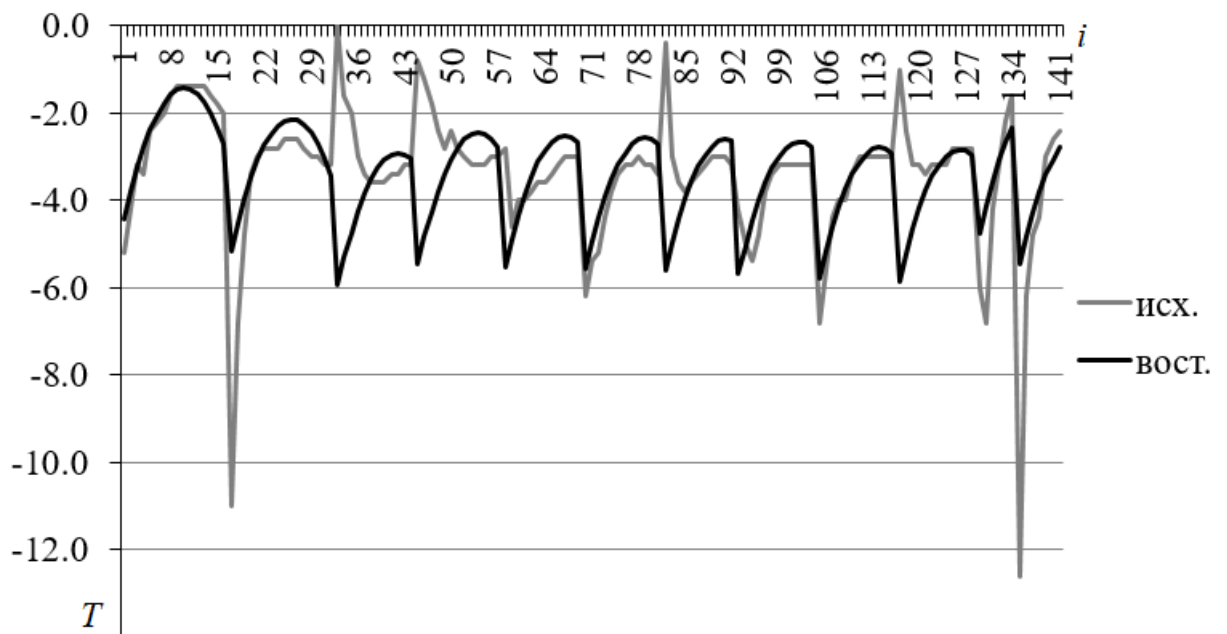


Рисунок 5.1 – Значения температуры в зависимости от наблюдения для сооружения № 1

Результаты моделирования данных мониторинга за температурными изменениями на рисунке 5.1 качественно совпадают с исходными данными,

представленными в таблицах Б.1 и Б.2 (см. приложение Б) и отображают перепады температурных состояний грунта.

В таблице Б.3 (см. приложение Б) представлены результаты температурных измерений для сооружения № 2. Данные собраны за период с 1989 года по 1990 год, в объеме, равном 16. Регрессионное уравнение температурных изменений представляется в виде (5.1). Поскольку имеющиеся данные были хуже сбалансированны [17] для данного сооружения размерность базиса функций, допускающих оценку, будет меньше, чем для сооружения № 1.

Результаты вычисления оценки функций, допускающих оценку, уравнения (5.1), вычисленные с помощью алгоритмов из главы 3, представлены в таблице 5.2 и обозначаются через $\hat{\phi}_i(\theta)$.

График оценок температурных изменений в зависимости от наблюдений представлен на рисунке 5.2. Результаты моделирования температурных изменений, представленные на рисунке 5.2, согласуются с исходными значениями в таблице Б.3 (см. приложение Б), реагируя на изменения температурных состояний грунта.

Таблица 5.2 – Оценки параметров регрессионной модели для сооружения № 2

$\hat{\theta}$	Алгоритм 1	Алгоритм 2	Алгоритм 3	Алгоритм 4	Алгоритм 5
$\hat{\phi}_5(\theta)$	-3.92863	-3.94215	-4.15624	-4.31109	-4.61976
$\hat{\phi}_6(\theta)$	-0.19313	-0.24047	-0.09593	0.11594	-0.28621
$\hat{\theta}_1$	1.989303	1.965482	1.768523	1.641265	1.272016
$\hat{\theta}_2$	-0.21785	-0.21219	-0.19903	-0.17204	-0.12731
$\hat{\theta}_3$	-10.1997	-10.1728	-10.0443	-10.483	-10.1743

В таблице Б.4 (см. приложение Б) представлены результаты температурных измерений для сооружения № 3. Данные собраны за период с 1991 года по 1996 год, объемом 58 наблюдений. Опираясь на представленные исходные данные, регрессионное уравнение температурных изменений было представлено в виде

(5.1). Поскольку имеющиеся данные также были хуже сбалансированы в сравнении с сооружением № 1 для данного сооружения размерность базиса функций, допускающих оценку, будет меньше.

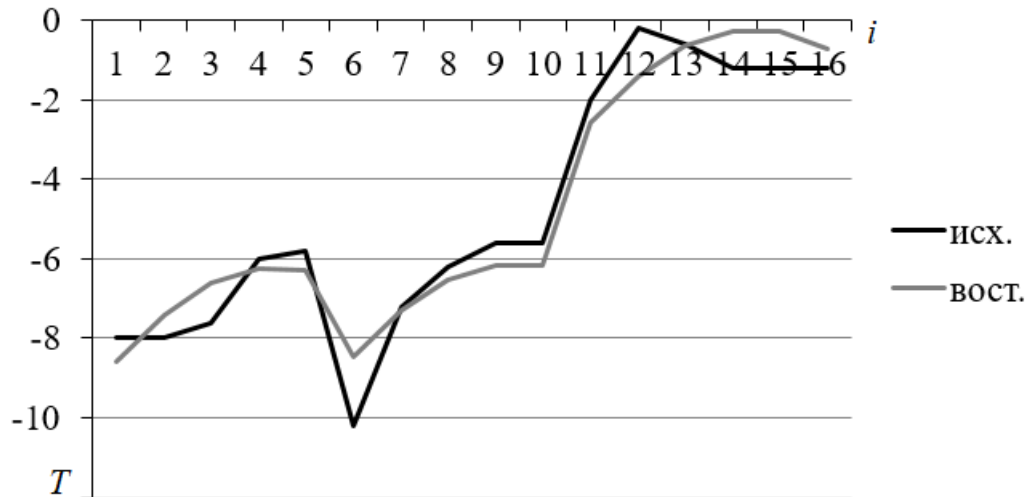


Рисунок 5.2 – Значения температуры в зависимости от наблюдения для сооружения № 2

Результаты вычисления оценки функций, допускающих оценку, уравнения (5.1), вычисленные с помощью алгоритмов из главы 3, представлены в таблице 5.3 и обозначаются через $\hat{\phi}_i(\theta)$.

Таблица 5.3- Оценки параметров регрессионной модели для сооружения № 3

$\hat{\theta}$	Алгоритм 1	Алгоритм 2	Алгоритм 3	Алгоритм 4	Алгоритм 5
$\hat{\phi}_7(\theta)$	-1.16807	1.739765	0.846225	0.846225	0.846078
$\hat{\phi}_8(\theta)$	0.884069	-0.30763	-0.14864	0.082337	0.80998
$\hat{\theta}_1$	-0.25486	-0.4869	-0.44995	-0.64083	-0.25884
$\hat{\theta}_2$	0.011628	0.020284	0.018688	0.026465	0.011491
$\hat{\theta}_3$	-0.19485	-0.52101	-0.46742	-0.51531	-0.21275

График оценок температурных изменений в зависимости от наблюдений представлен на рисунке 5.3. Температурные изменения, представленные на рисунке 5.2, были получены в результате моделирования, хорошо согласуются с исходными значениями температурных наблюдений и отражают резкие изменения температурных состояний грунта.

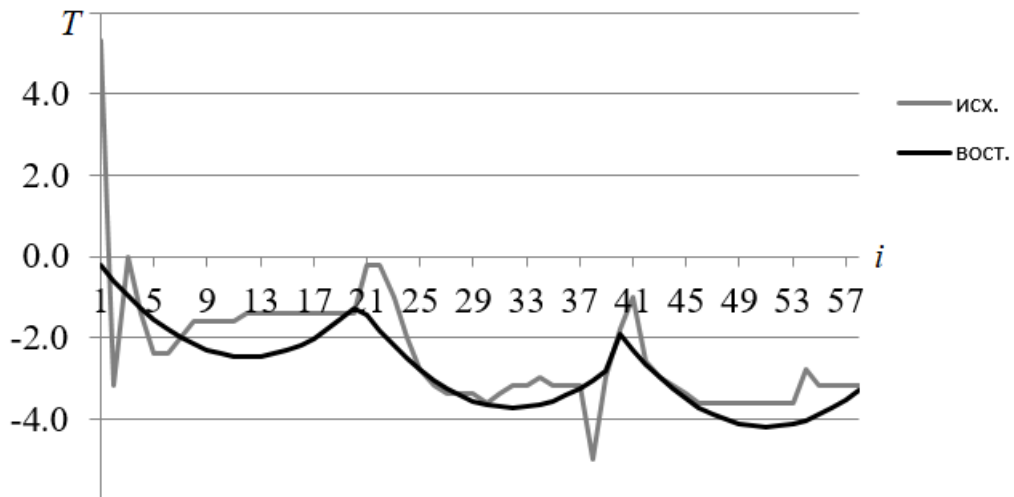


Рисунок 5.3 – Значения температуры в зависимости от наблюдения для сооружения № 3

Отметим, что оценки неизвестных параметров для прогноза температурных изменений, представленных в таблице 5.1 – 5.3, получены с использованием адаптивного алгоритма на основе материнского вейвлета DOG. Данный выбор материнского вейвлета обусловлен результатами исследования точности оценивания неизвестных параметров регрессионных зависимостей, проведенных в главе 3.

Построенные модели дают возможность прогнозировать температурные состояния грунтов, своевременно сигнализировать о нарушениях мерзлотно-грунтовых условий зданий и сооружений. Использование алгоритмов оценивания неизвестных параметров регрессионных моделей на основе преобразования Фурье и на основе вейвлетов, позволило получить более качественные оценки параметров модели, описывающих изменения температуры грунтов.

Выводы по главе 5

Основные результаты, полученные в данной главе:

1. Построена регрессионная модель, адекватно описывающая температурные изменения состояния грунтов в отдельных точках на сваях, где установлены многофазные цифровые датчики.
2. Проведена оценка неизвестных параметров регрессионной зависимости на основе разработанных адаптивных алгоритмов, основанных на преобразовании Фурье, а также на вейвлетах LITTLEWOOD&PALEY, Морле, DOG, «Мексиканская шляпа».
3. Выполнена графическая интерпретация результатов моделирования, которая дает возможность прогнозировать температурные состояния грунтов, своевременно сигнализировать о нарушениях мерзлотно-грунтовых условий для зданий и сооружений.

Заключение

Основные результаты, полученные в данной работе, заключаются в следующем:

1. Сформулирован и доказан ряд утверждений, на основе которых получены нормы для материнских вейвлетов LITTLEWOOD & PALEY, Морле, DOG, «Мексиканская шляпа» и построены системы базисных функций на их основе. Сформулирован и доказан ряд утверждений, позволяющих улучшить точность оценивания функции плотности распределения. Предложены три способа вычисления нормировочного коэффициента для построения системы базисных функций.

2. Предложены, разработаны и исследованы алгоритмы восстановления функции плотности с использованием ортогональных и неортогональных вейвлетов, а также на основе преобразования Фурье; сформулированы рекомендации относительно их использования.

3. Разработаны и исследованы алгоритмы адаптивного оценивания параметров линейно-параметризованных регрессионных моделей с использованием ортогональных и неортогональных вейвлетов и преобразования Фурье, получены выражения для логарифмической функции правдоподобия и ее производных.

4. Создан программный комплекс адаптивного оценивания параметров линейно-параметризованных регрессионных моделей и восстановления функции плотности распределения.

5. С использованием предложенных алгоритмов решена задача о прогнозе влияния геокриологических последствий глобального потепления климата на устойчивость и долговечность жилых зданий и сооружений.

Список литературы

1. *Айвазян С. А.* Прикладная статистика / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – Москва : Финансы и статистика, 1985. – 488 с.
2. *Айвазян С. А.* Прикладная статистика : основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – Москва : Финансы и статистика, 1983. – 471 с.
3. *Антонов А. В.* Системный анализ : учебник для вузов / А. В. Антонов. – Москва : Высшая школа, 2004. – 454 с.
4. *Астафьева Н. М.* Вейвлет-анализ : основы теории и примеры применения // Успехи физических наук. – 1996. – № 11. – С. 1145–1160.
5. *Банди Б.* Методы оптимизации. Вводный курс : пер. с англ. / Б. Банди. – Москва : Радио и связь, 1988. – 128 с.
6. *Барра Ж.-Р.* Основные понятия математической статистики / Ж.-Р. Барра. – Москва : Мир, 1974. – 275 с.
7. *Биккин Х. М.* Основы работы в статистическом пакете STADIA. – Екатеринбург: Уральский институт РАНХ и ГС. – 2012.
8. *Болдин М. В.* Знаковый статистический анализ линейных моделей / М. В. Болдин, Г. И. Симонова, Ю. Г. Тюрин. – Москва : Наука : Физматлит, 1997. – 208 с.
9. *Боровков А. А.* Математическая статистика. Оценка параметров. Проверка гипотез / А. А. Боровков. – Москва : Наука, 1984. – 472 с.
10. Вейвлет-анализ в примерах : учебное пособие / О. В. Нагорнов [и др.]. – Москва : НИЯУ МИФИ, 2010. – 120 с.
11. *Витязев В. В.* Вейвлет-анализ временных рядов / В. В. Витязев. – Санкт-Петербург : Изд-во С.-Петербур. ун-та, 2001. – 58 с.
12. *Воробьев В. И.* Теория и практика вейвлет-преобразования / В. И. Воробьев, В. Г. Грибунин : Воен. ун-т связи. – Санкт-Петербург : ВУС, 1999. – 203 с.

13. *Вучков И.* Прикладной линейный регрессионный анализ / И. Вучков, Л. Бояджилова, Е. Солаков ; пер. с болг. и предисл. Ю. П. Адлер. – Москва : Финансы и статистика, 1987. – 239 с.
14. *Геворкян П. С.* Теория вероятностей и математическая статистика / П. С. Геворкян, А. В. Потемкин, И. М. Эйсымонт. – Москва : Физматлит, 2016. – 176 с.
15. *Гмурман В. Е.* Теория вероятностей и математическая статистика : учебник для прикладного бакалавриата / В. Е. Гмурман. – Москва : Юрайт, 2016. – 479 с.
16. *Демиденко Е. З.* Линейная и нелинейная регрессии / Е. З. Демиденко. – Москва : Финансы и статистика, 1981. – 302 с.
17. *Денисов В. И., Полетаева И. А, Хабаров В. И.* Экспертная система для анализа многофакторных объектов. Дисперсионный анализ. Прецедентный подход : монография / В. И. Денисов, И. А. Полетаева, В. И. Хабаров. – : Новосибирский электротехнический институт, 1992. – 128 с.
18. *Денисов В. И.* Методы построения многофакторных моделей по неоднородным, негауссовским, зависимым наблюдениям : монография / В. И. Денисов, Д. В. Лисицин. – Новосибирск : Изд-во НГТУ, 2008. – 360 с.
19. *Денисов В. И.* Полупараметрическое восстановление функции плотности на основе обобщенного лямбда-распределения в задаче идентификации регрессионных моделей / В. И. Денисов, В. С. Тимофеев, Е. А. Хайленко // Сибирский журнал индустриальной математики. – 2014. – Т. 17, № 3. – С. 71–77.
20. *Добеши И.* Десять лекций по вейвлетам / И. Добеши. – Ижевск : Регулярная и хаотическая динамика, 2001. – 464 с.
21. *Дрейпер Н.* Прикладной регрессионный анализ / Н. Дрейпер, Н. Смит. – Москва : Статистика, 1973. – 392 с.
22. *Захарова Т. В.* Теория вейвлетов и ее применение в обработке сигналов : учебное пособие / Т. В. Захарова, О. В. Шестаков ; Моск. гос. ун-т им. М. В. Ломоносова, Фак. вычисл. математики и кибернетики, Каф. мат. статистики. – Москва : МастерПринт, 2018. – 178 с.

23. *Зорич В. А.* Математический анализ. Ч. 1 / В. А. Зорич. – 2-е изд., испр. и доп. – Москва : Фазис, 1997. – 554 с.
24. *Ивахненко А. Г.* Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степашко. – Киев : Наукова думка, 1985. – 216 с.
25. *Ивченко Г. И.* Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – Москва : Либроком, 2019. – 352 с.
26. Иллюстрированный самоучитель по SPSS. – URL: <http://www.hr-portal.ru/spss/index.php> (дата обращения: 03.05.2022).
27. *Исаева Е. В.* К вопросу об отслеживании смысловой целостности текста / Е. В. Исаева, В. А. Селезнев // Наука. Техника. Инновации : тезисы докладов : в 5 ч. – Новосибирск, 2001. – Ч. 1. – С. 116–117.
28. *Исаева Е. В.* О задаче идентификации смысловой целостности художественных текстов / Е. В. Исаева, В. А. Селезнев // Наука. Технологии. Инновации : труды. – Новосибирск, 2003. – С. 224–225.
29. *Исаева Е. В.* Оценивание функции плотности распределения с использованием вейвлета Морле = Estimation of the distribution density function using the Morlet wavelet / Е. В. Исаева. – DOI 10.17513/mjpf.13353 // Международный журнал прикладных и фундаментальных исследований. – 2022. – № 2. – С. 22–27.
30. *Кендалл М.* Теория распределений / М. Кендалл, А. Стьюарт. – Москва : Наука, 1966. – 587 с.
31. *Киселев А.* Основы теории вейвлет-преобразования / А. Киселев. – URL: <http://www.basegroup.ru/library/cleaning/intro-to-wavelets/> (дата обращения: 03.05.2022).
32. *Козлов П. В.* Вейвлет-преобразование и анализ временных рядов / П. В. Козлов, Б. Б. Чен // Вестник Кыргызско-Российского славянского университета. – 2002. – Т. 2, № 2. – С. 48–56.
33. *Колмогоров А. Н.* Элементы теории функций и функционального анализа / А. Н. Колмогоров, С. В. Фомин. – Москва : Физматлит, 2004. – 570 с.
34. *Крамер Г.* Математические методы статистики / Г. Крамер. – Москва : Мир, 1975. – 648 с.

35. *Кудрявцев Л. Д.* Курс математического анализа. Т. 1 / Л. Д. Кудрявцев. – Москва : Дрофа, 2003. – 704 с.
36. *Левкович-Маслюк Л.* Введение в вейвлет-анализ / Л. Левкович-Маслюк, А. Переберин // *ГрафиКон-98* : 8-я Международная конференция по компьютерной графике и визуализации : труды конференции. – Москва, 1998.
37. *Макаров А.* А. STADIA против Statgraphics, или Кто ваш «лоцман» в море статистических данных // *Мир ПК*. – 1992. – № 3. – С. 58 – 66.
38. *Максимов Ю. Д.* Математическая статистика : опорный конспект / Ю. Д. Максимов. – Москва : Проспект, 2016. – 104 с.
39. *Малашкевич И. А.* Вейвлет-анализ сигналов : от теории к практике : учебное пособие / И. А. Малашкевич ; Поволж. гос. технол. ун-т. – Йошкар-Ола : ПГТУ, 2016. – 276 с.
40. *Митропольский А. К.* Техника статистических вычислений / А. К. Митропольский. – Москва : Наука, 1971. – 576 с.
41. *Мудров В. И.* Метод наименьших модулей / В. И. Мудров, В. Л. Кушко. – Москва : Знание, 1971. – 61 с.
42. Оценивание функции плотности распределения на основе преобразований Фурье = Estimation of the distribution function based on the Fourier transform / В. С. Тимофеев, Е. В. Исаева, А. Э. Слободчикова, Е. Д. Малышкина. // *Обработка информации и математическое моделирование : материалы Рос. науч.-техн. конф., Новосибирск, 20–21 апр. 2022 г.* – Новосибирск : Изд-во СибГУТИ, 2022. – С. 170–177. – 20 экз. – ISBN 978-5-91434-070-1.
43. Оценивание функции плотности распределения с использованием вейвлета Литлвуда – Пэли / В. С. Тимофеев, Е. В. Исаева, Е. Д. Малышкина, А. Э. Слободчикова // *Обработка информации и математическое моделирование : материалы Рос. науч.-техн. конф., Новосибирск, 23–24 апр. 2020 г.* – Новосибирск, 2020. – С. 165–170.
44. *Переберин А. В.* О систематизации вейвлет-преобразований // *Вычислительные методы и программирование*. – 2002. – Т. 2. – С. 15–40.

45. *Петухов А. П.* Введение в теорию базисов всплесков / А. П. Петухов. – Санкт-Петербург : Изд-во СПбГТУ, 1999. – 132 с.
46. *Пеньков В. Б.* Сравнительный анализ процедур ортогонализации базисов евклидовых и гильбертовых пространств / В. Б. Пеньков, Л. В. Левина. // Международный журнал прикладных и фундаментальных исследований. – 2020. – № 3 – С. 103 – 107.
47. *Плошко Б. Г.* История статистики / Б. Г. Плошко, И. И. Елисеева. – Москва : Финансы и статистика, 1990. – 295 с.
48. *Поликар Р.* Введение в вейвлет-преобразование / Р. Поликар ; пер. В. Г. Грибунина. – Санкт-Петербург : АВТЭКС, [2012].
49. Программный комплекс для восстановления функции плотности и оценивания параметров регрессионных моделей на основе вейвлет анализа (WTiRM V1.0) : свидетельство о гос. регистрации программы для ЭВМ № 2022613412 : заявл. 24.02.2022 : зарег. 14.03.2022 / Исаева Е. В., Малышкина Е. Д., Слободчикова А. Э., Тимофеев В. С. ; правообладатель Новосиб. гос. техн. ун-т. – 1 с.
50. *Прокопенко Н. Ю.* Методы оптимизации / Н. Ю. Прокопенко. – Нижний Новгород : ННГАСУ, 2018 –121 с.
51. *Пугачев В. С.* Теория вероятностей и математическая статистика / В. С. Пугачев. – Москва : Наука, 1979. – 496 с.
52. *Рао С. Р.* Линейные статистические методы и их применение / С. Р. Рао. – Москва : Наука, 1968. – 548 с.
53. *Хампель Ф.* Робастность в статистике. Поход на основе функций влияния: Пер. с англ. / Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. – Москва : Мир, 1989. – 512 с.
54. *Самарский А. А.* Численные методы : учебное пособие для вузов / А. А. Самарский, А. В. Гулин. – Москва : Наука, 1989. – 432 с.
55. *Соболь И. М.* Численные методы Монте-Карло / И. М. Соболь. – Москва : Наука, 1973. – 312 с.

56. Справочник по теории вероятностей и математической статистике / В. С. Королюк, Н. И. Портенко, А. В. Скороходов, А. Ф. Турбин. – Москва : Наука, 1985. – 640 с.
57. Тимофеев В. С. Адаптивное оценивание параметров регрессионных моделей с использованием обобщенного лямбда-распределения / В. С. Тимофеев, Е. А. Хайленко // Доклады Академии наук высшей школы Российской Федерации. – 2010. – № 2 (15). – С. 25–36.
58. Тимофеев В. С. Об оценивании параметров регрессионных моделей с использованием вейвлетов / В. С. Тимофеев, Е. В. Исаева // Современные наукоемкие технологии. – 2022. – № 4. – С. 114–121.
59. Тимофеев В. С. Об оценивании функции плотности распределения случайной величины с использованием вейвлетов / В. С. Тимофеев, Е. В. Исаева // Научный вестник НГТУ. – 2019. – № 4 (77). – С. 71–84.
60. Тимофеев В. С. Особенности оценивания функции плотности распределения с помощью не ортонормированных вейвлетов / В. С. Тимофеев, Е. В. Исаева // Южно-Сибирский научный вестник. – 2022. – № 2 (42). – С. 87–94. – URL: <http://s-sibsb.ru/issues/122-2022-issues/issue-42/1348-12> (дата обращения: 03.05.2022).
61. Тимофеев В. С. Оценивание параметров регрессионных зависимостей с использованием кривых Пирсона. Ч. 1 // Научный вестник НГТУ. – 2009. – № 4 (37). – С. 57–66.
62. Тимофеев В. С. Оценивание параметров регрессионных зависимостей с использованием кривых Пирсона. Ч. 2 // Научный вестник НГТУ. – 2010. – № 1 (38). – С. 57–62.
63. Тимофеев В. С. Робастные оценки моментов при идентификации лямбда распределения в рамках адаптивного оценивания / В. С. Тимофеев, Е. А. Хайленко // Доклады Академии наук высшей школы Российской Федерации. – 2017. – № 4 (37). – С. 101–111.

64. Тимофеев В. С. Устойчивое оценивание параметров регрессионных моделей с использованием идей метода наименьших квадратов / В. С. Тимофеев, Е. А. Вострецова // Научный вестник НГТУ. – 2007. – № 2 (27). – С. 57–67.
65. Тимофеев В. С. Ядерные оценки плотности при идентификации уравнений регрессии // Научный вестник НГТУ. – 2010. – № 3 (40). – С. 41–50.
66. Фишер Р. А. Статистические методы для исследователей / Р. А. Фишер. – Москва : Госстатиздат, 1958. – 268 с.
67. Фрейзер М. Введение в вейвлеты в свете линейной алгебры / М. Фрейзер ; пер. с англ. Я. М. Жилейкина. – Москва : Бином. Лаборатория знаний, 2007. – 487 с.
68. Хеттманспергер Т. П. Статистические выводы, основанные на рангах / Т. П. Хеттманспергер ; пер. с англ. Д. С. Шмерлинга. – Москва : Финансы и статистика, 1987. – 333 с.
69. Чернов А. В. О применении квадратичных экспонент для дискретизации задач оптимального управления // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. – 2017. – Т. 27, вып. 4. – С. 558–575.
70. Чуи К. Введение в вейвлеты / К. Чуи. – Москва : Мир, 2001. – 412 с.
71. Шестаков О. В. Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов / О. В. Шестаков ; Моск. гос. ун-т им. М. В. Ломоносова, Фак. вычисл. математики и кибернетики. – Москва : Аргамак-Медиа, 2016. – 200 с.
72. Шурыгин А. М. Прикладная статистика : робастность, оценивание, прогноз / А. М. Шурыгин. – Москва : Финансы и статистика, 2000. – 224 с.
73. Электронный учебник StatSoft [Электронный ресурс]. – URL: https://http://statsoft.ru/resources/statistica_text_book.php/ (visited on 05/12/2021).
74. Aziz I. Quadrature rules for numerical integration based on Haar wavelets and hybrid functions / I. Aziz, Siraj-ul-Islam, W. Khan // Computers and Mathematics with Applications. – 2011. – Vol. 61 (9). – P. 2770–2781.
75. Aldroubi A. Wavelets on irregular grids with arbitrary dilation matrices, and frame atoms for $L^2(\mathbb{R}^d)$ / A. Aldroubi, C. Cabrelli, U. Molter // Applied and Computational Harmonic Analysis. – 2004. – Vol. 17 (2). – P. 119–140.

76. *Brook Q.* Lean Six Sigma and Minitab: The Complete Toolbox Guide for All Lean Six Sigma Practitioners – United Kingdom: OPEX Resources Ltd, 2010. – 293 p.
77. *Chen C.* Haar wavelet method for solving lumped and distributed-parameter systems / C. Chen, C. Hsiao // IEE Proceedings – Control Theory and Applications. – 1997.– Vol. 144 (1). – P. 87–94.
78. *Daubechies I.* Commutation for irregular subdivision / I. Daubechies, I. Guskov, W. Sweldens // Constructive Approximation. – 2001. – Vol. 17 (4). – P. 479–514.
79. Density estimation by wavelet thresholding / D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard // Annals of Statistics. – 1996. – Vol. 24 (2). – P. 508–539.
80. *Dong J.* Multinomial probability estimation by wavelet thresholding / J. Dong, R. Jiang // Communication in Statistics. Theory and Methods. – 2009. – Vol. 38. – P. 1486–1507.
81. *Dubeau F.* Non-uniform Haar wavelets / F. Dubeau, S. Elmejdani, R. Ksantini // Applied Mathematics and Computation. – 2004. – Vol. 159 (3). – P. 675–693.
82. *Fan J.* Local polynomial modelling and its application / J. Fan, I. Gijbels. – London : CRC Press, 1996. – 360 p.
83. *Foster G.* Wavelets for period analysis of unequally sampled time series // Astronomical Journal. – 1996. – Vol. 112 (4). – P. 1709 – 1729.
84. *Hogg R.* V A review of some adaptive statistical techniques / R. V. Hogg, R. V. Lenth // Communications in Statistics – Theory and methods, 1984. – Vol. 13. – P. 1551 – 157
85. IEEE transactions on information theory. – 1992. – Vol. 38, № 2. – Специальный номер по вейвлетам.
86. Installing NumPy [Электронный ресурс]. — URL: <https://numpy.org/install/> (visited on 12/27/2021).
87. *Lakhany A.* Estimation the parameters of the Generalized Lambda Distribution / A. Lakhany, H. Mausser // ALGO Research Quarterly. – 2000. – Vol. 3 (3). – P. 27–58.

88. *Lepik Ü.* Haar wavelets : with applications / Ü. Lepik, H. Hein. – Cham : Springer International Publishing, 2014. – 207 p. – (Mathematical Engineering).
89. *Lepik Ü.* Numerical solution of evolution equations by the Haar wavelet method // Applied Mathematics and Computation. – 2007. – Vol. 185 (1). – P. 695–704.
90. *Lutz M.* Learning Python [Текст] / M. Lutz. – 5th ed. – CA : O’Reilly Media, Inc, 2013. – 1648 p.
91. *Maleknejad K.* Using rationalized Haar wavelet for solving linear integral equations / K. Maleknejad, F. Mirzaee // Applied Mathematics and Computation. – 2005. – Vol. 160 (2). – P. 579–589.
92. *Malhotra H. K.* On scaling functions of non-uniform multiresolution analysis in $L^2(\mathbb{R})$ / H. K. Malhotra, L. K. Vashisht // International Journal of Wavelets, Multiresolution and Information Processing. – 2020. – Vol. 18 (02). – P. 1950055.
93. *Mallat S.* A wavelet tour of signal processing / S. Mallat. – New York, NY : Academic Press, 1999. – 851 p.
94. *Masry E.* Multivariate probability density estimation by wavelet methods : Strong consistency and rates for stationary time series // Stochastic Processes and Their Applications. – 1997. – Vol. 67 (2). – P. 177–193.
95. Mechanical signal analysis based on the Haar-type orthogonal matrix / X. Xiang, J. Zhou, J. Yang, L. Liu, X. An, C. Li // Expert Systems with Applications. – 2009. – Vol. 36 (6). – P. 9674–9677.
96. *Nkou E. D. de.* Wavelet-based estimation in a semiparametric regression model / E. de Dieu Nkou, G. M. Nkiet // International Journal of Wavelets, Multiresolution and Information Processing. – 2021. – Vol. 20 (02). – P. 202-203.
97. PySide2.QtWidgets [Электронный ресурс]. – URL: <https://doc.qt.io/qtforpython-5/PySide2/QtWidgets/index.html> (visited on 02/05/2022).
98. Qt for Python [Электронный ресурс]. – URL: <https://doc.qt.io/qtforpython> (visited on 02/05/2022).
99. SAS Visual Analytics [Электронный ресурс]. – URL: <https://soware.ru/products/sas-visual-analytics> (visited on 02/05/2022).

100. *Silverman B.* Density estimation for statistics and data analysis / B. Silverman. – London : Chapman and Hall, 1986. – 176 p.
101. *Siraj-ul-Islam.* A comparative study of numerical integration based on Haar wavelets and hybrid functions / Siraj-ul-Islam, I. Aziz, F. Hag // Computers and Mathematics with Applications. – 2010. – Vol. 59 (6). – P. 2026–2038.
102. *Stankovic R.* The Haar wavelet transform: its status and achievements / R. Stankovic, B. Falkowski // Computers and Electrical Engineering. – 2003. – Vol. 29 (1). – P. 25–44.
103. *Timofeev V. S.* Estimating the distribution density function using a DOG wavelet / V. S. Timofeev, E. V. Isaeva // Journal of Physics: Conference Series. – 2020. – Vol. 1661. – Art. 012084. – DOI 10.1088/1742-6596/1661/1/012084.
104. *Tribouley K.* Practical estimation of multivariate densities using wavelet methods // Statistica Neerlandica. – 1995. – Vol. 49 (1). – P. 41–62.
105. *Vidakovic B.* Statistical modeling by wavelets / B. Vidakovic. – New York, NY : Wiley, 1999. – 408 p.
106. Wavelets : a tutorial in theory and applications / ed. by C. K. Chui. – Boston : Academic Press, 1992. – 723 p.
107. Wavelets on irregular point sets / I. Daubechies, I. Guskov, P. Schröder, W. Sweldens // Philosophical Transactions of the Royal Society of London. Ser. A: Mathematical, Physical, and Engineering Sciences. – 1999. – Vol. 357 (1760). – P. 2397–2413.

Приложение А Интерфейс программного комплекса

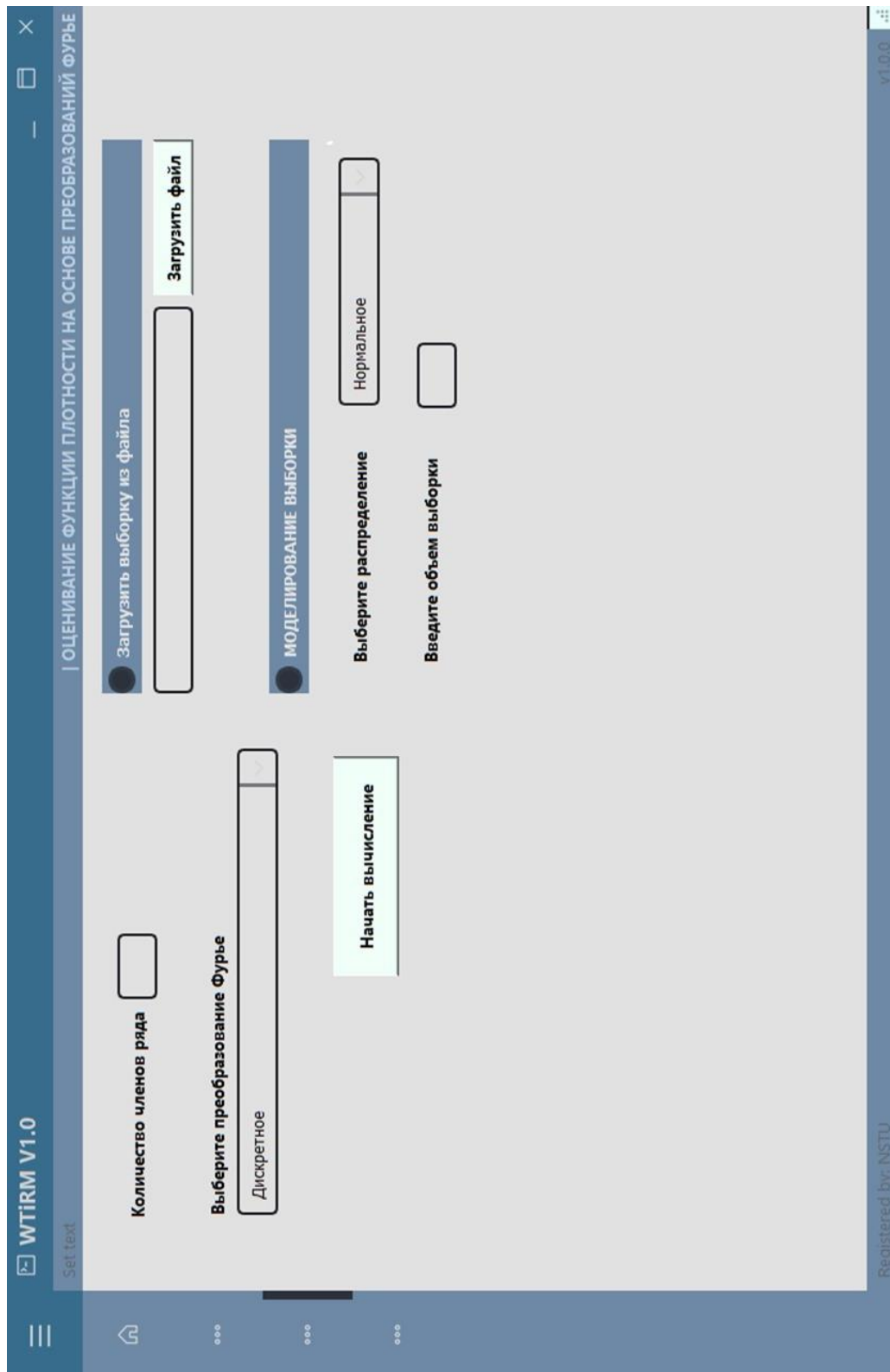


Рисунок А.1 – Интерфейс «Оценивание функции плотности с использованием преобразований Фурье»

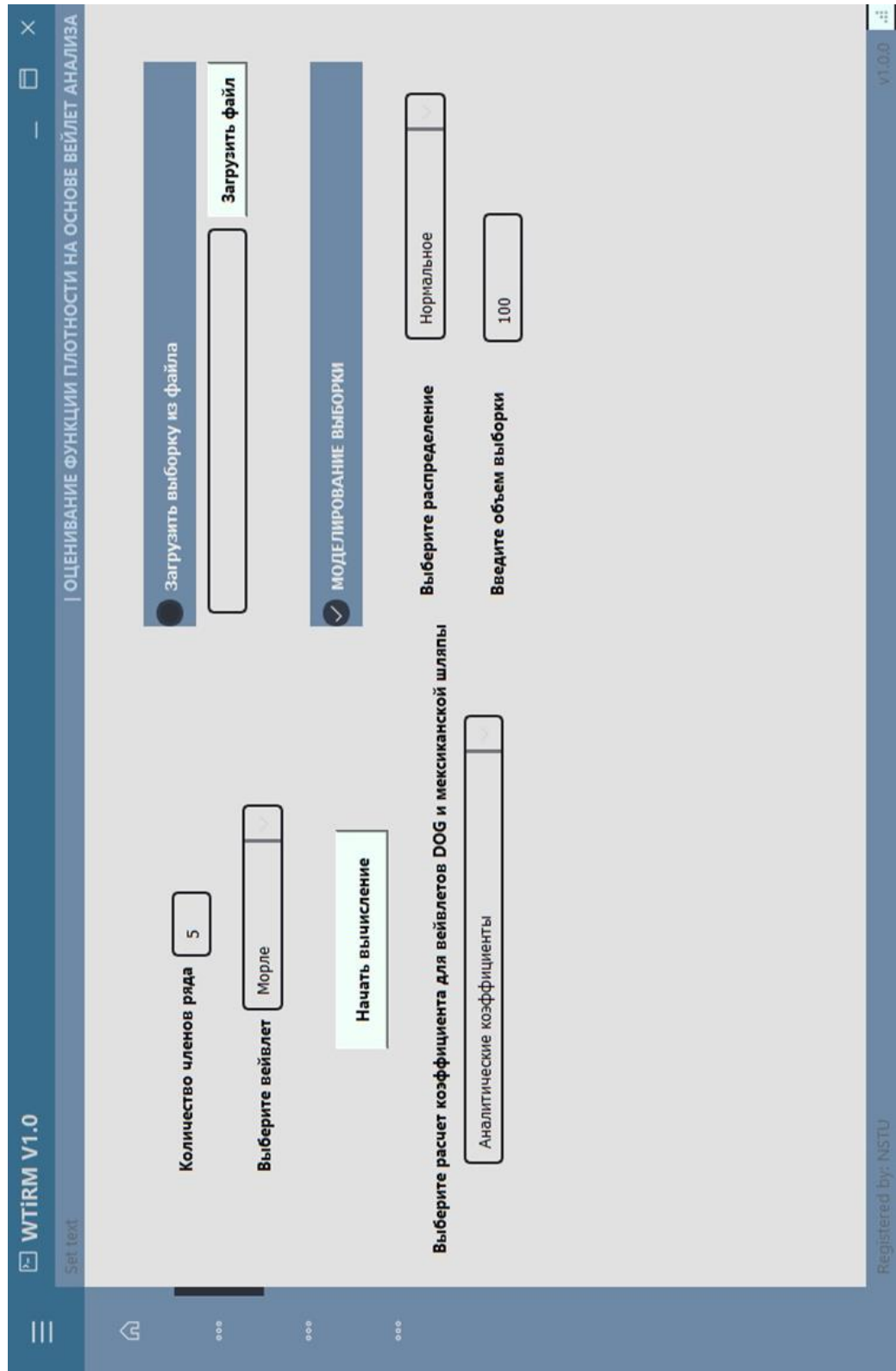


Рисунок А.2 – Интерфейс «Оценивание функции плотности с использованием вейвлетов»

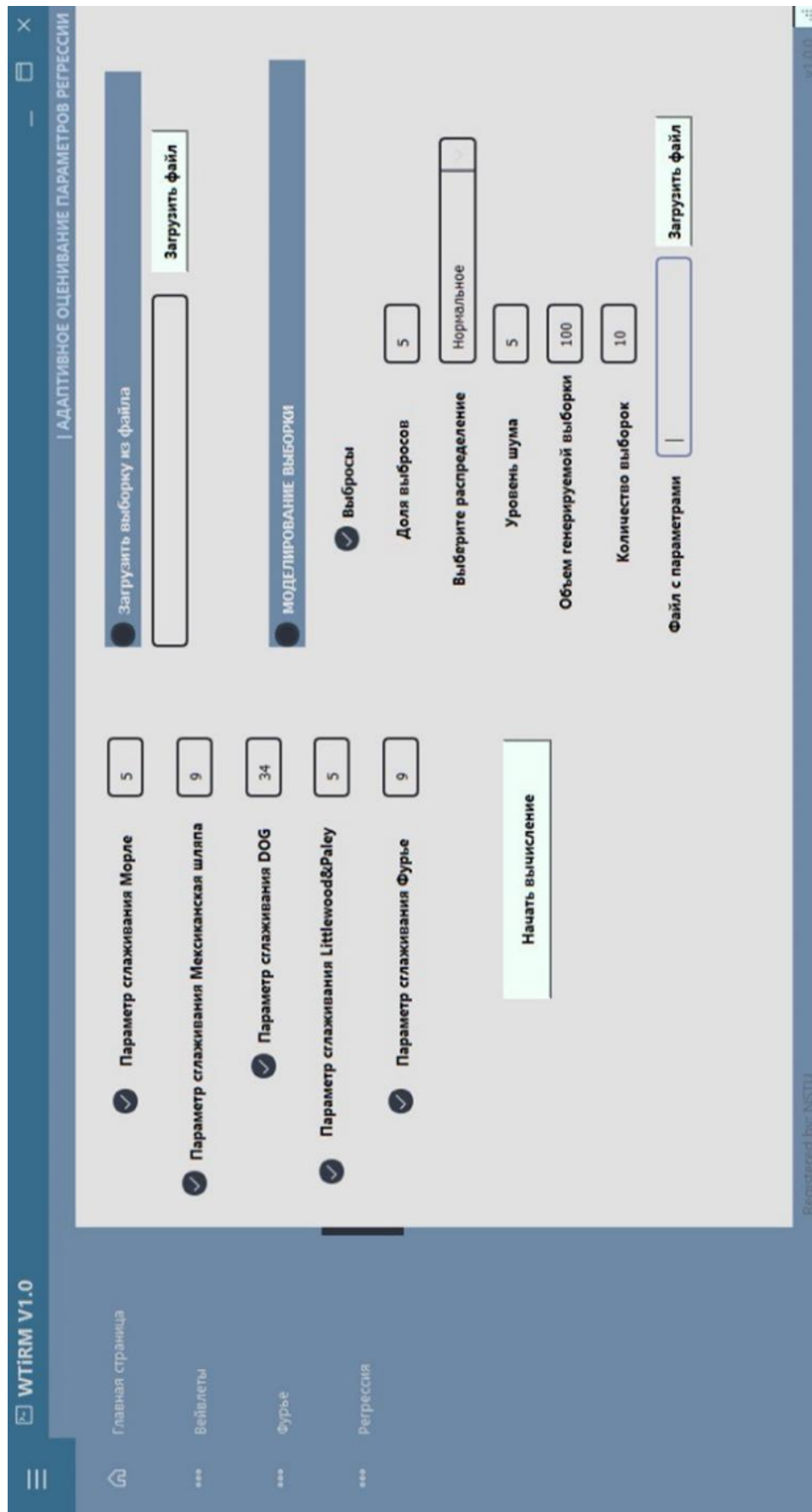


Рисунок А.3 – Интерфейс в режиме “Адаптивное оценивание параметров регрессии”

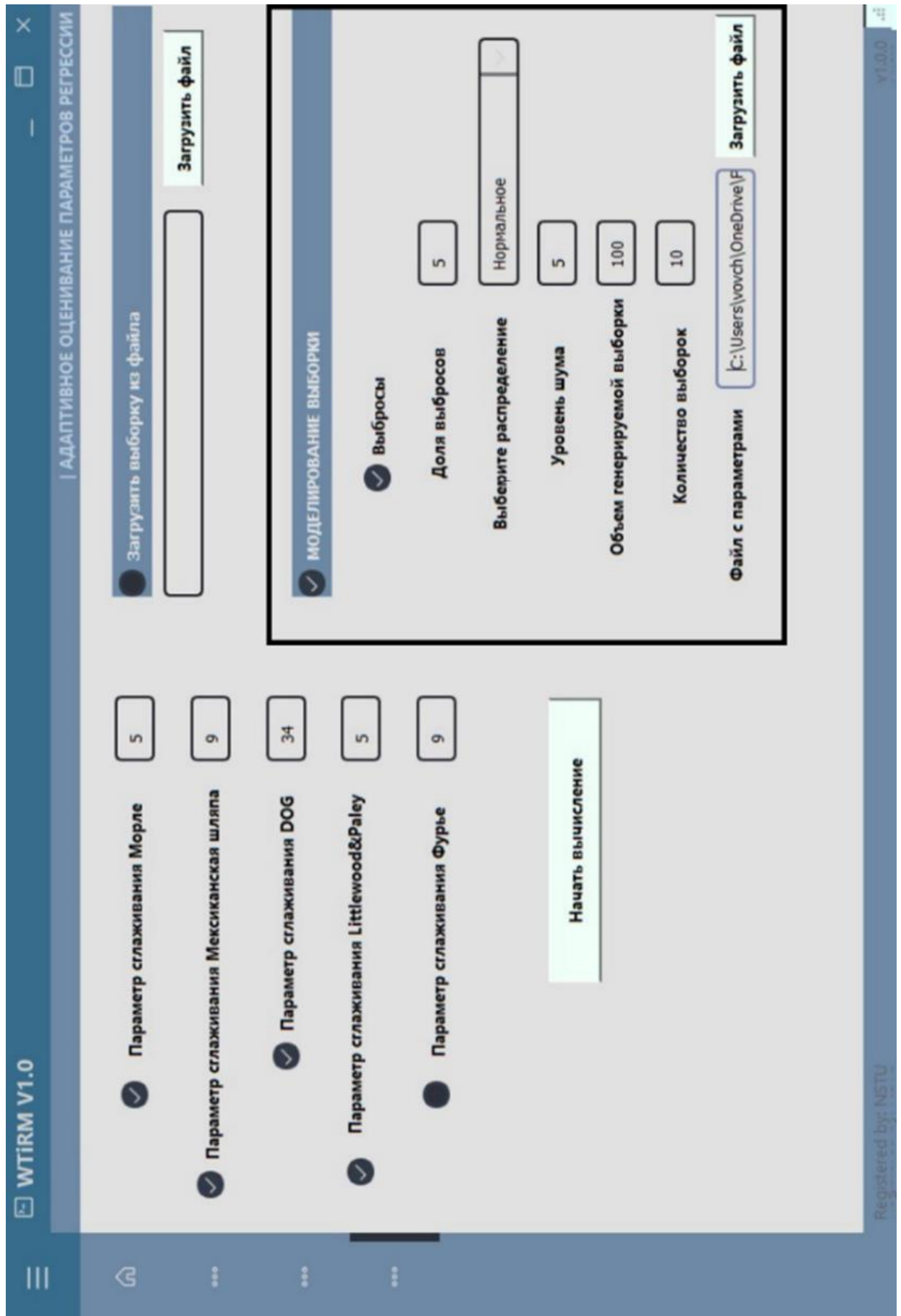


Рисунок А.4 – Область моделирования выборки

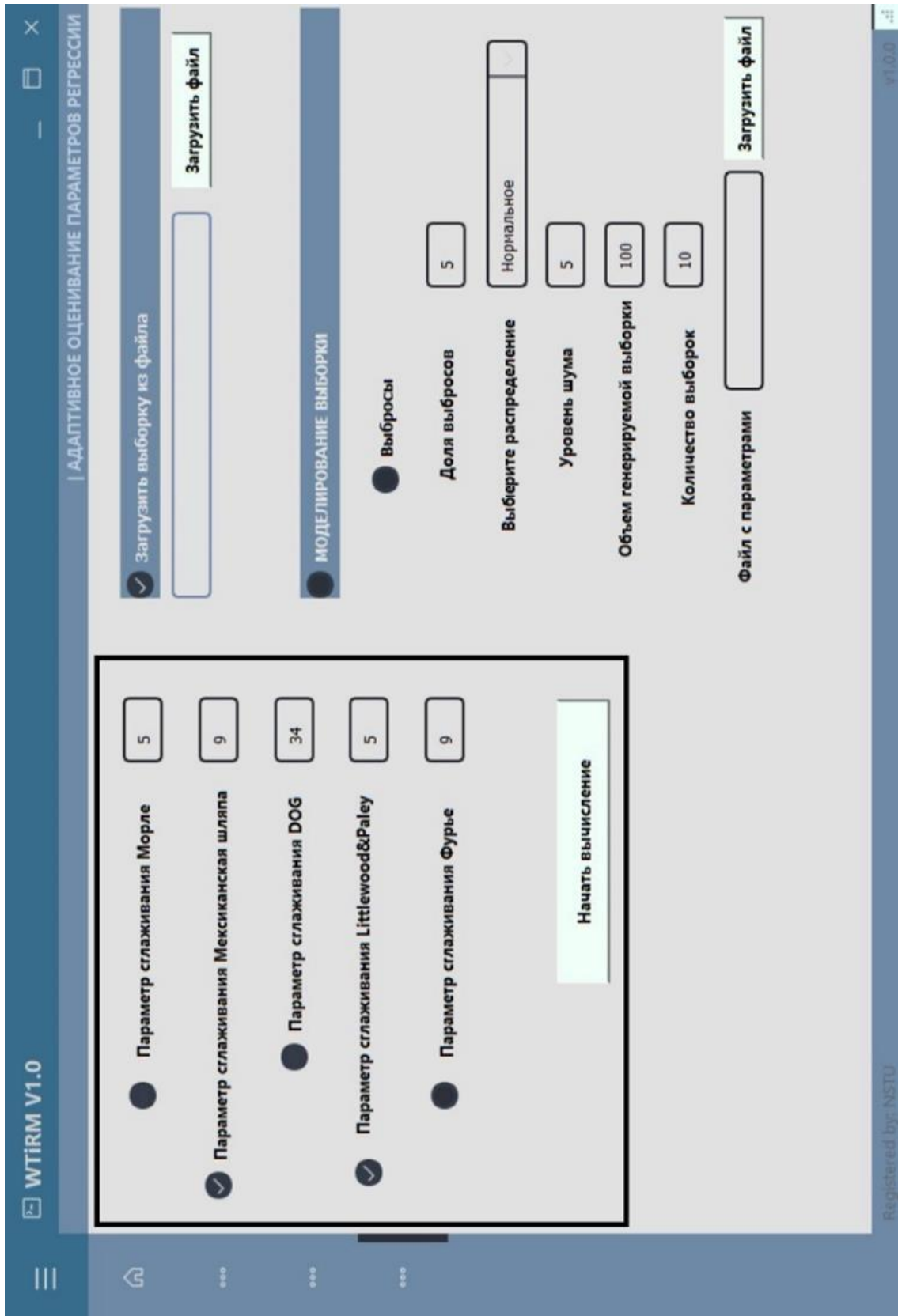


Рисунок А.5 – Область выбора функций вейвлетов и параметра сглаживания

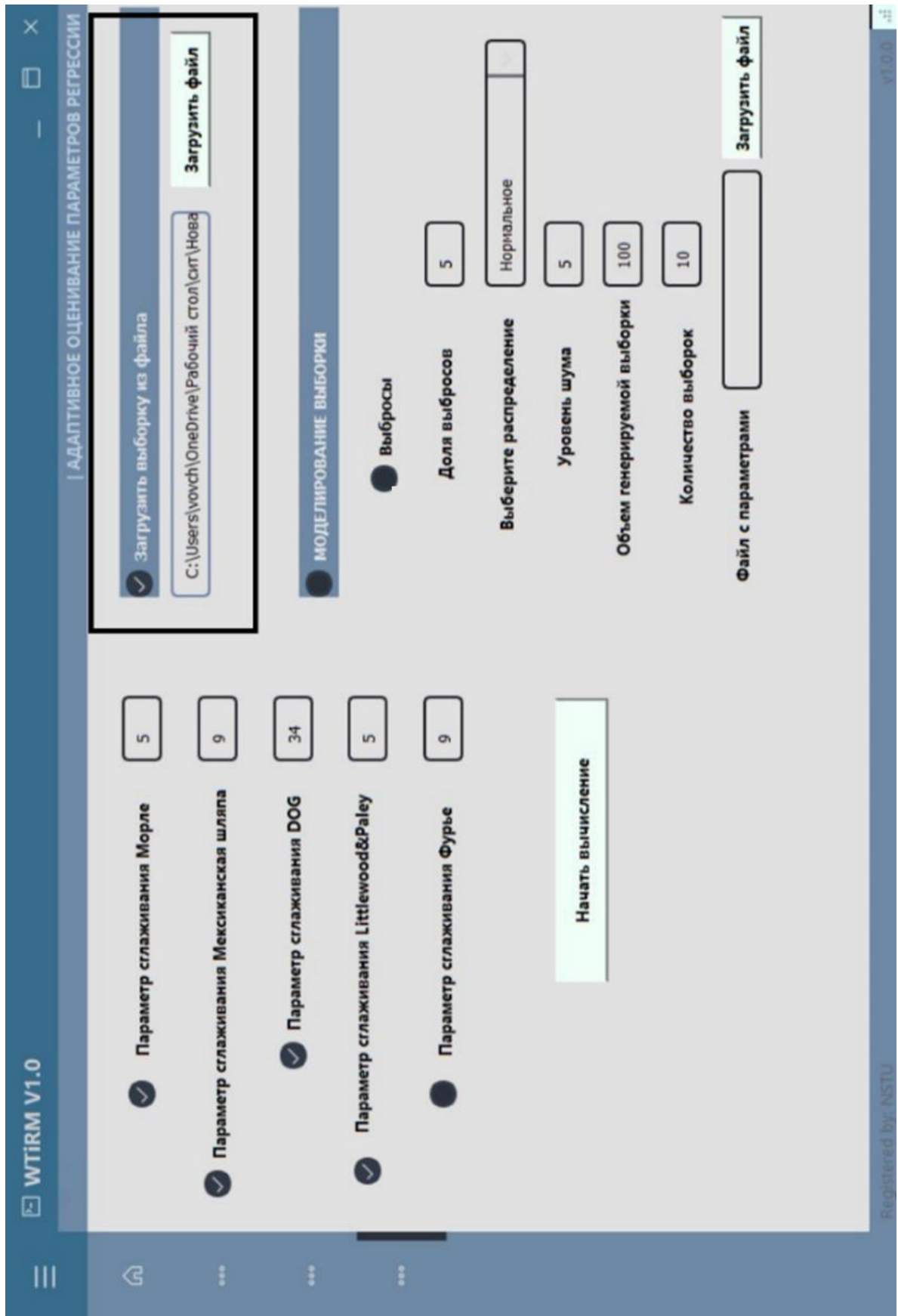


Рисунок А.6 – Область загрузки готовой выборки из файла

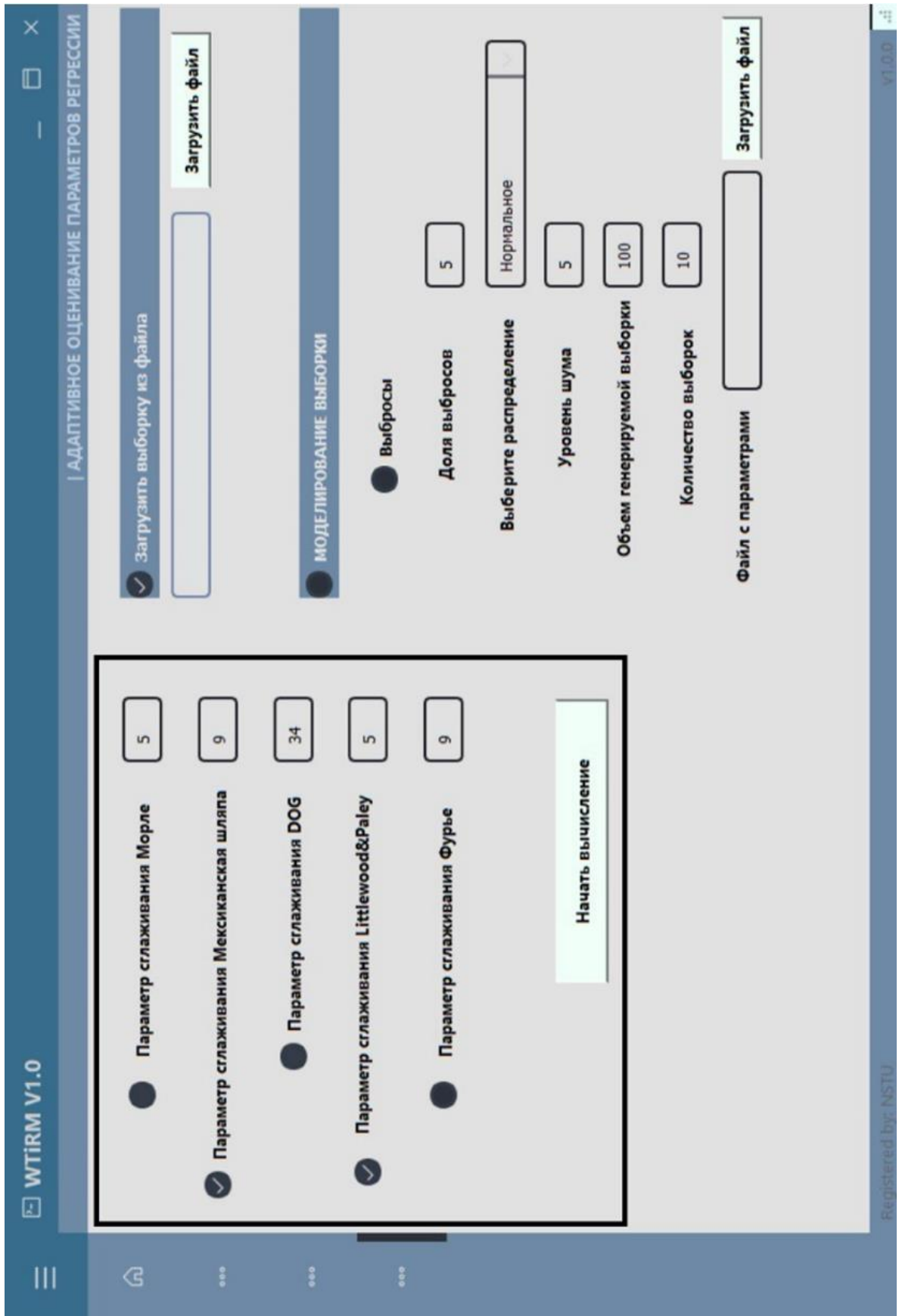


Рисунок А.7 – Выбор функций вейвлетов и параметра сглаживания

Таблица А.2 – Температурный мониторинг состояния грунта сваи В'-19

Шифр сваи В'-19										
Глубина замера	Дата									
	05.05.1989	28.12.1989	-	-	-	-	-	-	-	-
1	-6	-12.6	-	-	-	-	-	-	-	-
2	-6.8	-6.2	-	-	-	-	-	-	-	-
3	-4.2	-4.8	-	-	-	-	-	-	-	-
4	-3.2	-4.4	-	-	-	-	-	-	-	-
5	-2.2	-3	-	-	-	-	-	-	-	-
6	-1.6	-2.6	-	-	-	-	-	-	-	-
7	-	-2.4	-	-	-	-	-	-	-	-

Таблица А.3 – Температурный мониторинг состояния грунта

Температурный мониторинг состояния грунта Шифр сваи	Дата замера	Глубина					
		1	2	3	4	5	6
В'-12	18.04.1990	-8.0	-8.0	-7.6	-6.0	-5.8	-
В-8	18.04.1990	- 10.2	-7.2	-6.2	-5.6	-5.6	-
БВ-5/6	27.11.1989	-2.0	-0.2	-0.6	-1.2	-1.2	-1.2

Таблица А.4 – Температурный мониторинг состояния грунта

Шифр сваи Г-2/3	Дата		
	28.08.1991	04.11.1995	19.11.1996
Глубина замера			
1	5.3	-0.2	-1.8
2	-3.2	-0.2	-1.0
3	0.0	-1.0	-2.6
4	-1.4	-2.0	-3.0
5	-2.4	-2.8	-3.2
6	-2.4	-3.2	-3.4
7	-2.0	-3.4	-3.6
8	-1.6	-3.4	-3.6
9	-1.6	-3.4	-3.6
10	-1.6	-3.6	-3.6
11	-1.6	-3.4	-3.6
12	-1.4	-3.2	-3.6
13	-1.4	-3.2	-3.6
14	-1.4	-3.0	-3.6
15	-1.4	-3.2	-2.8
16	-1.4	-3.2	-3.2
17	-1.4	-3.2	-3.2
18	-1.4	-5.0	-3.2
19	-1.4	-3.0	-3.2
20	-1.4	–	–

Приложение В Свидетельства о регистрации программы для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022613412

Программный комплекс для восстановления функции плотности и оценивания параметров регрессионных моделей на основе вейвлет анализа (WTiRM V1.0)

Правообладатель: **ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» (RU)**

Авторы: **Тимофеев Владимир Семенович (RU), Исаева Елена Валерьевна (RU), Малышкина Елизавета Денисовна (RU), Слободчикова Александра Эдуардовна (RU)**

Заявка № **2022612390**

Дата поступления **24 февраля 2022 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **14 марта 2022 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 68b8007b14e3203a94e2b204145d5c7
Идентификатор: **Зубов Юрий Сергеевич**
Действителен с 14.03.2022 по 26.05.2023

Ю.С. Зубов



Приложение Г Акты внедрения



ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
«МЕРЗЛОТНЫЙ ИНЖЕНЕРНО-СТРОИТЕЛЬНЫЙ ЦЕНТР»

ОГРН 1112457002074, ИНН 2457072907, КПП 245701001, e-mail: miscenter@mail.ru
663300, Красноярский край, г. Норильск, ул. Завенягина, дом 4 офис 191 тел.: (3919) 36-05-60, 36-33-03 тел.: 89059790560, факс: (3919) 44-33-23

АКТ

о внедрении результатов научных исследований

Настоящим подтверждается, что Общество с ограниченной ответственностью «Мерзлотный инженерно-строительный центр» использует в аналитической работе адаптивные алгоритмы восстановления регрессионных зависимостей, разработанные и исследованные в кандидатской диссертации Исаевой Елены Валерьевны. Использование авторских алгоритмов, основанных на вейвлетах Морле, "Мексиканская шляпа", DOG, Littlewood & Paley, а также на преобразовании Фурье, позволило получить более качественные оценки параметров моделей, описывающих изменения температуры грунтов.

Построенные модели дают возможность прогнозировать температурные состояния грунтов, своевременно сигнализировать о нарушениях мерзлотно-грунтовых условий (отеплении грунтов) для зданий и сооружений, построенных в условиях крайнего севера, что крайне важно для своевременного принятия решений о защите строений.

Генеральный директор



В.Е.Петухов

22,04,2022



АКТ О ВНЕДРЕНИИ

результатов диссертационной работы Исаевой Е.В. в учебный процесс
кафедры математики и информатики

Результаты диссертационной работы Исаевой Елены Валерьевны, в частности, разработанные алгоритмы восстановления функции плотности распределения и оценивания неизвестных параметров регрессионных моделей, а также соответствующее программное обеспечение использованы при реализации учебного процесса МАОУ «Инженерный лицей НГТУ» в рамках факультативного курса по математике «Методы математического моделирования в различных прикладных задачах». Освоение учащимися соответствующих разделов дисциплины способствует приобретению необходимых знаний и умений для применения методов математического моделирования при решении практических задач.

Руководитель кафедры математики и информатики,
к.т.н., доцент

Подольян Е.В.

18.02.2022

УТВЕРЖДАЮ



Проректор по учебной работе НГТУ

Чернов С.С.

2022 г.

АКТ О ВНЕДРЕНИИ

результатов диссертационной работы Исаевой Е.В. в учебный процесс
кафедры теоретической и прикладной информатики

Результаты диссертационной работы Исаевой Елены Валерьевны, в частности, разработанные алгоритмы восстановления функции плотности распределения и оценивания неизвестных параметров регрессионных моделей, а также соответствующее программное обеспечение использованы при реализации образовательной программы по направлению 02.04.03 – «Математическое обеспечение и администрирование информационных систем» в рамках дисциплины «Статистические методы анализа данных». Освоение студентами соответствующих разделов дисциплины способствует приобретению необходимых знаний и умений для применения регрессионного анализа при решении практических задач.

Заведующий кафедрой ТПИ,
д.т.н., профессор

Чубич В.М.