

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

На правах рукописи

Демин Виктор Андреевич

**Выбор параметра размытости в непараметрической оценке условной
функции надёжности и её применение в критериях согласия**

05.13.17 – «Теоретические основы информатики»

Диссертация на соискание ученой степени кандидата технических наук

Научный руководитель:

кандидат технических наук, доцент

Чимитова Екатерина Владимировна

Новосибирск – 2016

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1 Непараметрическая оценка условной функции надежности по цензурированной выборке.....	16
1.1 Данные типа времени жизни.....	16
1.2 Регрессионные модели надежности.....	19
1.2.1 Параметрический подход	21
1.2.2 Непараметрический подход.....	25
1.3 Статистическое моделирование (метод Монте-Карло)	28
1.4 Влияние выбора параметра размытости на свойства оценки Берана	29
1.5 Факторы, оказывающие влияние на значение оптимального параметра размытости.....	32
1.6 Методы выбора оптимального значения параметра размытости	40
1.7 Выводы.....	43
ГЛАВА 2 Разработка алгоритмов выбора оптимального параметра размытости	44
2.1 Разработка адаптивного алгоритма выбора оптимального значения параметра размытости	44
2.1.1 Ядерная оценка обратной функции надёжности.....	45
2.1.2 Исследование свойств непараметрической оценки обратной функции надёжности	49

2.2 Исследование свойств оценки Берана с применением предложенного алгоритма	57
2.3 Выводы.....	68
ГЛАВА 3 Критерии согласия на основе оценки Берана	70
3.1 Проверка гипотез о виде параметрических моделей с помощью критериев согласия на основе выборок остатков	70
3.2 Критерии согласия на основе оценки Берана	75
3.3 Исследование распределений статистик и мощности критериев на основе оценки Берана.....	80
3.4 Выводы.....	88
ГЛАВА 4 Описание разработанного программного обеспечения и примеры его применения для анализа реальных данных.....	89
4.1 Описание разработанного программного обеспечения.....	89
4.2 Анализ данных о частичных разрядах, возникающих в жидких диэлектриках	92
4.2.1 Постановка задачи	92
4.2.2 Предварительный анализ	95
4.2.3 Построение модели зависимости распределения мгновенного напряжения от значения действующего	99
4.3 Анализ вероятности продления контракта рекламодателем	102
4.3.1 Введение в предметную область	102
4.3.2 Построение оценки Берана для анализа надёжности фирм-рекламодателей в городе Казань	105

4.4 Выводы.....	107
Заключение	109
СПИСОК ЛИТЕРАТУРЫ.....	111
ПРИЛОЖЕНИЕ А Акты о внедрении результатов диссертационной работы .	124

ВВЕДЕНИЕ

Современное состояние и актуальность темы исследования. На сегодняшний день невозможно представить эффективное производство без применения на различных этапах его создания и управления математических моделей и соответствующего математического аппарата. Методы прикладной статистики и раньше широко использовались в различных отраслях промышленности, в медицине, социологии, экономике и других науках, а теперь, когда существенно снизилась стоимость хранения информации, интерес к методам статистического анализа растёт с каждым накопленным терабайтом данных.

Важную роль в прикладной математической статистике играют методы анализа надёжности и выживаемости. Теория надёжности занимается изучением закономерностей появления отказов технических устройств, причинами и моделями их возникновения. Однако методы теории надёжности могут применяться не только в промышленности, но и в медицине, экономике – в любой области знаний, где предметом исследования является продолжительность жизни объектов до возникновения некоторого системного события, при котором объект перестает выполнять свои функции. Такое системное событие принято называть *отказом*. В промышленности отказом может быть выход из строя изделия, в медицине отказом может считаться смерть пациента, в экономике – банкротство компании. В задачах анализа данных об отказах нередко приходится иметь дело с цензурированными выборками, когда по каким-либо причинам время отказа для части объектов неизвестно. Разработке и исследованию статистических методов обработки цензурированных данных посвящено множество публикаций [5, 27, 28, 32, 34, 36, 46, 47, 62, 63, 67, 81].

Функция надёжности является одним из основных понятий в теории надёжности и определяет вероятность безотказной работы объекта за

некоторую наработку [96]. Для построения функции надёжности используются как непараметрические методы [19, 31, 35, 52, 62, 64, 66], так и параметрические вероятностные модели, основанные на предположении о принадлежности отказов, например, гамма-распределению, распределению Вейбулла, экспоненциальному, логнормальному и другим [6, 41, 43, 49].

Более того, при построении вероятностных моделей надёжности должна учитываться зависимость вероятности безотказной работы от значений объясняющих переменных, называемых также ковариатами [4, 35, 97, 56]. В качестве ковариат могут выступать как характеристики объекта (например, материал изделия в промышленности, возраст пациента в медицине и т.п.), так и факторы, воздействующие на объект в процессе наблюдения, которые оказывают влияние на его надёжность (например, температура и давление, при котором эксплуатируется изделие, тип лечения больных и т.д.).

Одной из наиболее распространённых моделей зависимости вероятности безотказной работы от ковариат является модель пропорциональных интенсивностей Кокса [13]. Данная модель получила широкую популярность благодаря тому, что для неё существует простая процедура оценивания неизвестной базовой функции риска и параметров, предложенная в [13]. Основное предположение, необходимое для применимости модели Кокса, заключается в том, что отношение функций риска для объектов с разными значениями ковариат остаётся постоянным во времени.

Модель пропорциональных интенсивностей имеет несколько обобщений. Например, в [26] предложена модель Ксяя, которая не требует выполнения предположения пропорциональности рисков и позволяет при разных значениях ковариаты описывать пересекающиеся функции надёжности. В [4, 7] описана модель с пересечением функций надёжности (SCE – simple cross-effect model), позволяющая описывать не только пересекающиеся функции надёжности, но и приближающиеся или отдаляющиеся друг от друга функции с разными

значениями ковариаты. Сложностью построения обобщенных моделей является весьма трудоёмкая процедура оценивания параметров [8].

Другой широко распространённой вероятностной моделью является модель ускоренных испытаний (AFT – accelerated failure time model) [5, 46, 47]. Основным предположением, обуславливающим возможность использования AFT-модели, является то, что изменение значения ковариаты должно влиять на масштаб функции надёжности, но не должно влиять на её форму.

Таким образом, качество построенной параметрической модели (её адекватность и корректность сопровождаемых выводов) зависит от того, выполняются ли предположения, необходимые для возможности её использования. Поэтому проверка выполнения такого рода предположений является обязательным этапом построения любой параметрической модели.

Для проверки соответствия модели исходным данным и выполнения предположений используют различные критерии согласия. Как для модели пропорциональных интенсивностей [9, 28, 57, 65], так и для модели ускоренных испытаний [38, 37, 62, 84, 85, 92, 93, 89, 98] существует множество критериев согласия. В работах [1, 28, 39, 65, 86, 3] предлагается подход к проверке гипотез о виде параметрических регрессионных моделей надёжности, основанный на применении классических критериев согласия к выборкам остатков. Такой подход позволяет проверить предположение о распределении отказов, однако при проверке гипотезы относительно вида регрессионной зависимости мощность критериев согласия оказывается очень низкой. Таким образом, актуальной задачей оказывается разработка критерия согласия, имеющего более высокую мощность при проверке гипотез относительно вида регрессионной зависимости.

Непараметрические методы построения моделей функции надёжности рассматривались в работах многих авторов. В частности, можно выделить работы И. Ван Кейлегома (I. Van Keilegom), М. Аркитаса (M. Akritas), Н. Веравербеке (N. Veraverbeke), В. Хардле (V. Hardle) [1, 59, 58, 104], среди

отечественных авторов – Ф.П. Тарасенко, Г.А. Кошкина, А.В. Медведева [82, 94]. Непараметрические методы отличаются своей «неприхотливостью». Они не требуют никаких априорных сведений о виде модели или иных специальных условий для данных.

Наиболее распространённой непараметрической оценкой функции надёжности является оценка Каплана–Мейера [31]. Основным преимуществом данной оценки является то, что она позволяет учитывать цензурированные наблюдения. Благодаря этому оценка Каплана-Мейера получила широкое распространение, и на её основе построен ряд критериев согласия. Оценки Каплана-Мейера, как и критерии согласия на её основе, рассматривались различными авторами [3, 25, 33, 44, 54], реализованы в большинстве статистических пакетов.

Другой распространённый непараметрический метод связан с построением оценки функции риска Нельсона–Аалена [4], которая также позволяет учитывать цензурированные наблюдения. На основе данной оценки также возможно построение критериев согласия [24, 40], однако критерии на основе оценки Нельсона–Аалена обладают меньшей мощностью, что было показано в [80].

Вместе с тем, ни оценка Нельсона–Аалена, ни оценка Каплана–Мейера не позволяют учесть влияние объясняющих переменных в функции надёжности. В этом смысле представляются актуальными разработка и развитие более широкого класса непараметрических методов, учитывающих влияние ковариат. К таким методам относится предложенное в 1981 году Р. Бераном обобщение оценки Каплана-Мейера на случай построения регрессионных моделей надёжности [10].

Несмотря на преимущества оценки Берана при использовании её в качестве непараметрической модели функции надёжности [10], можно указать весьма ограниченный перечень публикаций, посвященных исследованию статистических свойств данной оценки. Среди таких публикаций можно

выделить работы [15, 22, 42, 58], в которых свойства оценок изучаются для случайного плана эксперимента, когда значение ковариаты не фиксировано. А в [1] рассматриваются свойства оценки для неслучайного плана, когда значения ковариаты определяются заранее.

Важнейшую роль при построении непараметрических оценок играет выбор параметра размытости, от которого существенно зависит точность получаемых оценок [99, 104]. Для случая построения непараметрических регрессионных моделей с аддитивной ошибкой разработано множество методов определения оптимальных значений параметра размытости (метод кросс-валидации, метод минимума интегральной среднеквадратической ошибки, методы асимптотической оценки и другие [99, 104]). Применяя тот или иной метод выбора оптимального параметра размытости, важно понимать, что смысловая интерпретация этого параметра в оценке Берана существенно отличается от его интерпретации в классических моделях регрессии. В оценке Берана на основе параметра размытости могут быть получены лишь весовые коэффициенты, которые впоследствии используются при построении оценки. Среди публикаций на эту тему можно отметить [20], где предлагается решение данной задачи при помощи бутстреп-метода в предположении, что ковариата является случайной величиной и определяется некоторым законом распределения. Однако для задач теории надёжности такой вариант совершенно нетипичен, поскольку чаще всего объясняющие переменные являются детерминированными. Таким образом, разработка метода выбора оптимального параметра размытости при построении оценки Берана является актуальной задачей.

Цель и задачи исследования. Целью данной диссертационной работы является разработка и исследование адаптивного алгоритма выбора оптимального параметра размытости для оценки Берана, а также разработка критерия согласия на основе оценки Берана для проверки гипотезы о виде параметрической модели надёжности.

В соответствии с поставленной целью предусмотрено решение следующих задач:

- разработка адаптивного алгоритма выбора оптимального параметра размытости для оценки Берана;
- исследование статистических свойств оценок Берана, построенных с использованием предложенного алгоритма выбора оптимального параметра размытости;
- разработка критериев согласия на основе оценки Берана для проверки сложных гипотез о виде параметрических регрессионных моделей надежности;
- исследование распределений статистик и мощности предложенных критериев согласия;
- разработка программного обеспечения построения оценки Берана на основе оптимального параметра размытости;
- решение практических задач с использованием оценки Берана.

Методы исследования. Для решения поставленных задач использовались методы математической статистики, теории вероятностей, математического программирования и статистического моделирования.

Научная новизна. Научная новизна диссертационной работы заключается в следующем:

- впервые предложен адаптивный алгоритм выбора оптимального параметра размытости для непараметрической оценки Берана условной функции надёжности;
- методами компьютерного моделирования показано, что применение предложенного алгоритма выбора оптимального параметра размытости позволяет существенно повышать точность оценки Берана при различных планах эксперимента;

- предложены новые критерии согласия на основе оценки Берана, позволяющие проверять простые и сложные гипотезы о виде параметрических регрессионных моделей надежности;
- на основе результатов исследования распределений статистик и мощности предложенных критериев согласия сформулированы рекомендации по их использованию для различных планов эксперимента.

Основные положения, выносимые на защиту. На защиту выносятся:

- результаты исследования статистических свойств оценки Берана в зависимости от вида выбираемых ядерных функций, от объёма выборок, от числа опорных точек плана эксперимента, от вида регрессионной зависимости;
- предложенный и реализованный адаптивный алгоритм выбора оптимального значения параметра размытости, позволяющий строить непараметрическую оценку Берана, более точно описывающую условную функцию надежности по результатам экспериментальных наблюдений;
- предложенные критерии согласия, статистики которых представляют собой отклонение непараметрической оценки Берана от параметрической регрессионной модели надежности;
- результаты исследования распределений статистик предложенных критериев согласия, сравнительного анализа мощности критериев при проверке близких конкурирующих гипотез, методика применения предложенных критериев.

Обоснованность и достоверность научных положений, выводов и рекомендаций. Обоснованность и достоверность научных положений, выводов и рекомендаций обеспечивается:

- корректным применением математического аппарата и методов статистического моделирования для исследования свойств и распределений статистик критериев;

- совпадением результатов статистического моделирования с известными теоретическими результатами.

Личный творческий вклад автора. Личный творческий вклад автора заключается:

- в разработке адаптивного алгоритма выбора оптимального параметра размытости и проведении исследований свойств оценки Берана с использованием предложенного алгоритма;
- в разработке статистического критерия согласия на основе оценки Берана и проведении исследований распределений статистик и мощности предложенных критериев;
- в разработке программного обеспечения, реализующего предложенные алгоритмы построения оценки Берана и проверку гипотез о виде параметрических регрессионных моделей с использованием предложенных критериев согласия;
- в решении задач анализа реальных данных с применением разработанных алгоритмов и программного обеспечения.

Практическая ценность и реализация результатов. Практическая ценность результатов заключается в разработке адаптивного алгоритма выбора параметра размытости для оценки Берана, в формировании рекомендаций по использованию предложенного метода и оценки Берана для построения непараметрической оценки условной функции надёжности. Предложен универсальный критерий согласия на основе оценки Берана, предложено несколько статистик критерия, сформированы рекомендации по использованию критерия.

Разработанный алгоритм выбора параметра размытости и построение оценок Берана по цензурированным выборкам реализованы в программной системе статистического анализа данных типа времени жизни «LiTiS» (Свидетельство о государственной регистрации программы для ЭВМ № 2014661905, 2015 г. – М.: Федеральная служба по интеллектуальной

собственности (Роспатент)). Результаты внедрены в практику деятельности ООО «ДГ-Софт», филиала ПАО «Электросетьсервис ЕНЭС», а также нашли практическое применение в учебном процессе на факультете прикладной математики и информатики ФГБОУ ВО «Новосибирский государственный технический университет», что подтверждается соответствующими актами о внедрении.

Соответствие диссертации паспорту научной специальности. Содержание диссертации соответствует п. 5 области исследований «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» паспорта специальности научных работников 05.13.17 – «Теоретические основы информатики» по техническим наукам.

Апробация результатов диссертации. Результаты работы докладывались на Международном семинаре “Applied methods of statistical analysis” (Новосибирск, 2013, 2015 г.); Международном семинаре “International Workshop on Simulation” (Римини, Италия, 2013г.); международной научно-технической конференции “Актуальные проблемы электронного приборостроения” (Новосибирск, 2012, 2014 г.); Российской научно-технической конференции “Обработка информационных сигналов и математическое моделирование” (Новосибирск, 2012 г.); Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”, (Новосибирск, 2010, 2011г.); всероссийской научной конференции молодых ученых “Наука. Технология. Инновации”, (Новосибирск, 2011, 2015 г.); всероссийском научном симпозиуме “Непараметрика-XIV” (Томск, 2012 г.); международном семинаре по моделированию «International Workshop on Simulation», (Вена, Австрия, 2015 г.); Международном форуме по стратегическим технологиям IFOST-2016 (Новосибирск, 2016 г.).

Публикации. По результатам диссертационных исследований опубликовано 13 печатных работ, в том числе четыре статьи в научных журналах и изданиях, рекомендуемых ВАК РФ, восемь публикаций в материалах Международных и Российских конференций, получено свидетельство о государственной регистрации программы для ЭВМ.

Структура и объем диссертации. Общий объем диссертационной работы составляет 127 страниц, основная часть изложена на 123 страницах и состоит из введения, четырех глав основного содержания, включая 20 таблиц и 34 рисунка, заключения, списка использованных источников из 113 наименований и приложения.

Краткое содержание работы. В первой главе представлены основные определения, необходимые теоретические выкладки, приведены результаты исследований, касающиеся влияния параметра размытости на точность оценки Берана, сформулированы задачи исследования.

Во второй главе предлагается адаптивный алгоритм выбора параметра размытости, исследуются точность непараметрической оценки Берана с использованием предложенного алгоритма при различных планах эксперимента, приводится сравнительный анализ точности оценки Берана при использовании предложенного алгоритма и референтного правила выбора параметра.

В третьей главе предлагаются критерии согласия на основе оценки Берана, проводятся исследования распределений статистик предложенных критериев, проводится сравнительный анализ мощности предложенных критериев и классических критериев согласия, применяемых к выборкам остатков для проверки гипотез о виде параметрических регрессионных моделей надежности.

В четвёртой главе представлено описание разработанного программного обеспечения для построения непараметрической оценки Берана, решаются

задачи статистического анализа реальных данных с использованием разработанных алгоритмов и программного обеспечения.

В заключении приводится перечень основных результатов исследований.

ГЛАВА 1 Непараметрическая оценка условной функции надежности по цензурированной выборке

В данной главе диссертационной работы рассматривается непараметрическая регрессионная модель зависимости вероятности дожития до времени t от некоторого фактора, влияющего на продолжительность жизни. Исследуются свойства получаемых непараметрических оценок условной функции надежности.

1.1 Данные типа времени жизни

В задачах анализа данных типа времени жизни объектом исследования является группа объектов, для каждого из которых определено некоторое системное событие, часто называемое отказом. Например, проверка эффективности лекарств при лечении больных или тестирование изделий на надёжность. Данные любого случайного эксперимента, в результате которого получены отказы объектов, можно считать данными типа времени жизни.

Отказом называется событие, после возникновения которого, характеристики технического объекта (параметры) выходят за допустимые пределы. Очевидно, что это понятие является субъективным и определяется пользователем. Отказы обусловлены закономерными явлениями, вызывающими постепенное накопление повреждений: усталость, износ, старение, коррозия материалов и другие.

Одной из особенностей данных типа времени жизни является их неполнота. Во время испытаний может выйти из строя лишь некоторый процент исследуемых объектов, ряд объектов по каким-то причинам может быть снят с испытаний, время испытаний может быть ограничено. Таким образом, к концу эксперимента часть объектов может остаться в работоспособном состоянии. В медицине, чаще всего, оказывается возможным

наблюдать пациентов только некоторый ограниченный период времени, и системное событие, например, изменение некоторых жизненных показателей или смерть, может и не произойти за наблюдаемый период. Такие данные называют цензурированными. Это связано с ограниченностью времени наблюдения, времени испытаний на надежность.

Цензурированной справа называют выборку вида:

$$(Y_1, \delta_1), (Y_2, \delta_2), \dots, (Y_n, \delta_n), \quad (1.1)$$

где Y_i – значение наблюдения;

$$Y_i = \min(T_i, C_i),$$

где T_i – время наступления отказа;

C_i – время цензурирования (время завершения наблюдения за i -м объектом);

δ_i – индикатор события, который содержит информацию о причине прекращения наблюдения, здесь $i = 1, 2, \dots, n$.

Если в ходе эксперимента было зафиксировано время отказа, то $Y_i = T_i$, $\delta_i = 1$, и данное наблюдение называется *полным*. Если же нам неизвестно T_i по причине окончания наблюдения в момент $C_i \leq T_i$, то $Y_i = C_i$, $\delta_i = 0$, и наблюдение называется *цензурированным справа*. Цензурированные справа выборки, встречающиеся на практике, можно разделить на три основных типа и их комбинации.

Выделяют несколько типов цензурирования справа [48]:

– цензурирование I типа. В этом случае все объекты тестируются до заранее определенного момента времени T_C . Иными словами, момент i -го отказа T_i будет наблюдаем тогда и только тогда, когда он не превосходит момент цензурирования: $T_i \leq T_C$;

– цензурирование II типа. Испытания продолжают до наступления $k < n$ первых отказов, где n – число всех наблюдавшихся объектов (полный

объем выборки). Для оставшихся объектов момент цензурирования $T_C = T_{(k)}$, где $T_{(k)}$ – момент отказа k -го объекта;

– цензурирование III типа (случайное цензурирование). Моменты отказов T_1, \dots, T_n и моменты цензурирования C_1, \dots, C_n являются независимыми неотрицательными случайными величинами.

Одной из важнейших задач статистического анализа надёжности является изучение зависимости времени до наступления отказа от объясняющих переменных, которые также называются ковариатами. *Ковариатой* является величина, описывающая какое-либо свойство, характеристику объекта, либо степень внешнего воздействия на объект, условия проведения эксперимента. Так, в промышленных испытаниях ковариатой может являться материал, из которого сделан объект (характеристика объекта), либо температура, давление, при котором данный объект эксплуатируется (характеристика внешних условий, воздействия). В медицине ковариатами являются такие характеристики как возраст, пол, тип лечения, тип болезни и другие.

Область значений ковариаты x определяется условиями проведения эксперимента и представляет собой отрезок числовой прямой. В настоящем диссертационном исследовании рассматривается случай скалярной ковариаты, и предполагается, что ковариата является постоянной по времени величиной.

Планом эксперимента называется совокупность

$$\xi_n^m = \begin{Bmatrix} x^1 & \dots & x^m \\ n_1 & \dots & n_m \end{Bmatrix}, \quad (1.2)$$

где x^1, \dots, x^m – опорные точки плана (обязательно различные);

n_i – число объектов, исследуемых при значении ковариаты x^i , $i = \overline{1, m}$,

$$m \leq n, \quad \sum_{i=1}^m n_i = n.$$

Таким образом, все объекты выборки разделяются на m групп, соответствующих различным значениям ковариаты (опорным точкам плана).

Обычно для удобства значения ковариат в плане эксперимента линейным преобразованием приводят к отрезку $[0, 1]$.

Если опорные точки плана выбираются случайно, в соответствии с некоторым законом распределения, то такой план эксперимента называют *случайным*. При этом если распределение ковариаты является непрерывным (например, равномерное на заданном отрезке), то число групп m будет равно объёму выборки n .

Данные об отказах с ковариатами представляются в следующем виде:

$$(Y_1, x_1, \delta_1), (Y_2, x_2, \delta_2), \dots, (Y_n, x_n, \delta_n), \quad (1.3)$$

где Y_i – время безотказной работы или момент цензурирования i -го объекта;

δ_i – индикатор цензурирования i -го объекта, который принимает значение 1, если наблюдение полное, и 0, если цензурированное;

x_i – значение ковариаты, при котором наблюдался i -й объект.

1.2 Регрессионные модели надежности

Свойство объекта сохранять работоспособное состояние в течение некоторого времени или некоторой наработки называется надёжностью. Или более формально: *надёжность* – это свойство объекта сохранять во времени в установленных пределах значения всех параметров, характеризующих способность выполнять требуемые функции в заданных режимах и условиях применения, технического обслуживания и транспортирования (ГОСТ 27.002-89).

Надёжность объектов описывается, характеризуется функцией надёжности. Обозначим через T_x длительность жизни (время наработки до отказа), которая является неотрицательной случайной величиной с непрерывным законом распределения. Поскольку длительность безотказной работы объектов может зависеть от их свойств или условий проведения эксперимента (от значений ковариаты x), будем рассматривать условную функцию надёжности:

$$S(t | x) = P(T_x \geq t) = 1 - F(t | x), \quad (1.4)$$

где $F(t | x)$ – условная функция распределения случайной величины T_x .

В теории надёжности регрессионную модель принято записывать через функцию надёжности следующим образом:

$$S(t | x) = H(S_0(t); x), \quad (1.5)$$

где $S_0(t) = S(t | x = 0)$ – базовая функция надёжности; оператор $H(\cdot): [0,1] \rightarrow [0,1]$ определяет изменение базовой функции надёжности для различных значений объясняющих переменных.

Закон распределения может быть также определён функцией интенсивности

$$\lambda_x(t) = \lim_{h \rightarrow 0} (P\{t \leq T_x < t + h | T_x > t\} / h)$$

и кумулятивной функцией риска

$$\Lambda_x(t) = \int_0^t \lambda_x(u) du = -\ln(S_x(t)).$$

На основе данных об отказах объектов, полученных в результате исследования (эксперимента), можно построить вероятностные модели для оценки показателей надёжности.

В зависимости от характера привлекаемой априорной информации различают два основных подхода к построению оценки вероятностной модели надёжности: *параметрический* и *непараметрический*.

1.2.1 Параметрический подход

Если рассматриваются данные без ковариат, то при *параметрическом* подходе предполагается, что неизвестное теоретическое распределение (функция надёжности) принадлежит какому-либо семейству, зависящему от конечного числа параметров (например, семейству распределений Вейбулла). По результатам наблюдений можно оценивать неизвестные параметры и проверять те или иные гипотезы относительно их значений. Наиболее популярные семейства распределений, используемые в моделях надёжности, приведены в таблице 1.1.

Таблица 1.1 – Функции надёжности некоторых семейств распределений

Распределение	Функция надёжности, $S_0(t; \theta)$
Экспоненциальное	$\exp(-t/\theta_1)$
Вейбулла	$\exp\left(- (t/\theta_1)^{\theta_2}\right)$
Обобщенное Вейбулла	$\exp\left(1 - \left(1 + (t/\theta_1)^{\theta_2}\right)^{1/\theta_3}\right)$
Гамма	$1 - \frac{\Gamma(t/\theta_1, \theta_2)}{\Gamma(\theta_2)}$
Обобщенное гамма	$1 - \frac{\Gamma\left((t/\theta_1)^{\theta_2}, \theta_2/\theta_3\right)}{\Gamma(\theta_2/\theta_3)}$
Логнормальное	$\frac{1}{2} - \frac{1}{2\sqrt{\pi}} \Gamma\left(\frac{1}{2\theta_2^2} \ln^2(t/\theta_1), \frac{1}{2}\right)$

В случае наблюдений с ковариатами помимо параметризации функции распределения (базового распределения), также предполагается некоторый вид зависимости базовой функции надежности от ковариат. Для этого вводится функция от ковариат с неизвестными параметрами. В таблице 1.2. приведены некоторые часто используемые на практике функции от ковариат.

Таблица 1.2 – Функции от ковариаты

Название	Функция от ковариаты, $r(x;\beta)$
Аррениуса	$e^{\beta_0+\beta_1/x}$
Степенная	$e^{\beta_0+\beta_1 \ln x}$
Логлинейная	$e^{\beta_0+\beta_1 x}$

Функция от ковариаты и базовая функция распределения могут по-разному учитываться в модели. Поэтому существует множество параметрических моделей надёжности. Рассмотрим наиболее популярные из них.

Одной из наиболее популярных вероятностных моделей, описывающих зависимость надёжности от независимых ковариат, является модель пропорциональных интенсивностей, которая также называется «пропорциональной моделью Кокса». Функция надёжности для модели пропорциональных интенсивностей имеет вид [13]:

$$S_x(t;\beta) = (S_0(t))^{r(x;\beta)}. \quad (1.6)$$

В основе данной модели лежит предположение, что отношение функций интенсивности при разных значениях ковариаты x^1 и x^2 не зависит от времени:

$$\frac{\lambda_{x^2}(t)}{\lambda_{x^1}(t)} = \frac{\exp(\beta \cdot x^2)}{\exp(\beta \cdot x^1)} = \text{const}. \quad (1.7)$$

Модель пропорциональных интенсивностей имеет несколько обобщений, рассмотрим два из них. Первое из рассмотренных предложил Ксай [26]. Согласно его идее, возможным методом получения немонотонного поведения отношений функций риска является возведение базовой функции риска в некоторую степень:

$$\Lambda_x(t; \beta, \gamma) = \exp(\beta \cdot x) \{ \Lambda_0(t) \}^{\exp(\gamma \cdot x)}. \quad (1.8)$$

Функции интенсивности, соответствующие данной модели, при разных значениях ковариат пересекаются для любых значений параметров β и $\gamma \neq 0$ [26].

SCE-модель (Simple Cross-Effect model) или модель с пересечением функций надёжности, предложенная в [5] и позволяющая получить не только пересекающиеся, но и расходящиеся функции интенсивности, может быть записана следующим образом:

$$\Lambda_x(t; \beta, \gamma) = \left(1 + \exp((\beta + \gamma) \cdot x) \cdot \Lambda_0(t) \right)^{\exp(-\gamma \cdot x)} - 1. \quad (1.9)$$

Отношение функций интенсивности

$$\frac{\lambda_{x^2}(t)}{\lambda_{x^1}(t)} = \exp(\beta^T \cdot (x^2 - x^1)) \frac{\left(1 + \exp((\beta + \gamma) \cdot x^2) \Lambda_0(t) \right)^{\exp(-\gamma \cdot x^2) - 1}}{\left(1 + \exp((\beta + \gamma) \cdot x^1) \Lambda_0(t) \right)^{\exp(-\gamma \cdot x^1) - 1}}$$

монотонно, больше 1 в точке $t = 0$: $\lambda_{x^2}(0) / \lambda_{x^1}(0) = \exp(\beta \cdot (x^2 - x^1)) = c_0 > 1$.

Если $\gamma < 0$, то $\lambda_{x^2}(\infty) / \lambda_{x^1}(\infty) = \infty$, функции интенсивности расходятся, и функции надёжности не пересекаются.

Не менее распространённой моделью является модель ускоренных испытаний (Accelerated Failure Time). AFT-модель предназначена для оценивания функции надёжности изделий (систем), функционирующих в

нормальных условиях эксплуатации (при воздействии нормальных нагрузок), по данным об отказах, полученным в результате ускоренных испытаний (при использовании повышенных нагрузок). В общем виде АFT-модель ускоренных испытаний выглядит следующим образом [5, 6]:

$$S_x(t) = S_0\left(\frac{t}{r(x;\beta)}\right), \quad (1.10)$$

где $r(x;\beta)$ – функция от ковариаты;

$S_0(\cdot)$ – базовая функция надёжности.

При параметрическом подходе всегда делаются некоторые предположения о виде зависимости функции надёжности от ковариат. Важнейшим этапом построения параметрической регрессионной модели является проверка того, согласуются ли сделанные предположения с данными об отказах. Другими словами требуется проверить сложную гипотезу

$$H_0 : S(t|x) = H(S_0(t;\theta); x, \beta), \quad (1.11)$$

где $S_0(t;\theta)$ – базовая функция надёжности, β – вектор регрессионных параметров.

Для проверки гипотезы (1.11) используются различные критерии согласия, в основе которых чаще всего лежит оценка расстояния между непараметрической оценкой и функцией надёжности, соответствующей проверяемой гипотезе. В работах Чимитовой Е.В. предлагается подход к проверке данной гипотезы, основанный на применении классических критериев согласия к выборкам остатков [1, 3, 28, 39, 65, 86]. Такой подход позволяет проверить предположение о базовом распределении отказов, однако при проверке гипотезы относительно вида регрессионной зависимости мощность критериев согласия оказывается очень низкой. Таким образом, актуальной задачей оказывается разработка критерия согласия, имеющего более высокую

мощность при проверке гипотез относительно вида регрессионной зависимости.

1.2.2 Непараметрический подход

На практике не всегда существуют априорные предположения о виде распределения отказов или о функциональной зависимости функции надёжности от ковариат. К тому же, предположение может быть ошибочным, что может привести к неверным выводам. Или данных может не хватать для построения параметрической модели. Таким образом, возникает необходимость в использовании непараметрических методов.

Непараметрические методы оценивания используют только выборку и не требуют никаких априорных предположений или каких-либо специальных условий, что является несомненным плюсом непараметрического подхода перед параметрическим. Пожалуй, самым распространённым непараметрическим методом для оценивания функции надёжности является оценка Каплана-Мейера [31]:

$$S(t) = \prod_{j: \delta_j=1, Y_j \leq t} \left(1 - \frac{d_j}{r_j} \right), \quad (1.12)$$

где δ_i – индикатор цензурирования;

r_j – число объектов, наблюдаемых в момент t ;

d_j – число объектов, отказавших в момент t .

Оценка Каплана–Мейера не учитывает влияние ковариат и не позволяет оценить условную функцию надёжности. Поэтому в 1981 году Бераном была предложена оценка, обобщающая оценку Каплана-Мейера и имеющая вид [10]:

$$\tilde{S}_{b_n}(t|x) = \prod_{Y_{(i)} \leq t} \left\{ 1 - \frac{W_{n(i)}(x; b_n)}{1 - \sum_{j=1}^{i-1} W_{n(j)}(x; b_n)} \right\}^{\delta_i}, \quad (1.13)$$

где x – значение ковариаты, для которой оценивается функция надёжности;

$W_{n(i)}(x; b_n), i = 1, \dots, n$ – веса Надарая–Ватсона, которые вычисляются по формуле [104]:

$$W_{n(i)}(x; b_n) = K\left(\frac{x - x_i}{b_n}\right) / \sum_{j=1}^n K\left(\frac{x - x_j}{b_n}\right), \quad (1.14)$$

где $K(\cdot)$ – ядерная функция, удовлетворяющая условиям регулярности:

$$K(y) = K(-y), \quad 0 \leq K(y) < \infty, \quad \int_{-\infty}^{\infty} K(y) dy = 1;$$

b_n – параметр размытости такой, что $\lim_{n \rightarrow \infty} b_n = 0$, $\lim_{n \rightarrow \infty} nb_n = \infty$.

Такая непараметрическая модель способна служить оценкой функции надёжности для наблюдений с ковариатами. Основная идея оценки Берана заключается в том, что каждому наблюдению присваивается свой вес в зависимости от близости значения ковариаты этого наблюдения к значению ковариаты, для которого строится оценка Берана. Чем это расстояние меньше, тем больше вес. При значениях весов Надарая–Ватсона $W_{n(i)}(x; b_n) = n^{-1}$, то есть когда всем наблюдениям присваивается одинаковый вес, оценка Берана сводится к оценке Каплана-Мейера.

Для расчёта весов Надарая–Ватсона используются ядерные функции. В данной работе рассматривались следующие функции:

а) Квартическая ядерная функция:

$$K(y) = \begin{cases} \frac{15}{16}(1-y^2)^2, & |y| \leq 1; \\ 0, & |y| > 1. \end{cases}$$

б) Ядерная функция Гаусса:

$$K(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, & |y| \leq 1; \\ 0, & |y| > 1. \end{cases}$$

в) Прямоугольная ядерная функция:

$$K(y) = \begin{cases} \frac{1}{2}, & |y| \leq 1; \\ 0, & |y| > 1. \end{cases}$$

г) Ядерная функция Епанечникова:

$$K(y) = \begin{cases} 0.75(1-y^2), & |y| \leq 1; \\ 0, & |y| > 1. \end{cases}$$

На рисунке 1.1 представлены графики рассмотренных ядерных функций.

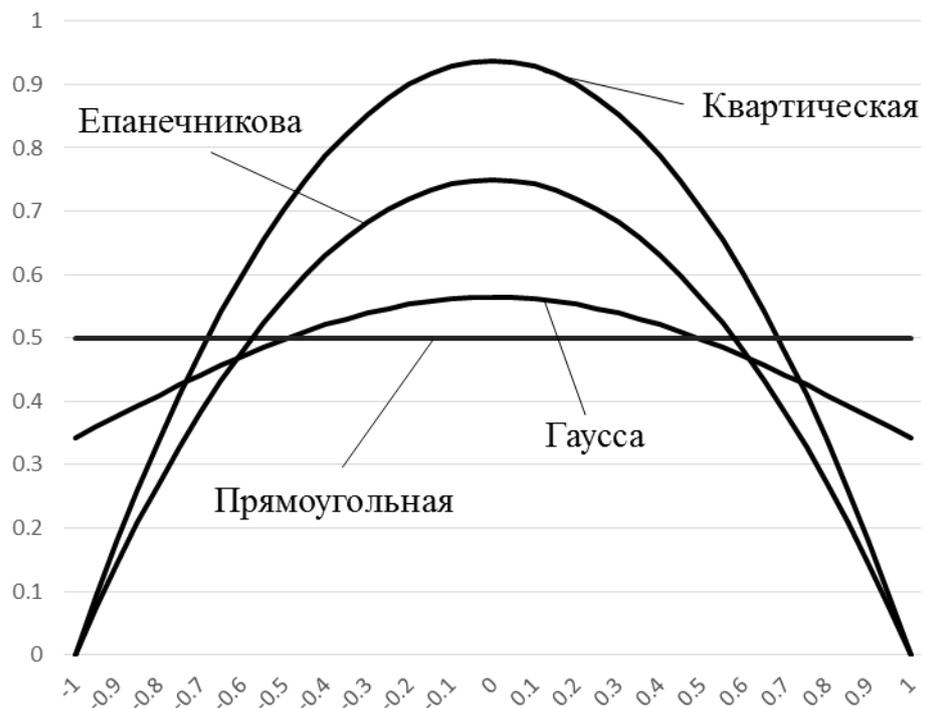


Рисунок 1.1 – Вид различных ядерных функций

Как известно, в ядерном сглаживании ключевую роль играет параметр размытости, от которого зависит то, по каким элементам выборки будет происходить сглаживание. При построении оценки Берана, главным образом, от выбора этого параметра зависит качество полученной модели.

1.3 Статистическое моделирование (метод Монте-Карло)

В данной работе большинство исследований выполнены с использованием метода Монте-Карло. Данный метод применяется для исследования статистических свойств непараметрических оценок условной функции надёжности, а также для исследования распределений статистик и мощности критериев согласия.

Методика компьютерного моделирования статистических закономерностей предусматривает статистическое моделирование эмпирических распределений статистик, вычисляемых по выборкам псевдослучайных величин, построение аналитических моделей, наилучшим образом сглаживающих (выравнивающих) полученные эмпирические распределения, уточнение построенных моделей по серии экспериментов.

Рассмотрим использование метода Монте-Карло на примере проверки статистической гипотезы о виде распределения. Моделирование распределения статистики критерия согласия при проверке сложной гипотезы определяется следующей последовательностью действий.

Выбор требуемого количества экспериментов N производится на основании заданной доверительной вероятности γ . В частности, для приближения функции $F(t)$ эмпирической функцией распределения $F_N(t)$ с погрешностью моделирования, не превышающей величину ε , потребуется [93]

$$N = \left[t_\gamma^2 \frac{F(t)(1-F(t))}{\varepsilon^2} \right] + 1$$

повторений, где $[\cdot]$ означает целую часть числа, $t_\gamma = \Phi^{-1}\left(\frac{\gamma+1}{2}\right)$ – квантиль стандартного нормального распределения. Абсолютная погрешность моделирования зависит от величины $F(t)$ и достигает максимума при $F(t)=0.5$. Поэтому необходимое количество повторений можно оценить следующим неравенством:

$$N \leq \left[t_\gamma^2 \frac{1}{4\varepsilon^2} \right] + 1.$$

Так для $\varepsilon=0.01$ и доверительной вероятности $\gamma=0.99$ необходимое количество экспериментов составит величину $N=16600$. В основе метода Монте-Карло лежит использование последовательностей псевдослучайных величин, подчиняющихся различным законам распределения. При этом качество результатов моделирования зависит и от свойств датчика, генерирующих псевдослучайные величины, равномерно распределённые на интервале $[0, 1]$. Поэтому важно применять датчики псевдослучайных чисел, обладающие хорошими свойствами.

В настоящей диссертационной работе при выполнении исследований использовался датчик «WELL» (Well Equidistributed Long-period Linear) в модификации WELL512, который имеет период $2^{512}-1$ и обладает высокой производительностью [50].

1.4 Влияние выбора параметра размытости на свойства оценки Берана

В настоящей работе для исследования статистических свойств непараметрических оценок и исследования распределений статистик и мощности критериев согласия использовалась методика статистического моделирования. Методика компьютерного моделирования статистических закономерностей предусматривает статистическое моделирование

эмпирических распределений статистик, вычисляемых по выборкам псевдослучайных величин [30].

На примерах ниже показана зависимость оценки Берана от параметра размытости. В качестве расстояния между истинной функцией надёжности и оценкой Берана рассчитывалась статистика типа Колмогорова вида:

$$D_{b_n} = \sup_{j=1..m, t < \infty} \left| \tilde{S}_{b_n}(t | x_j) - S_{x_j}(t) \right|, \quad (1.15)$$

где n – объём выборки;

m – количество групп в плане эксперимента;

$\tilde{S}_{b_n}(t | x_j)$ – значение оценки Берана в точке t ;

$S_{x_j}(t)$ – истинное значение функции надёжности в точке t при заданном значении ковариаты x_j .

В качестве истинной модели была выбрана модель пропорциональных интенсивностей Кокса с базовой функцией надёжности, соответствующей экспоненциальному распределению. Моделировались выборки объёмом 100 элементов, точки спектра плана выбирались случайно из равномерного распределения на отрезке $[0,1]$. Для каждой выборки строилась оценка Берана с заданным значением параметра размытости, после чего оценивалось расстояние (1.15), которое усреднялось по $N=16600$ экспериментам. Из таблицы 1.3 видно, что существует оптимальное значение параметра размытости, при котором точность оценки максимальна.

Таблица 1.3 – Зависимость оценки Берана от параметра размытости, $n = 100$, случайный план.

Значение параметра размытости b_n	Среднее отклонение \bar{D}_{b_n}
0.1	0.41
0.2	0.33
0.3	0.35
0.4	0.39
0.5	0.43
0.6	0.47
0.7	0.50
0.8	0.53
0.9	0.56

В частности, видно, что при значении параметра размытости равном 0.2 отклонение вида (1.15) принимало наименьшее значение (среди рассмотренных). Таким образом, можно сделать вывод, что существует оптимальное значение параметра размытости, при котором среднее отклонение оценки Берана от истинной функции надёжности минимально.

Ниже продемонстрировано влияние значения параметра размытости на оценку Берана. На рисунке 1.2 приведены условная функция надёжности, соответствующая модели (1.6) со значением ковариаты $x = 0.2$, оценка Берана с параметром размытости $b_n = 0$ и оценка Берана с параметром $b_n = 0.1$. Оценки Берана были построены на основе выборки, смоделированной в соответствии с моделью пропорциональных интенсивностей Кокса (1.6) с базовой функцией надёжности, соответствующей экспоненциальному распределению, объёмом $n = 100$, опорные точки плана эксперимента выбирались случайно из равномерного распределения на отрезке $[0,1]$.

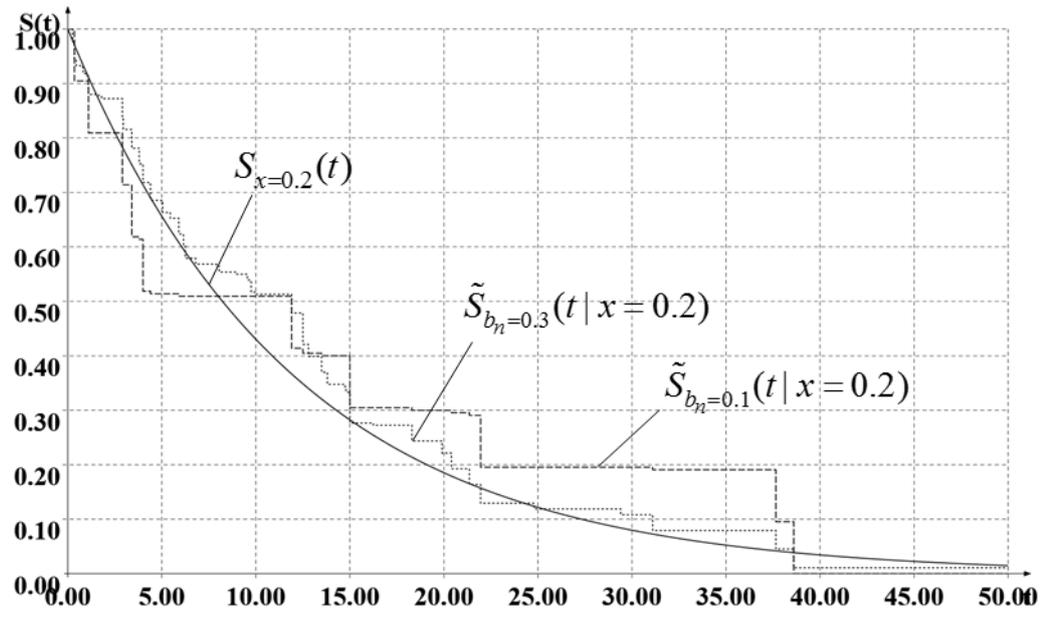


Рисунок 1.2 – Теоретическая функция надёжности и оценка Берана при различных значениях параметра размытости

Из рисунка 1.2 видно, как сильно меняется оценка Берана в зависимости от значения параметра размытости.

Таким образом, исходя из таблицы 1.3 и рисунка 1.2, можно сделать вывод, что для каждой выборки существует оптимальный параметр размытости, при котором оценка Берана максимально точно описывает условную функцию надёжности.

1.5 Факторы, оказывающие влияние на значение оптимального параметра размытости

Оптимальный параметр размытости зависит от нескольких факторов: от плана эксперимента, от числа групп, от объёма выборки, от степени влияния ковариаты на функцию надёжности, выбора ядерной функции, от того, для какого значения ковариаты строится условная функция надёжности. Здесь и далее, в качестве оптимального значения параметра размытости принимается то, при котором достигается максимальная точность оценки Берана:

$$b_{pp}^j = \arg \min_{b_{pp}^j} \left(\sup_{t < \infty} \left| \tilde{S}_{b_{pp}^j} (t | x_j) - S_{x_j} (t) \right| \right). \quad (1.16)$$

Для демонстрации зависимости оптимального параметра b_{pp}^j от n моделировались выборки объёмов $n = 10, 50, 100, 250$ из некоторой известной модели, после чего для каждой выборки вычислялось оптимальное значение параметра размытости (1.16).

В таблице 1.4 представлены значения усреднённого оптимального параметра b_{pp}^j по $N=16600$ выборкам каждого объёма.

В качестве истинной модели рассматривалась модель Кокса с логарифмически линейной функцией от ковариат и параметром $\beta = 2$, с логнормальным базовым распределением с параметрами $\theta_1 = 21.5$, $\theta_2 = 1.6$. Значения ковариаты моделировались в соответствии с равномерным распределением на интервале $[0, 1]$. Оценка Берана считалась для значения ковариаты $x = 0.5$.

Таблица 1.4 – Зависимость оптимального параметра размытости b_{pp}^j от объёма выборки, случайный план, ядро Епанечникова

Объём выборки	Усреднённый оптимальный параметр
10	0.55
50	0.43
100	0.36
250	0.29

Как показано в таблице 1.4, с увеличением объёма выборки средний оптимальный параметр размытости b_{pp}^j уменьшается. Данную зависимость можно объяснить тем, что с увеличением объёма выборки число наблюдений с

близкими значениями ковариаты увеличивается, поэтому и средний оптимальный параметр уменьшается, чтобы в построении оценки участвовали наиболее “близкие” (по значению ковариаты) наблюдения. Для демонстрации зависимости оценки Берана от количества опорных точек в плане проведено аналогичное исследование. В качестве истинной модели также взята модель Кокса, только на этот раз значение оптимального параметра b_{pp}^j рассматривалось при фиксированном объёме выборки $n=100$ и различном числе групп: $m=4, 10, 20, 50, 100$. Результаты представлены в таблице 1.5.

Таблица 1.5 – Зависимость оптимального параметра размытости b_{pp}^j от числа групп, ядро Епанечникова

Количество групп	Усреднённый оптимальный параметр
4	0.55
10	0.43
20	0.36
50	0.9
100	0.9

Из таблицы 1.5 видно, что оптимальный параметр размытости увеличивается с ростом количества групп (числа опорных точек плана эксперимента).

В таблице 1.6 приведены значения среднего оптимального параметра размытости для прямоугольной ядерной функции в зависимости от объема выборки.

Сравнивая результаты, представленные в таблицах 1.4 и 1.6, можно увидеть, что при использовании прямоугольной ядерной функции значение среднего оптимального параметра меньше чем при ядерной функции Епанечникова.

Таблица 1.6 – Зависимость оптимального параметра размытости b_{pp}^j от объёма выборки, случайный план, прямоугольная ядерная функция

Объём выборки	Усреднённый оптимальный параметр
10	0.37
50	0.30
100	0.25
250	0.25

Кроме того, оптимальный параметр размытости зависит от того, насколько сильно ковариата влияет на распределение отказов, то есть от того, как сильно будет меняться функция надёжности при изменении значения ковариаты. Степень влияния ковариаты на функцию надёжности рассчитывалась в следующем виде:

$$R = \sup_t |S(t | x = 1) - S(t | x = 0)|. \quad (1.17)$$

В таблице 1.7 представлены усредненные по $N = 2000$ выборкам объёма $n = 100$ значения оптимального параметра размытости b_{pp}^j при различных значениях регрессионного параметра модели Кокса β . Большим значениям параметра β соответствует большее значение R . Опорные точки плана моделировались в соответствии с равномерным распределением на отрезке $[0, 1]$.

Таблица 1.7 – Зависимость оптимального параметра размытости b_{pp}^j от степени влияния ковариаты на функцию надёжности

R	Усредненный оптимальный параметр
0.382	0.714
0.682	0.491
0.866	0.338
0.975	0.235
0.997	0.194
0.999	0.156

Ниже зависимость от степени влияния ковариаты на функцию надёжности демонстрируется на следующем примере. Рассмотрим две выборки времен отказов Y_1 и Y_2 с ковариатой, принимающей значения 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, и объемом $n=150$ (по 15 наблюдений для каждого значения ковариаты). Выборки смоделированы с помощью метода обратной функции в соответствии с моделью ускоренных испытаний, где функция от ковариаты $r(x, \beta) = e^{\beta x}$. Выборка Y_1 моделировалась при значении параметра $\beta = 2$, выборка Y_2 – при $\beta = 5$. На рисунке 1.3 представлены оценки Берана для функции надёжности при $x=0.1$ и $x=0.6$, полученные по выборке Y_1 . Для сравнения также приведены соответствующие истинные функции надёжности. Оценки Берана, построенные по выборке Y_2 , и соответствующие истинные функции надёжности приведены на рисунке 1.3. Параметр размытости b_n при построении оценки Берана в обоих случаях взят равным 0.5. В качестве меры отклонения оценки Берана от истинной функции надёжности при заданном значении ковариаты рассматривается статистика вида

$$D_n = \frac{1}{n} \sum_{i=1}^n \left| \tilde{S}_{b_n}(Y_i | x) - S(Y_i | x) \right|. \quad (1.18)$$

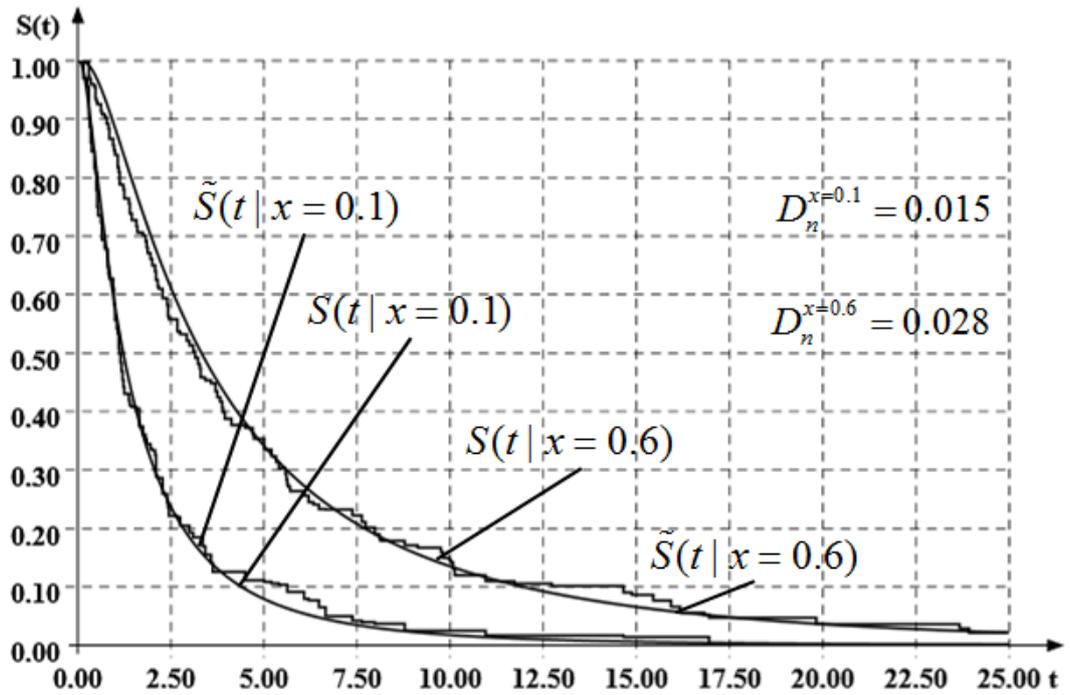


Рисунок 1.3 – Функции надёжности и оценки Берана, $\beta = 2$

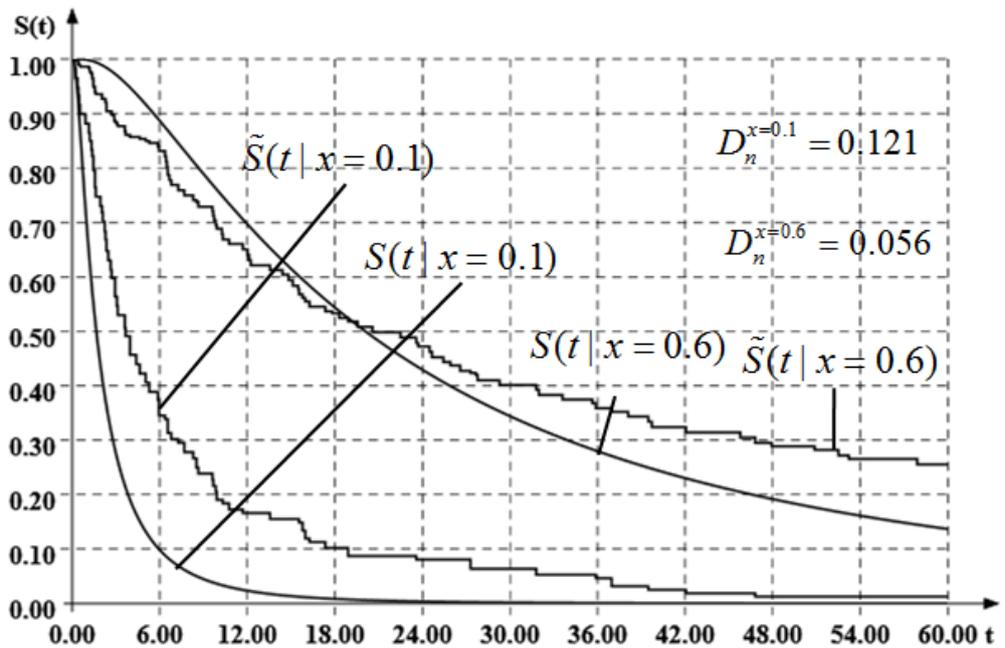


Рисунок 1.4 – Функции надёжности и оценки Берана, $\beta = 5$

Как видно из рисунка 1.3, оценки Берана достаточно близки к соответствующим функциям надежности. Однако при таком же плане эксперимента в случае большего воздействия x (при большем значении параметра β), как показано на рисунке 1.4, наблюдается существенное отклонение оценок Берана от истинных функций надёжности.

Методами статистического моделирования проведено аналогичное исследование влияния на оптимальный параметр размытости значения ковариаты, для которой строится оценка Берана. Ниже данная зависимость иллюстрируется на примере. В качестве истинной модели взята модель пропорциональных интенсивностей с функцией от ковариат $r(x; \beta) = \ln(1 + e^{\beta x})$ и логнормальной базовой функцией распределения с параметрами $\theta_1 = 21.5$, $\theta_2 = 1.6$. На рисунках 1.5 и 1.6 в зависимости от значений ковариаты представлены усреднённые по $N = 2000$ экспериментам значения оптимального параметра размытости при различных объёмах выборки. В первом случае ковариата принимала 4 значения и спектр плана эксперимента имеет вид: $\text{supp}(\xi_n^m) = \{0, 0.33, 0.67, 1\}$, во втором $\text{supp}(\xi_n^m) = \{0, 0.11, 0.22, 0.33, 0.44, 0.56, 0.67, 0.78, 0.89, 1\}$, размеры выборок $n = 20, 40, 80, 120, 200, 300, 400$, количества наблюдений в группах равны.

Из рисунков 1.5 и 1.6 видно, что оптимальный параметр размытости зависит от того, для какой условной функции надёжности строится оценка. Оптимальные значения параметра размытости зависят от значений ковариаты. Для различных групп эти различия могут быть значимыми. Например, оптимальное значение параметра размытости для крайних групп (0 и 1) отличаются на 80-100% от значений параметра для центральных групп. В случае четырёх групп разница между значениями параметра для соседних групп может меняться в интервале от 3 до 70%. В случае 10 групп разница для соседних групп колеблется в интервале от 3 до 21%. Оптимальные значения для

параметра размытости соседних значений ковариаты ближе друг к другу для соседних групп, так как условные функции надёжности для таких значений также близки.

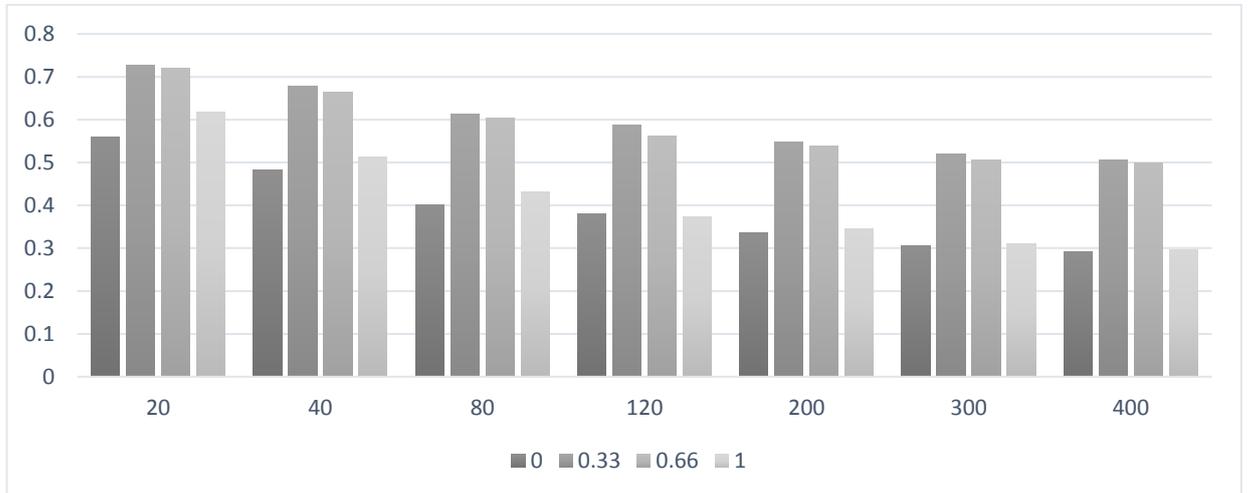


Рисунок 1.5 – Значения оптимального параметра размытости b_{pp}^j в зависимости от значения ковариаты (для 4 групп) и объёма выборки

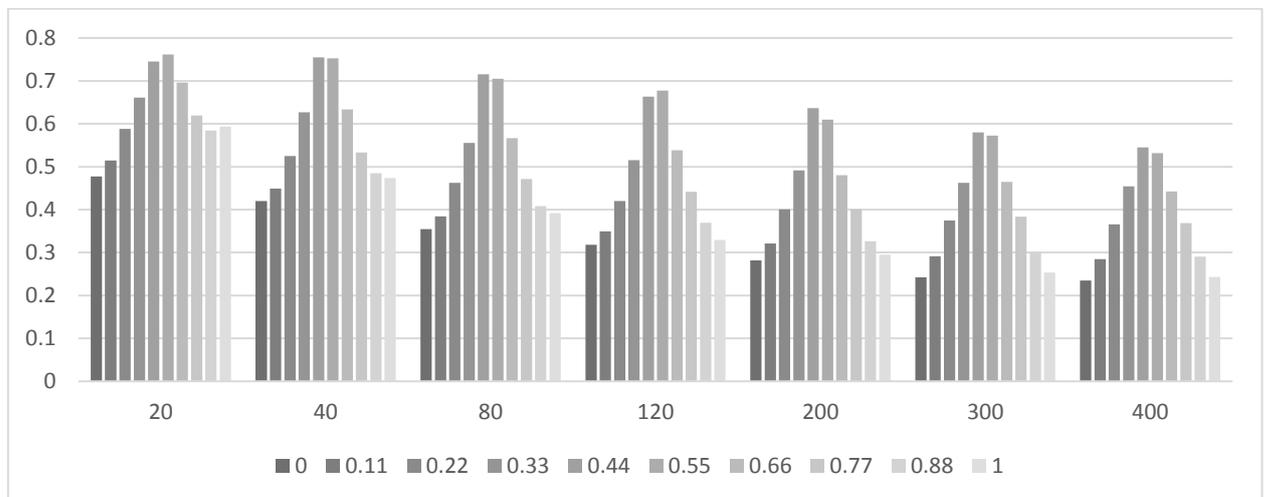


Рисунок 1.6 – Значения оптимального параметра размытости b_{pp}^j в зависимости от значения ковариаты (для 10 групп) и объёма выборки

Из рисунков 1.5 и 1.6 можно заметить, что оптимальные значения параметров распределены симметрично для равномерного плана. Таким образом, достаточно подобрать оптимальные значения параметров размытости только для половины групп. Подтверждается также зависимость оптимального

значения параметра размытости от числа групп в плане. Например, для объёма выборки 80 и четырёх групп оптимальное значение параметра размытости для значения ковариаты 0.33 равно 0.61. Однако для того же объёма выборки, но для 10 групп оптимальное значение параметра равно 0.55.

Таким образом, выбор оптимального значения параметра размытости должен осуществляться с учётом объёма выборки, числа групп, степени влияния ковариаты на функцию надёжности, а также значения ковариаты, для которой строится оценка Берана.

1.6 Методы выбора оптимального значения параметра размытости

Существует множество различных методов для выбора оптимального значения параметра размытости. В качестве наиболее популярных можно указать:

- Метод кросс-проверки, который чаще всего рассматривается как наиболее точный, однако требующий значительных вычислительных затрат [76].
- Метод минимума средней интегральной ошибки, в соответствии с которым параметр размытости вычисляется по формуле [104]:

$$b_{NS} = \left[\frac{8\pi^{1/2}R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \hat{\sigma},$$

где $\mu_2(K) = \int y^2 K(y) dy$; $R(K) = \int K^2(y) dy$; $\hat{\sigma}$ – оценка среднеквадратического отклонения, которую можно вычислять различными способами, но чаще всего используется несмещенная выборочная дисперсия.

- Референтный выбор параметра размытости. Преимуществом данного метода является вычислительная простота расчёта, однако он является весьма неточным и обычно применяется для выбора начального приближения.

Параметр размытости для оценки Берана определяет только вес наблюдения. По сути, при построении оценки Берана процедура ядерного сглаживания отсутствует как таковая, т.к. не производится сглаживание (восстановление) регрессионной зависимости отклика от ковариаты. Поэтому метод кросс-проверки в данном случае неприменим. По той же причине оказывается неприменимым и метод минимума средней интегральной ошибки, в соответствии с которым необходимо вычислять функции $\mu_2(K)$ и $R(K)$, что не представляется возможным. Кроме того расчёт $\hat{\sigma}$ крайне затруднителен для неслучайного плана.

Пожалуй, единственным методом, который можно применить для выбора параметра размытости оценки Берана, является референтный метод. Сложность применения такого подхода для оценки Берана была описана в ряде работ. Например, в [59] предлагается теоретическая оценка для оптимального значения параметра размытости на основе референтного метода:

$$b_{n,opt} = Cn^{-1/5},$$

$$C = \left(\frac{\Gamma(t, t | x)}{4\bar{b}^2(t | x)} \right)^{1/5},$$

$$\bar{b}(t | x) = \frac{1}{2} \mu_2^K (1 - F(t | x)) \int_0^t \left\{ \frac{\ddot{H}(s | x) dH^u(s | x)}{(1 - H(s | x))^2} + \frac{d\ddot{H}^u(s | x)}{1 - H(s | x)} \right\} ds,$$

где $H(s | x)$ – функция распределения наблюдаемой случайной величины $Y_x = \min(T_x, C_x)$,

$H^u(s | x) = P\{T_x < s | \delta = 1\}$ – условная функция распределения отказов.

Однако, как далее пишет сам автор, на практике такую формулу использовать нельзя, так как функции $H(t | x)$, $\ddot{H}(t | x)$, $H^u(t | x)$, $\ddot{H}^u(t | x)$ неизвестны. Их

можно попробовать оценить, однако для оценки требуется введение ещё несколько параметров сглаживания и разработка метода для выбора значения и этих параметров. Таким образом, предложенным методом пользоваться не представляется возможным. Автор предложил использовать метод кросс-проверки для определения данных параметров. Однако, во-первых, процедура кросс-проверки для четырёх функций крайне трудоёмка, а, во-вторых, это не приводит к положительному результату [15].

В [20] предложено оценивать оптимальное значение параметра сглаживания по схожей формуле:

$$b_{n,opt} = Cn^{-1/5}.$$

Однако алгоритм подбора константы C в работе не приведен. Говорится только, что данная константа может быть получена эмпирически.

Таким образом, на данный момент референтный метод является единственным способом выбора оптимального параметра размытости. Очевидно, в нём есть рядов минусов. Во-первых, это отсутствие какого-либо алгоритма для подбора константы C . Поэтому, в качестве константы в дальнейших исследованиях возьмём единицу. Понятно, что данная формула взята из ядерного сглаживания, где число одинаковых наблюдений обычно крайне мало. Поэтому второй минус заключается в том, что данный алгоритм предназначен только для случайного плана. В-третьих, данный подход не учитывает число групп и степень влияния значения ковариаты на функцию надёжности. В-четвертых, данный метод не позволяет выбрать значение параметра размытости для конкретного значения ковариаты, т.е. не является адаптивным.

Таким образом, актуальной задачей является разработка алгоритма выбора оптимального значения параметра размытости для оценки Берана, который бы удовлетворял следующим требованиям:

- алгоритм должен быть адаптивным, и способным находить оценку оптимального значения параметра размытости для любого значения ковариаты;
- алгоритм должен работать для любого числа опорных точек плана эксперимента, в том числе для случайного плана;
- алгоритм должен учитывать особенности используемой модели, в том числе степень влияния ковариаты на функцию надёжности;
- алгоритм должен учитывать объём выборки.

1.7 Выводы

В данной главе проведен обзор основных понятий и определений, используемых в диссертационной работе. Рассмотрены параметрический и непараметрический подходы к оцениванию функции надёжности.

Методами статистического моделирования проведено исследование точности непараметрической оценки Берана для функции надёжности, в результате которого показано, что:

- основным фактором, влияющим на точность оценки Берана, является выбор параметра размытости;
- на выбор оптимального значения параметра размытости оказывают влияние следующие факторы: число опорных точек в плане эксперимента, выбор ядерной функции, степень влияния ковариаты на функцию надёжности, значение ковариаты, для которой строится оценка Берана.

Таким образом, показана актуальность разработки адаптивного алгоритма для выбора оптимального значения параметра размытости для оценки Берана, а также разработки критериев согласия для проверки адекватности параметрических регрессионных моделей надёжности.

ГЛАВА 2 Разработка алгоритмов выбора оптимального параметра размытости

В данной главе предлагается адаптивный алгоритм выбора оптимального параметра размытости, который учитывает особенности плана эксперимента и степень влияния ковариаты на функцию надёжности. Методами статистического моделирования исследуется точность оценки Берана, строящейся с использованием предложенного алгоритма для выбора оптимального значения параметра размытости.

Результаты исследований, представленные в данной главе, опубликованы в [11, 17, 18, 74, 75, 76].

2.1 Разработка адаптивного алгоритма выбора оптимального значения параметра размытости

Как было показано в главе 1 настоящей диссертации, ключом к построению «качественной» непараметрической регрессионной модели надёжности с помощью оценки Берана является выбор оптимального значения параметра размытости. В данной главе предлагается алгоритм выбора оптимального значения параметра размытости b_n , основанный на минимизации среднего отклонения времен отказов Y_1, Y_2, \dots, Y_n от непараметрической оценки обратной функции надёжности $S_x^{-1}(p)$. Разработанный метод позволяет учесть объём выборки, число групп, степень влияния ковариаты на функцию надёжности. Кроме того, метод является адаптивным, то есть позволяет подобрать оценку параметра размытости для произвольного значения ковариаты. Суть метода заключается в следующем.

Обозначим обратную функцию надёжности через $g(p|x)$. Тогда модель (1.4) можно переписать в следующем виде:

$$T_x = g(p|x) + \varepsilon, \quad (2.1)$$

где $p \in (0,1)$; ε – ошибка наблюдения, которая в общем случае может зависеть от p и x .

Для такой постановки задачи можно построить ядерную оценку для модели (2.1), имеющую вид:

$$\hat{g}(\hat{p}_{b_n}(Y_i|x)) = \frac{1}{n} \sum_{i=1}^n \omega_i(\hat{p}_{b_n}(Y_i|x)) \cdot Y_i, \quad (2.2)$$

где ω_i – вес i -го наблюдения, для вычисления которого можно использовать различные весовые функции. Вероятности $\hat{p}_{b_n}(Y_i|x)$ вычисляются с использованием оценки Берана по формуле:

$$\hat{p}_{b_n}(Y_i|x) = \tilde{S}_{b_n}(Y_i|x).$$

Найти оптимальное значение параметра размытости можно в результате решения задачи вида:

$$b_{opt}^j = \arg \min_{b_n} \sum_{i=1}^{n^*} \delta_i \cdot \left| \hat{g}(\hat{p}_{b_n}(Y_i|x)) - Y_i \right|. \quad (2.3)$$

где n^* – количество наблюдений участвующих при построении оценки, в зависимости от параметра b_n .

Таким образом, при изменении параметра размытости b_n изменяется и оценка Берана, которая в свою очередь является параметром для ядерной оценки обратной функции надёжности. Минимизируя разницу между оценкой обратной функции надёжности и временами отказов, которые нам известны, получаем оптимальное значение параметра размытости.

2.1.1 Ядерная оценка обратной функции надёжности

В предложенном подходе используются методы ядерного сглаживания для вычисления оценки (2.2), поэтому необходимо провести ряд исследований, чтобы выбрать наиболее предпочтительную ядерную функцию, весовую

функцию, а также способ выбора параметра сглаживания (параметр сглаживания входит в весовую функцию).

Задача выбора параметра сглаживания значительно легче выбора оптимального значения параметра размытости, в виду того, что мы находимся в контексте ядерного сглаживания, и можем использовать любой из разработанных методов выбора оптимального параметра сглаживания для ядерной оценки регрессии.

Считается [99, 104], что вид ядерной функции не оказывает значимого влияния на ядерную оценку. Однако, как показано выше, от вида ядерной функции зависит оптимальное значение параметра размытости, а от значения параметра размытости зависит оценка Берана. Таким образом, вид ядерной функции влияет на точность оценки. В данном диссертационном исследовании рассмотрены ядерные функции Епанечникова, Гаусса, четвертая и прямоугольная.

Веса ω_i могут рассчитываться различными способами. В данной работе рассматривались веса первого порядка Надарая – Ватсона

$$\omega_j^{(1)}(\hat{p}_i) = W_n^j(\hat{p}_i; h_n) = K\left(\frac{\hat{p}_i - \hat{p}_j}{h_n}\right) / \sum_{k=1}^n K\left(\frac{\hat{p}_i - \hat{p}_k}{h_n}\right)$$

и веса второго порядка Пристли – Чао [104]:

$$\omega_j^{(2)}(\hat{p}_i) = n(\hat{p}_{(i)} - \hat{p}_{(i-1)}) K\left(\frac{\hat{p}_i - \hat{p}_j}{h_n}\right),$$

где h_n – параметр сглаживания.

Для выбора оптимального параметра сглаживания h_n можно воспользоваться уже разработанными подходами для выбора оптимального параметра сглаживания ядерной оценки регрессии. Однако есть ряд требований, которым должен удовлетворять метод. Во-первых, при его использовании должно находиться оптимальное значение параметра

размытости. То есть конечная цель применения метода выбора оптимального параметра сглаживания – это не получение точной ядерной оценки обратной функции надёжности, а получение точной оценки Берана. Во-вторых, он должен удовлетворять условиям быстрогодействия, так как функция, в которой он используется, будет многократно вызываться. И, в-третьих, он должен быть устойчивым к выбросам, так как конечный алгоритм выбора оптимального значения параметра размытости должен работать с различными моделями. Также необходимо помнить, что здесь в качестве регрессионного параметра будет выступать оценка Берана, что также может наложить ограничения на методы.

Рассматривались следующие методы выбора параметра сглаживания h_n .

– Метод кросс-проверки, который чаще упоминается как наиболее точный, однако требующий значительных вычислительных затрат [76, 104]. Так как для того, чтобы получить оптимальный параметр сглаживания необходимо минимизировать функцию вида:

$$CV(h) = n^{-1} \sum_{j=1}^n \left[Y_j - \hat{m}_{h,j}(\hat{p}_j) \right]^2 \omega(\hat{p}_j),$$

где

$$\hat{m}_{h,j}(\hat{p}_j) = n^{-1} \sum_{i \neq j} W_{h,i}(\hat{p}_j) Y_i.$$

– Метод минимума средней интегральной ошибки, в соответствии с которым параметр сглаживания вычисляется по формуле

$$h_{NS} = \left[\frac{8\pi^{1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \hat{\sigma},$$

где $\mu_2(K) = \int y^2 K(y) dy$, $R(K) = \int K^2(y) dy$, $\hat{\sigma}$ – оценка среднеквадратического отклонения, которую можно вычислять различными способами, чаще всего для этого используется несмещенная выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{p}_i - \bar{\hat{p}})^2. \quad (2.4)$$

Однако оценка дисперсии (2.4), во-первых, не является робастной, а во-вторых, обладает «хорошими» статистическими свойствами лишь в случае законов распределения, близких к нормальному. Поэтому в данной работе будем рассматривать также робастную оценку среднеквадратического отклонения вида:

$$\hat{\sigma}_{robust} = 1.4826 med_{i=1..n} \left| \hat{p}_i - med_{j=1..n, k=j+1..n} \left(\frac{\hat{p}_j + \hat{p}_k}{2} \right) \right|. \quad (2.5)$$

Такая оценка представляет собой комбинацию известной робастной оценки Ходжеса-Леманна (Hodges-Lehmann) для параметра сдвига и робастной оценки Русьё (Rousseeuw), называемой медианой абсолютных отклонений, для параметра масштаба [55].

Выбор оценки Ходжеса-Леманна вместо медианы обусловлен следующими соображениями. Если половина наблюдений равна 0, то оценка параметра сдвига с использованием медианы также оказывается равной 0, что приводит конечном итоге к нулевому значению оценки среднеквадратического отклонения и, следовательно, к нулевому значению параметра размытости. Кроме того, данная оценка обладает меньше дисперсией. Ниже в таблице 2.1 приведены оценки дисперсии для медианы (med) и оценки Ходжеса-Леманна (HL) для выборок различных объёмов из равномерного распределения.

Таблица 2.1 – Оценки дисперсии для медианы и оценки Ходжеса-Леманна

Объём выборки	20	40	80	120	200
med	0.0110	0.0060	0.0030	0.0021	0.0012
HL	0.0049	0.0024	0.0011	0.0008	0.0004

2.1.2 Исследование свойств непараметрической оценки обратной функции надёжности

Из рассмотрения пришлось сразу исключить метод кросс-проверки, так как временные затраты при её использовании слишком велики. В целом, время расчёта оценки Берана увеличилось примерно в шесть раз при неслучайном плане и малом объёме выборок, при случайном же плане это время возрастет ещё в несколько раз. Применение метода кросс-проверки для выбора параметра сглаживания не позволит использовать оценку Берана для построения критерия согласия. Причина заключается в том, что время, потраченное на моделирование распределения статистики критерия, необходимого для принятия решения о результатах проверки соответствующей гипотезы, на современной элементной базе может занять несколько дней. Кроме вычислительных сложностей кросс-проверка обладает ещё одним недостатком, не позволяющим её использовать для решения задачи: кросс-проверка минимизирует оценку обратной функции надёжности по параметру сглаживания, а не по параметру размытости, и, как следствие, параметр размытости вырождается в 0, что не является оптимальным значением. Таким образом, полученные с помощью метода кросс-проверки оценки параметра сглаживания не позволили бы получить более точную оценку Берана.

Исследование свойств получаемых оценок проводилось с помощью методов статистического моделирования. В качестве отклонения оценки Берана от истинной условной функции надёжности использовалась статистика:

$$D_{avg}(b) = \sup_{j=1..n} \left| \left(\tilde{S}_b(Y_j | x_j) - S(Y_j | x_j) \right) \right|. \quad (2.6)$$

Результаты исследования представлены на примере модели надёжности, соответствующей параметрической модели пропорциональных интенсивностей Кокса [13]:

$$S_x(t) = (S_0(t))^{r(x;\beta)} \quad (2.7)$$

с функцией от ковариат вида $r(x;\beta) = \ln(1 + e^{\beta x})$ и базовой функцией надёжности, соответствующей логнормальному распределению с функцией плотности

$$f_0(t) = \frac{1}{\sqrt{2\pi\theta_1 t}} \exp\left(-\frac{1}{2\theta_1^2} \ln^2\left(\frac{t}{\theta_2}\right)\right) \quad (2.8)$$

с параметрами $\theta_1 = 21.5$, $\theta_2 = 1.6$.

Для используемых при моделировании весовых функций, методов оценки дисперсии и истинного значения параметра регрессии β использовались следующие обозначения:

- 1 – веса Пристли-Чао, робастная оценка дисперсии (2.5), $\beta = 2$;
- 2 – веса Пристли-Чао, неробастная оценка дисперсии (2.4), $\beta = 2$;
- 3 – веса Надарая-Ватсона, робастная оценка дисперсии (2.5), $\beta = 2$;
- 4 – веса Надарая-Ватсона, неробастная оценка дисперсии (2.4), $\beta = 2$;
- 5 – веса Пристли-Чао, робастная оценка дисперсии (2.5), $\beta = 5$;
- 6 – веса Пристли-Чао, неробастная оценка дисперсии (2.4), $\beta = 5$;
- 7 – веса Надарая-Ватсона, робастная оценка дисперсии (2.5), $\beta = 5$;
- 8 – веса Надарая-Ватсона, неробастная оценка дисперсии (2.4), $\beta = 5$.

На рисунках 2.1 – 2.3 для указанных вариантов условий 1÷8 приведены усредненные значения отклонения (2.6), оптимального значения параметра размытости b_n^{opt} и значения параметра сглаживания h_{NS} , соответственно, при объемах выборок $n = 100, 200, 300$. Ковариата x принимала $m = 10$ значений, спектр плана эксперимента $\text{supp}(\xi_n^m) = \{0, 0.11, 0.22, 0.33, 0.44, 0.56, 0.67, 0.78, 0.89, 1\}$. Количество наблюдений, соответствующих различным значениям ковариаты, во всех экспериментах выбиралось одинаковым.

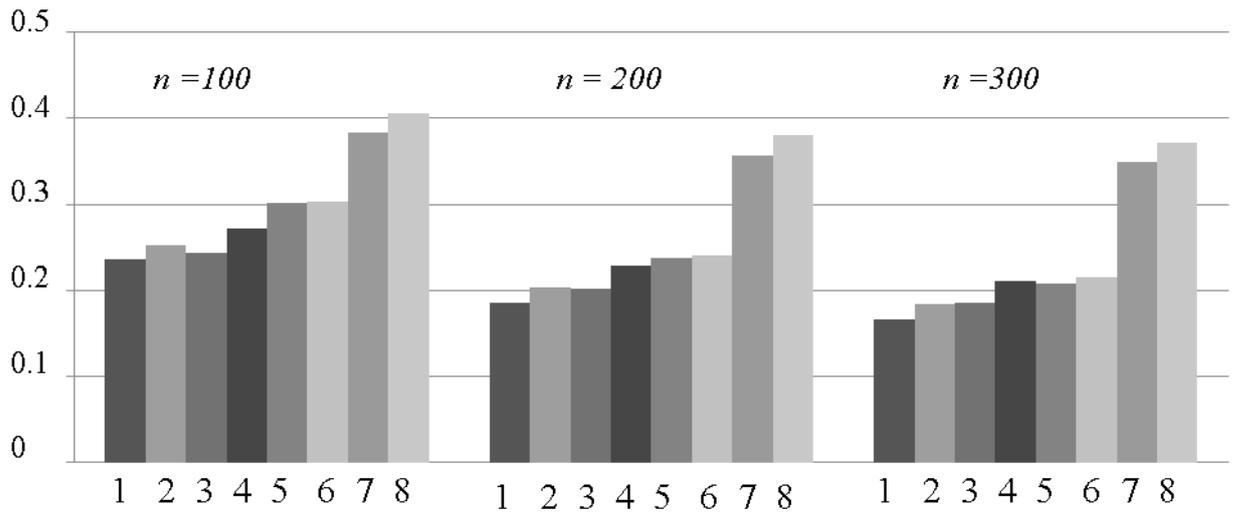


Рисунок 2.1 – Усреднённое отклонение оценки Бернана от заданной условной функции надёжности при различных объёмах выборки

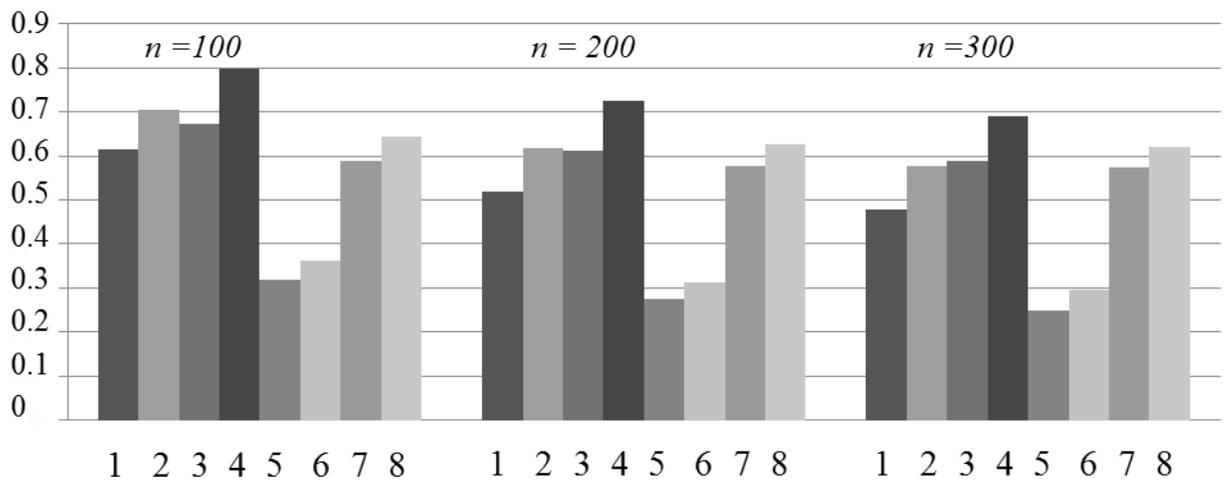


Рисунок 2.2 – Усреднённое значение параметра размытости b_n^{opt} при различных объёмах выборки

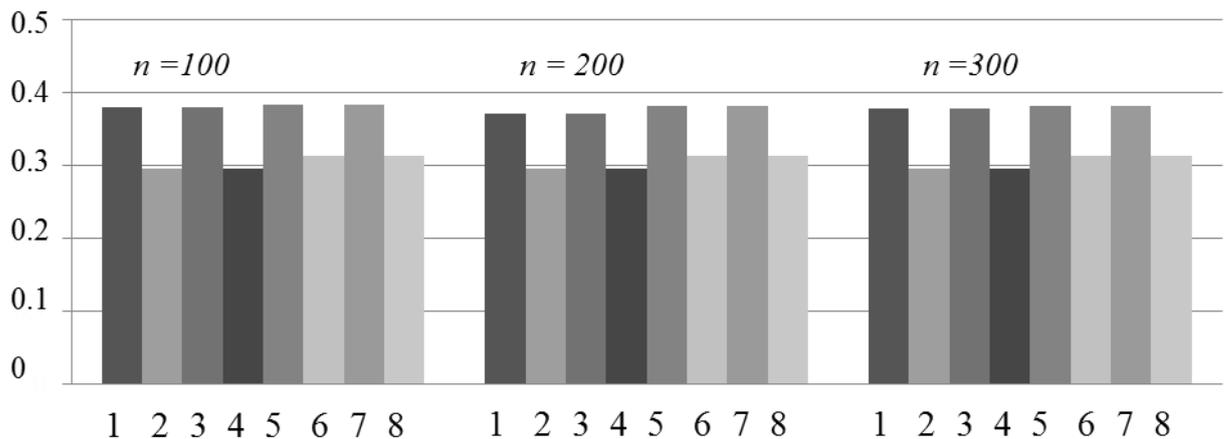


Рисунок 2.3 – Усреднённое значение параметра сглаживания h_{NS} при различных объёмах выборки

Из рисунка 2.1 видно, что использование весовой функции Пристли-Чао позволяет получить более точную оценку Берана. При объёме выборки $n = 100$ выигрыш в точности оценок Берана с использованием в выражении (2.2) весов Пристли-Чао составляет примерно 3% от средней величины отклонения (2.6) по сравнению с использованием весов Надарая-Ватсона, при $n = 200$ выигрыш составляет около 8%, а при $n = 300$ – 11%. Ухудшение точности оценки Берана при использовании весов Надарая-Ватсона может быть вызвано тем, что данная весовая функция не предполагает упорядочивания весов, в результате чего при построении оценки Берана используются «лишние» наблюдения и соответствующая дисперсия увеличивается. Использование робастной оценки дисперсии при вычислении параметра размытости повышает точность оценки Берана, которая также растёт с увеличением объёма выборки.

Из картины, представленной на рисунке 2.2, следует, что с увеличением объёма выборки параметр размытости b_n^{opt} уменьшается, что вполне логично, так как число наблюдений при разных значениях ковариаты увеличивается, а следовательно растёт и число «лишних» наблюдений, то есть таких наблюдений при использовании которых построение оценки приводит к

ухудшению точности. Важно отметить, что с увеличением параметра β в 2.5 раза (т.е. при увеличении влияния ковариаты на функцию надёжности), значение b_n^{opt} уменьшилось почти в два раза при использовании весов Пристли-Чоа. Однако в случае весов Надарая-Ватсона такой ситуации не наблюдается. Поэтому в данном случае точность оценок Берана с использованием весов Надарая-Ватсона существенно ухудшается. Как показано на рисунке 2.3, усреднённое значение параметра размытости практически не изменяется при увеличении объёма выборки. Такой результат согласуется с выводами, полученными в работе [82], где говорится о независимости значения параметра h_{NS} от вида весовой функции.

В рассматриваемых условиях дискретный план эксперимента будет зависеть от числа значений m , которые может принимать ковариата. Следующая часть исследований посвящена изучению зависимости искомой оценки ее свойств от m .

На рисунках 2.4-2.5 представлены аналогичные результаты исследования отклонения оценки Берана от заданной (моделируемой) условной функции надёжности и параметра сглаживания h_n^{opt} для объёма выборок $n = 200$ при различном числе значений ковариаты m .

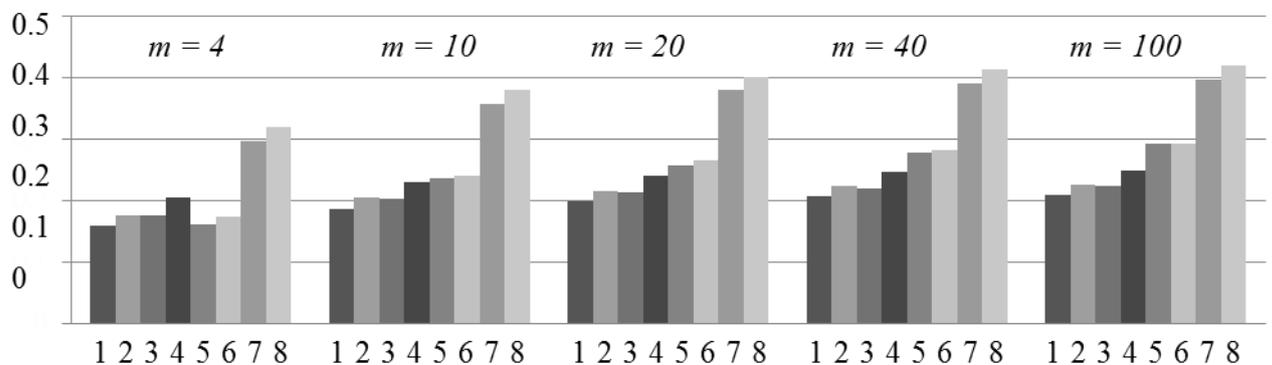


Рисунок 2.4 – Усреднённое отклонение оценки Берана от заданной условной функции надёжности при различном числе значений ковариаты

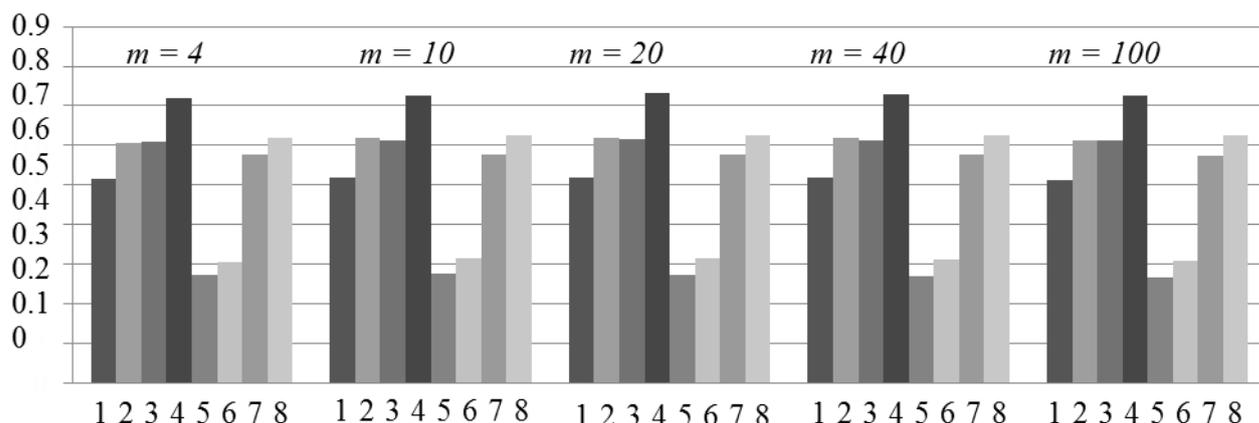


Рисунок 2.5 – Усреднённый параметр сглаживания h_n^{opt} при различном числе значений ковариаты

Как видно из рисунка 2.4, с увеличением числа возможных значений ковариаты точность оценки Берана снижается. В частности, при использовании весов Пристли-Чоа и робастной оценки дисперсии при $m=10$ относительно $m=4$ ухудшение составляет около 17%, а при $m=20$ относительно $m=10$ – 6%. Этот результат является вполне логичным, поскольку количество наблюдений, соответствующих фиксированному значению ковариаты, уменьшается, следовательно, объем информации о каждом наблюдении также становится меньше, что приводит к потере точности.

Анализ значений параметра сглаживания на основе рисунка 2.5 показывает, что оптимальные значения параметра практически не изменяются. Аналогичная ситуация наблюдалась при изучении усреднённых значений параметра сглаживания h_{NS} . Данное свойство оптимальных параметров размытости b_n^{opt} и сглаживания h_{NS} распространяется и на случай других значений регрессионного параметра.

Исследования свойств оценки Берана при различных условиях моделирования были проведены также для экспоненциальной модели Кокса.

Выявленные закономерности оказались такими же, поэтому конкретные численные результаты для этого случая опускаются.

В большинстве исследований, посвященных непараметрическим методам, отмечается, что выбор ядерной функции практически не влияет на точность получаемых оценок [99, 104]. Однако в данной постановке задачи это свойство может нарушаться. Для проверки этого предположения было проведено исследование точности оценки Берана при использовании различных ядерных функций при решении оптимизационной задачи (2.3) и оценивании параметра размытости. В качестве заданной модели была взята модель Кокса (2.7) с базовой функцией надёжности (2.8). Объём выборки $n=100$, опорные точки плана выбирались случайно из равномерного распределения на отрезке $[0,1]$. Результаты представлены в таблице 2.2.

Таблица 2.2 – Усреднённое отклонение (2.6) при использовании различных ядерных функций

b_{NS}	h_n	Квартическая	Епанечникова	Прямоугольная	Гаусса
Квартическая		0.20	0.20	0.21	0.23
Епанечникова		0.21	0.22	0.23	0.26
Прямоугольная		0.25	0.26	0.29	0.30
Гаусса		0.32	0.35	0.36	0.38

Как видно из таблицы 2.2, наилучшая точность оценки Берана достигается при использовании квартического ядра и ядра Епанечникова как при оценивании параметра сглаживания, так и при вычислении параметра размытости. Таким образом, изменение в постановке задачи, т.е. при применении ядерных функций для вычисления весовых коэффициентов в непараметрической оценке (1.13), приводит к тому, что выбор формы ядра оказывает существенное влияние на свойства оценок Берана. Косвенно такое

влияние обуславливается наличием взаимосвязи между параметрами размытости и сглаживания, определяющими значения оптимизируемого функционала в (2.3).

Любопытно, что наиболее ядерные функции, при которых достигнута наилучшая точность оценки Берана, являются более чувствительными к изменениям параметрам: как видно из рисунка 1.1, ядра Епанечникова и квадратическое сильнее меняют свои значения при изменении аргумента, чем более пологие ядра Гаусса и, тем более, прямоугольное.

Таким образом, адаптивный алгоритм выбора оптимального параметра размытости можно сформулировать следующим образом.

1. Нормировать значения ковариат:

$$x_i = \frac{x_i}{x_{\max}}, \quad i = \overline{1, n},$$

где $x_{\max} = \max(\text{supp}(\xi_n^m))$ – наибольшее значение ковариаты в выборке.

2. Задать $k = 1$.

3. Решить задачу одномерной оптимизации на интервале:

$$b_n^k = \arg \min_{b_n \in \left[\frac{k-1}{m}, \frac{k}{m} \right]} \sum_{i=1}^n \delta_i \cdot |\hat{g}(\hat{p}_i | x_i) - Y_i|,$$

где параметр размытости b_n входит в оценку Берана, которая является параметром в ядерной оценке обратной функции надёжности:

$$\hat{g}(\hat{p}_i | x_i) = \sum_{j=1}^n \omega_j(\hat{p}_i) \cdot Y_j$$

В качестве $\omega_j(\hat{p}_i)$ рекомендуется использовать веса Пристли-Чоа:

$$\omega_j^{(2)}(\hat{p}_i) = n(\hat{p}_{(i)} - \hat{p}_{(i-1)}) K\left(\frac{\hat{p}_i - \hat{p}_j}{h_{NS}}\right),$$

где для оценки параметра размытости h_{NS} используется метод минимума средней интегральной ошибки:

$$h_{NS} = \left[\frac{8\pi^{1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \hat{\sigma}_{robust}$$

с робастной оценкой среднеквадратического отклонения вида:

$$\hat{\sigma}_{robust} = 1.4826 \operatorname{med}_{i=1..n} \left| \hat{p}_i - \operatorname{med}_{j=1..n, k=j+1..n} \left(\frac{\hat{p}_j + \hat{p}_k}{2} \right) \right|.$$

4. Если $k < m$, то $k = k + 1$ и перейти к пункту 3, иначе перейти к пункту 5.
5. Вычислить оптимальный параметр размытости:

$$b_n^{opt} = \arg \min_{b_n \in \{b_n^1, b_n^2, \dots, b_n^m\}} \sum_{i=1}^n \delta_i \cdot |\hat{g}(\hat{p}_i | x_i) - Y_i|.$$

2.2 Исследование свойств оценки Берана с применением предложенного алгоритма

Необходимо проверить, удовлетворяет ли предложенный алгоритм требованиям, описанным в главе 1. Исследование оценки Берана с использованием предложенного алгоритма проводилось в сравнении с референтным методом [99], так как, несмотря на свои недостатки, данный подход всё же может применяться в некоторых случаях. Значение константы C примем равным единице. Таким образом, в следующих исследованиях референтное правило выглядит следующим образом:

$$b_n = n^{-1/5}. \quad (2.9)$$

Оценка оптимального значения параметра размытости b_{pp}^j определялась в соответствии с формулой (1.16). В качестве оценки точности подобранного параметра рассматривалась величина:

$$Db_{pp}^j(b) = \frac{|b_{pp}^j - b|}{b_{pp}^j}. \quad (2.10)$$

В качестве оценки точности получаемых оценок Берана рассматривалось максимальное отклонение при заданном значении ковариаты вида (2.6).

В первую очередь было рассмотрено, насколько близко к оптимальному предложенный метод подбирает оценку параметра размытости, и какая точность оценки Берана при этом достигается в зависимости от объёма выборки. В качестве истинной модели была выбрана модель пропорциональных интенсивностей Кокса с экспоненциальной базовой функцией и линейной функцией от ковариат. Исследования проводились как для случайного плана, когда ковариата x принимает случайное значение на интервале $[0,1]$, так и для неслучайного плана. В последнем случае ковариата x принимает значения спектра плана $\text{supp}(\xi_n^m) = \{0, 0.11, 0.22, 0.33, 0.44, 0.56, 0.67, 0.78, 0.89, 1\}$, $m=10$. Количество наблюдений, соответствующее различным точкам спектра плана, во всех экспериментах считалось одинаковым. Поскольку оптимальное значение параметра размытости зависит от того, для какой группы строится оценка, то рассматривались оценки точности при значениях ковариаты $x=0.11$ и $x=0.56$. В таблицах 2.3 – 2.9 во втором столбце приведены отклонения $Db_{pp}^j(b_n^{opt})$ параметра размытости b_n^{opt} , вычисленного с помощью предложенного алгоритма, от оптимального b_{pp}^j , в третьем столбце приведены значения статистики $D_{avg}(b_n^{opt})$, вычисленные по формуле (2.6) при параметре b_n^{opt} , в четвёртом столбце приведены отклонения $Db_{pp}^j(b_n)$ параметра размытости b_n , вычисленного с помощью референтного алгоритма, от оптимального b_{pp}^j , в пятом столбце приведены значения статистики $D_{avg}(b_n)$ при параметре b_n , в шестом столбце точность оценки

Берана при оптимальном параметре размытости b_{pp}^j . Полученные значения усреднялись по $N = 2000$ выборкам.

В таблицах 2.3 – 2.4 приведены результаты исследований для неслучайного плана эксперимента, при различных объемах выборок для значений ковариаты $x = 0.11$ и $x = 0.56$, соответственно.

Таблица 2.3 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от объема выборки при неслучайном плане эксперимента и значении ковариаты $x = 0.11$

Объём выборки	$Db_{pp}^{x=0.11}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.11})$
20	0.33	0.24	0.41	0.24	0.19
40	0.33	0.20	0.33	0.20	0.17
100	0.33	0.13	0.33	0.13	0.10
200	0.33	0.10	0.29	0.10	0.08
300	0.33	0.08	0.29	0.08	0.06

Из таблицы 2.3 видно, что при значении ковариаты $x = 0.11$ при обоих способах выбора параметра размытости отклонение от оптимального значения параметра незначительны при малых объемах выборки (20...40 наблюдений). Однако при больших объемах выборки предложенный метод позволяет получить более точную оценку Берана, близкую к наилучшей.

Из таблицы 2.4 видно, что для значения ковариаты $x = 0.56$, предложенный метод позволяет построить оценку, достаточно близкую к оптимальной и во всех рассмотренных случаях превосходит по точности референтный метод выбора параметра. При больших объемах выборки (от 100 наблюдений) точность оценки Берана такая же, как для значения $x = 0.11$, однако для меньших объемов точность выше. Это связано с тем, что для

центральных групп больше «соседних» наблюдений, поэтому информации для построения оценки больше. При большем же объёме данная разница в количестве соседних наблюдений перестаёт быть значимой.

Таблица 2.4 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от объёма выборки при неслучайном плане эксперимента и значении ковариаты $x = 0.56$

Объём выборки	$Db_{pp}^j(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{true})$
20	0.17	0.19	0.45	0.22	0.17
40	0.09	0.13	0.52	0.16	0.12
100	0.30	0.09	0.45	0.11	0.08
200	0.45	0.07	0.45	0.09	0.06
300	0.57	0.07	0.47	0.08	0.06

В таблицах 2.5 – 2.6 приведены результаты исследований для случайного плана эксперимента, при различных объёмах выборок для значений ковариаты $x = 0.11$ и $x = 0.56$, соответственно.

Для случайного плана оба метода подбирают параметр размытости с одинаковой точностью. Одинаковую точность методов, а также достаточно точный подбор параметра размытости референтным методом для значения ковариаты 0.11 для группированного плана в данном случае объясняется следующим. Во-первых, случайный план для оценки Берана близок к постановке задачи для ядерного сглаживания, а именно для этих целей построен референтный метод выбора параметра. Во-вторых, ковариата x принимала значения в соответствии с равномерным распределением. Так при малых объёмах выборок при 10 группах план близок к случайному, то точность референтного метода высока.

Таблица 2.5 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от объёма выборки для случайного плана при значении ковариаты $x = 0.11$.

Объём выборки	$Db_{pp}^{x=0.11}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.11})$
20	0.86	0.26	0.48	0.26	0.18
40	0.97	0.21	0.51	0.21	0.15
100	0.32	0.14	0.32	0.14	0.11
200	0.27	0.11	0.27	0.11	0.08
300	0.27	0.09	0.27	0.09	0.08

Таблица 2.6 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от объёма выборки для случайного плана при значении ковариаты $x = 0.56$.

Объём выборки	$Db_{pp}^{x=0.56}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.56})$
20	0.68	0.27	0.44	0.26	0.15
40	0.80	0.23	0.51	0.15	0.12
100	0.71	0.13	0.60	0.12	0.08
200	0.51	0.09	0.47	0.09	0.06
300	0.46	0.07	0.46	0.07	0.05

В таблице 2.7 приведены исследования для плана эксперимента, в котором опорные точки выбирались неравномерно: $\text{supp}(\xi_n^q) = \{0.11, 0.44, 0.5, 0.6, 0.7, 0.8, 0.90, 0.95, 1\}$, число групп $m=10$ осталось прежним.

Таблица 2.7 – Точность подбора параметра размытости и оценки Берана при предложенном методе при неравномерном плане при значении ковариаты $x = 0.11$

Объём выборки	$Db_{pp}^{x=0.11}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.11})$
20	0.20	0.32	0.35	0.38	0.26
40	0.30	0.28	0.36	0.33	0.22
100	0.22	0.20	0.40	0.24	0.16
200	0.45	0.17	0.45	0.17	0.12

Как видно из таблицы 2.7, точность подбора референтным методом упала и сравнивается с точностью подбора предложенного алгоритма только при 200 наблюдений, когда оба метода дают такие значения параметра размытости, при которых учитывается только одна группа. Однако, как видно из точности оценки при оптимальном параметре, это не всегда справедливо. В целом заметно, что оценка Берана стало менее точной, и это связано со сложностью плана.

Ниже, в таблицах 2.8-2.9 приводятся результаты исследований при подборе параметра размытости и анализе достигаемой точности оценки Берана в зависимости от количества групп. В качестве заданной модели бралась та же модель, что и выше. Объём выборки $n = 100$. Количество наблюдений, соответствующее различным значениям ковариаты, во всех экспериментах считалось одинаковым. Значение ковариаты распределено равномерно на отрезке $[0,1]$.

Таблица 2.8 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от числа групп при значении ковариаты $x = 0.11$

Количество групп	$Db_{pp}^{x=0.11}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.11})$
4	0.20	0.10	0.45	0.14	0.09
10	0.23	0.11	0.27	0.12	0.09
20	0.26	0.12	0.27	0.12	0.09
50	0.35	0.14	0.35	0.14	0.12
100	0.32	0.14	0.32	0.14	0.11

Таблица 2.9 – Точность подбора параметра размытости и оценки Берана при предложенном методе в зависимости от числа групп при значении ковариаты $x = 0.56$

Количество групп	$Db_{pp}^{x=0.56}(b_n^{opt})$	$D_{avg}(b_n^{opt})$	$Db_{pp}^j(b_n)$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.56})$
4	0.14	0.10	0.45	0.14	0.09
10	0.20	0.09	0.45	0.12	0.08
20	0.30	0.09	0.50	0.11	0.08
50	0.46	0.10	0.53	0.11	0.08
100	0.45	0.17	0.45	0.17	0.12

Во всех рассмотренных случаях предложенный метод выбора параметра размытости показал точность выше, чем референтный метод, а также точность оценки Берана, близкую к оптимальной. Из таблиц 2.7 и 2.8 также видно, что для значения ковариаты $x = 0.11$, с увеличением числа групп точность оценки падает, так как становится меньше наблюдений, соответствующих одному значению ковариаты.

В таблице 2.10. демонстрируется использование предложенного метода выбора параметра размытости при различных типах моделей. Рассматривались модели ускоренных испытаний и пропорциональных интенсивностей при различных значениях регрессионного параметра β .

Таблица 2.10 – Точность оценки Берана при предложенном методе в зависимости от типа модели

Модель	Значение регрессионного параметра β	$D_{avg}(b_n^{opt})$	$D_{avg}(b_n)$	$D_{avg}(b_{pp}^{x=0.56})$
AFT	$\beta = 2$	0.11	0.12	0.09
	$\beta = 5$	0.15	0.16	0.13
Cox	$\beta = 2$	0.11	0.12	0.09
	$\beta = 5$	0.15	0.16	0.13

Точность подбора параметра размытости для обоих методов не зависит от типа модели. С увеличением регрессионного параметра точность оценки Берана падает, так как, чем больше регрессионный параметр, тем сильнее различаются условные функции надёжности при различных значениях ковариаты.

На рисунке 2.6 показано, насколько точно алгоритм подбирает оптимальное значение параметра размытости для каждой группы. Полученные результаты сравнивались с подобранными параметрами размытости (1.16). В данных сравнениях не приводятся значения для референтного способа выбора параметра, так как его значение не меняется в зависимости от номера группы.

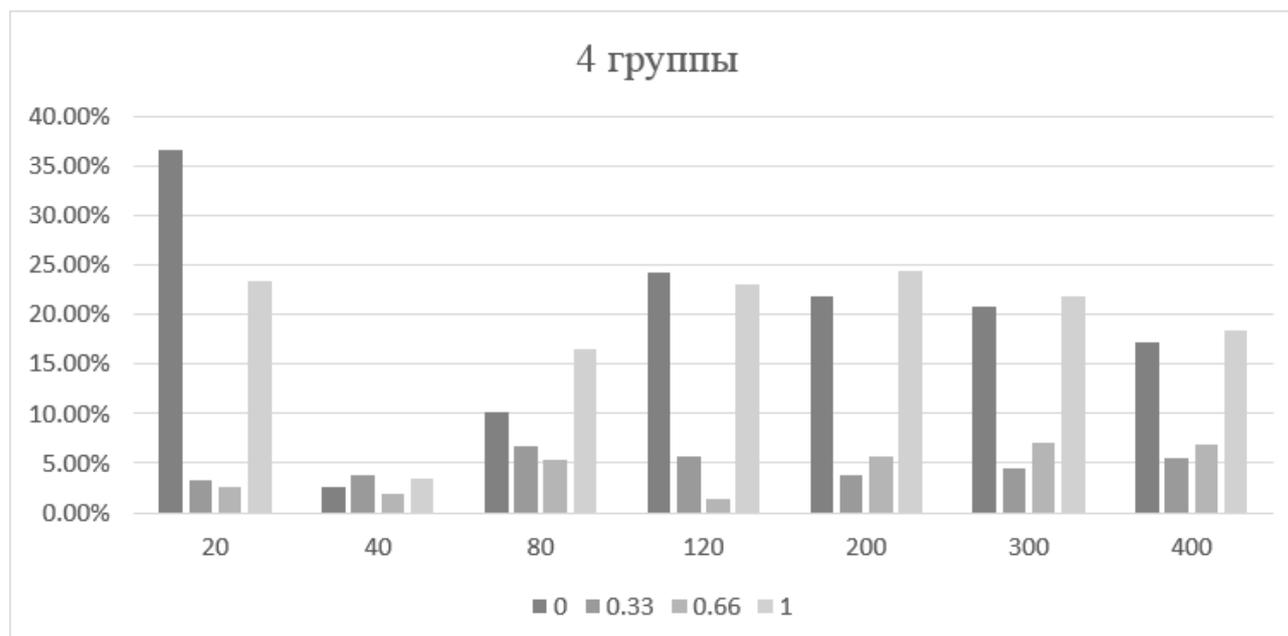


Рисунок 2.6 – Относительное отклонение параметра b_n^{opt} от оптимального параметра размытости в зависимости от номера группы и объёма выборки

Как видно из графика, предложенный подход позволяет достаточно точно подобрать оптимальный параметр размытости. Так, максимальное отклонение составляет порядка 37 %, при достаточно малом объёме выборок, когда на каждую группу приходится только пять наблюдений. При большем количестве групп максимальное отклонение составляет не более 24% для крайних групп и не более 7% – для центральных. Большее отклонение для крайних групп при объёмах 120-400 наблюдений объясняется так же тем, что оптимальное значение параметра размытости при таких объёмах почти всегда меньше, либо равно 0.33. А это означает, что оценка Берана вырождается в оценку Каплана–Мейера. В этом случае подобранный параметр размытости может принимать любое значение в интервале $(0, 0.33]$. И при любом значении из этого интервала значение параметра будет оптимальным. Для всех случаев, начиная с объёма $n=120$, для крайних наблюдений значение параметра размытости выбирается меньше 0.33.

Ниже на рисунке 2.7 представлен более сложный пример с десятью группами.

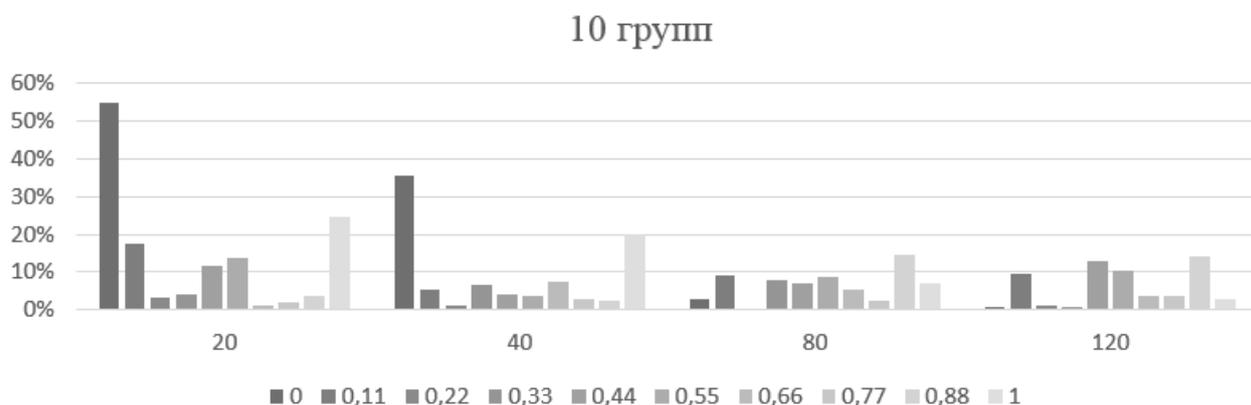


Рисунок 2.7 – Относительное отклонение параметра b_n^{opt} от оптимального параметра размытости в зависимости от номера группы и объёма выборки

При десяти группах наблюдается примерно схожая картина. При малых объёмах выборок, когда оценивание параметра происходит по двум наблюдениям, наихудшие оценки получаются для крайних групп. С увеличением объёма выборки точность оценок растёт и уже для объёма в 80 наблюдений отклонение не превосходит 15%.

Следует добавить, что увеличение объёма выборки может и не приводить к улучшению точности подбора оптимального значения параметра размытости. С увеличением объёма выборки увеличивается не только число элементов в группе, соответствующей значению ковариаты, для которой строится оценка Берана, но и число элементов в других группах. Этим можно объяснить то, что для крайних значений ковариаты точность оценки Берана увеличивается с увеличением объёма выборки. При этом для центральных точек спектра плана такой тенденции не наблюдается. Для того, чтобы увеличить точность подбора параметра размытости необходимо увеличивать объём только исследуемой группы, не увеличивая другие группы.

На рисунках 2.8 и 2.9 представлены теоретические функции надежности и оценки Берана, построенные при использовании оптимального значения параметра размытости, выбранного с помощью предложенного метода, для выборок Y_1 и Y_2 , по которым строились оценки на рисунках 1.3 и 1.4.

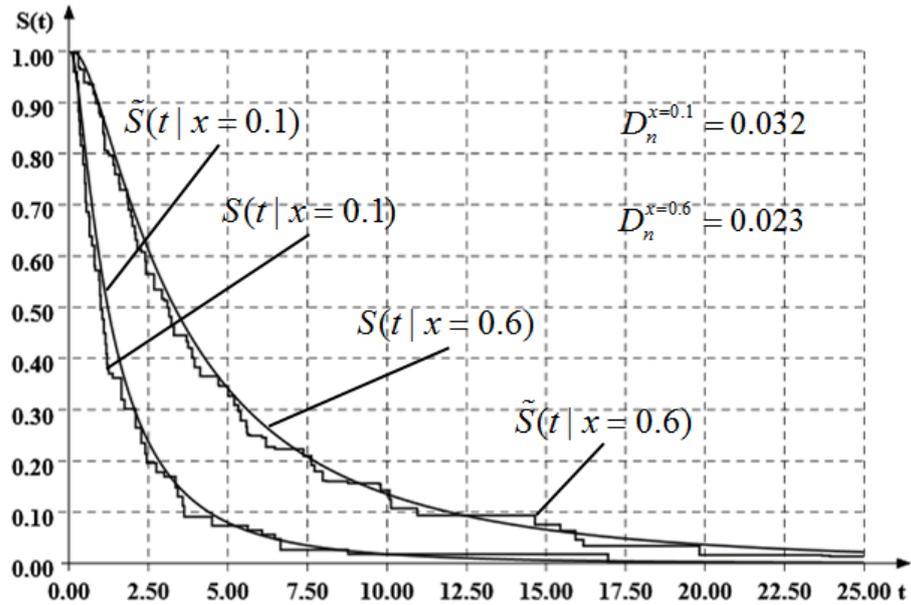


Рисунок 2.8 – Функция надёжности и оценки Берана с параметром b_n^{opt} , $\beta=2$

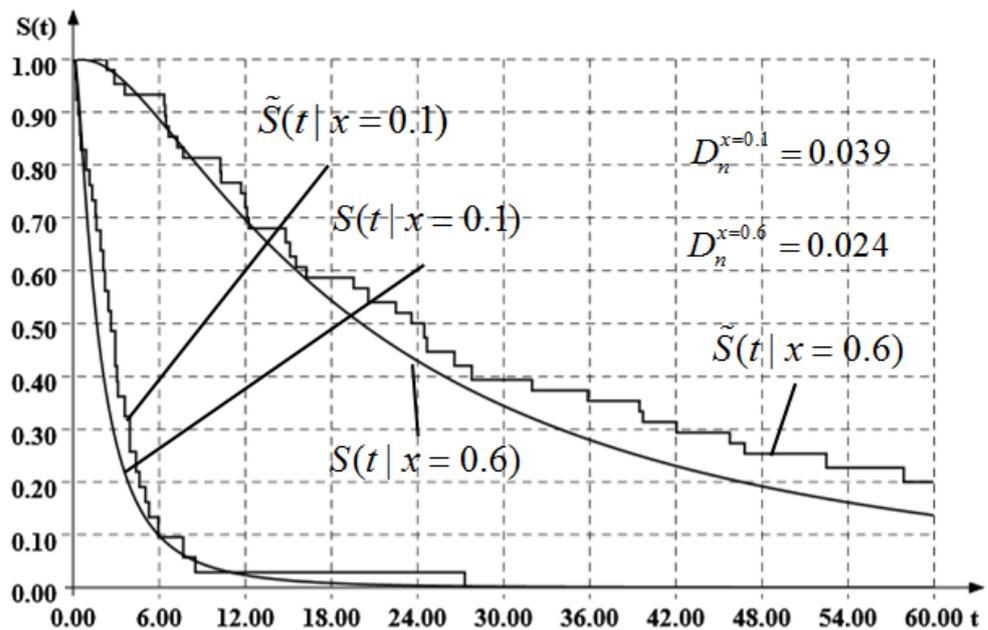


Рисунок 2.9 – Функция надёжности и оценки Берана с параметром b_n^{opt} , $\beta=5$

Как видно на рисунке 2.8, оценки Берана достаточно близки к соответствующим функциям надёжности: отклонения вида (1.18) при $x=0.1$ и $x=0.6$ равны 0.032 и 0.023, соответственно. Впрочем, на рисунке 1.3 наблюдалась аналогичная картина.

Напротив, на рисунке 2.9 наблюдается другая картина. Здесь оценки Берана значительно ближе к соответствующим теоретическим функциям надёжности ($D_n^{x=0.1} = 0.039$ и $D_n^{x=0.6} = 0.024$), чем на рисунке 1.4, что свидетельствует о том, что применение алгоритма выбора оптимального значения параметра размытости позволяет существенно повысить точность оценок Берана по сравнению с подходами к выбору параметра размытости, основанными на объеме выборки и особенностях плана эксперимента.

2.3 Выводы

В данной главе предложен адаптивный алгоритм выбора оптимального значения параметра размытости, используемого при построении непараметрической оценки Берана регрессионной модели надёжности. Алгоритм основан на минимизации среднеквадратического отклонения времен отказов от непараметрической оценки обратной функции надёжности.

Методами статистического моделирования проведено исследование свойств оценки Берана, получаемой с использованием предложенного алгоритма, в результате которого показано следующее.

1. На точность оценки Берана оказывают влияние объём выборки, количество опорных точек плана эксперимента, а также вид ядерных функций и способ вычисления параметра сглаживания, используемые при построении непараметрической оценки обратной функции надёжности.

2. Наилучшие результаты достигаются при использовании весовой функции Пристли-Чао с параметром сглаживания h_{NS} , вычисляемого на основе медианы абсолютных отклонений $\hat{\sigma}_{robust}$.
3. Среди рассмотренных вариантов ядерных функций предпочтительней использовать ядро Епанечникова и квартическую функцию, поскольку они позволяют получать более точные оценки Берана.
4. Оценки Берана, получаемые на основе предложенного алгоритма выбора оптимального значения параметра размытости, оказываются точнее, чем при использовании референтного правила выбора параметра размытости, который применим только при случайном плане эксперимента.

ГЛАВА 3 Критерии согласия на основе оценки Берана

В данной главе для проверки гипотез о виде параметрических регрессионных моделей надёжности предлагаются непараметрические критерии согласия, в статистиках которых используются оценки Берана. Методами статистического моделирования исследуются распределения статистик и мощность предложенных критериев согласия, проводится сравнение с классическими критериями согласия, применяемыми к выборкам остатков. Результаты, представленные в данной главе, опубликованы в [16, 73, 78, 79].

3.1 Проверка гипотез о виде параметрических моделей с помощью критериев согласия на основе выборок остатков

Существующие критерии проверки гипотезы о виде параметрической вероятностной модели пропорциональных интенсивностей основаны на построении остатков Кокса-Снелла, которые можно вычислить следующим образом [28]:

$$R_i = \Lambda_{x_i}(Y_i; \hat{\beta}), \quad i = 1, 2, \dots, n, \quad (3.1)$$

где $\Lambda_{x_i}(\cdot)$ – кумулятивная функция риска, $\hat{\beta}$ – ОМП параметров предполагаемой модели.

Если гипотеза о виде модели верна, то полученная выборка остатков $\mathbf{R}_n = \{(R_1, \delta_1), \dots, (R_n, \delta_n)\}$ принадлежит стандартному экспоненциальному распределению [28].

Подобный подход применяется и для моделей ускоренных испытаний. Для проверки гипотезы о виде построенной параметрической AFT-модели также анализируют выборку остатков. Остатки для параметрической AFT-модели вычисляются иначе, чем остатки для моделей пропорциональных

интенсивностей. Кроме того, для AFT моделей ковариата может быть – как постоянной, так и ступенчатой. Однако в данной работе исследуется только случай с постоянной по времени ковариатой, поэтому рассмотрим остатки только для этого случая:

$$R_i = \frac{Y_i}{r(x_i; \hat{\beta})}, \quad i = \overline{1, n}. \quad (3.2)$$

Если данные хорошо описываются построенной AFT-моделью, остатки распределены в соответствии с базовым законом распределения отказов $F_0(t; \hat{\theta})$, стандартизованным по параметру масштаба (при параметре масштаба $\theta_1 = 1$).

Для проверки гипотезы о принадлежности выборки остатков экспоненциальному распределению в случае моделей пропорциональных интенсивностей с базовым распределением $F_0(t; \theta)$ в случае AFT-модели можно использовать критерии согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлингга [3].

Если гипотеза проверяется по выборке без цензурированных наблюдений, статистика критерия Колмогорова с поправкой Большева имеет вид [65]

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \quad (3.3)$$

где $D_n = \sup_{0 \leq t \leq \infty} |F_n(t) - F(t)|$.

На практике значение статистики вычисляют, используя величины $D_n = \max\{D_n^+, D_n^-\}$, $D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(R_{(i)}) \right\}$, и $D_n^- = \max_{1 \leq i \leq n} \left\{ F(R_{(i)}) - \frac{i-1}{n} \right\}$, где $R_{(1)}, R_{(2)}, \dots, R_{(n)}$ – упорядоченные по возрастанию выборочные значения.

Статистика Крамера–Мизеса–Смирнова определяется следующим образом [9]:

$$S_{\omega} = n \int_0^{\infty} (F(t) - F_n(t))^2 dF(t) \quad (3.4)$$

и вычисляется как $S_{\omega} = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(R_{(i)}) - \frac{(2i-1)}{2n} \right\}^2$.

Статистика критерия Андерсона–Дарлинга задается выражением [3]

$$S_{\Omega} = n \int_0^{\infty} \frac{(F(t) - F_n(t))^2}{F(t) \cdot (1 - F(t))} dF(t). \quad (3.5)$$

В выражениях (3.3), (3.4) и (3.5) $F(t)$ представляет собой функцию распределения стандартного экспоненциального закона в случае модели пропорциональных интенсивностей и функцию распределения базового закона распределения отказов в случае АFT-модели, $F_n(t)$ – эмпирическая функция распределения остатков.

Следует заметить, что при построении моделей с учетом ковариат, проверяемая гипотеза о согласии выборки остатков со стандартным экспоненциальным законом или с базовым распределением отказов, является сложной, так как проверка осуществляется по той же выборке, по которой были оценены параметры модели.

В случае проверки сложной гипотезы о согласии, непараметрические критерии согласия теряют свойство свободы от распределения [84], и распределения статистик $G(y|H_0)$ непараметрических критериев согласия зависят от вида проверяемой вероятностной модели, от количества параметров, оцениваемых по выборке, от метода оценивания параметров, в некоторых случаях от значений параметров и других факторов.

Модели распределений статистик непараметрических критериев согласия Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга при проверке сложных гипотез с использованием оценок максимального правдоподобия для различных комбинаций параметров в случае полных выборок (без цензурированных наблюдений) представлены в работах [12, 22].

Для проверки гипотезы о виде распределения выборки остатков при наличии цензурированных измерений можно воспользоваться модифицированными критериями согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга, в которых в выражениях для статистик вместо эмпирической функции распределения используется непараметрическая оценка Каплана–Мейера [31].

В случае проверки гипотез о виде распределения выборки остатков при наличии цензурированных измерений распределения статистик рассматриваемых критериев зависят от степени цензурирования и распределения моментов цензурирования в вариационном ряду выборки остатков. Необходимо четко понимать, что наличие множества указанных факторов, влияющих на распределения статистик вышеперечисленных критериев, не позволяет заранее построить распределения статистик, соответствующие конкретным условиям испытаний.

Следовательно, распределения статистик модифицированных критериев согласия на основе выборок остатков могут быть построены с применением подхода, использованного в [12, 22], опирающегося на компьютерные технологии и статистическое моделирование.

Интерактивный алгоритм моделирования распределений статистик критериев согласия на основе выборок остатков определяется следующим образом.

1. Смоделировать полную выборку отказов $(Y_1, \delta_1 = 1, x_1), \dots, (Y_n, \delta_n = 1, x_n)$ в соответствии с проверяемой моделью $F_x(t; \hat{\beta}, \hat{\theta})$, где $\hat{\beta}$ и $\hat{\theta}$ – оценки максимального правдоподобия параметров модели по исходной выборке.

При этом момент отказа объекта, находящегося под постоянным воздействием x_i , моделируется в соответствии с выражением

$$Y = S^{-1}\left(\zeta; \hat{\theta} | r(x_i; \beta)\right), \quad (3.6)$$

где ζ – псевдослучайная величина, равномерно распределенная на интервале $[0,1]$.

2. При необходимости, в соответствии с заданной схемой цензурирования, преобразовать полную выборку в цензурированную $(Y_1, \delta_1, x_1), \dots, (Y_n, \delta_n, x_n)$.

3. Вычислить оценки максимального правдоподобия параметров β и θ по выборке $(Y_1, \delta_1, x_1), \dots, (Y_n, \delta_n, x_n)$.

4. Вычислить остатки для проверяемой модели (по (3.1) или (3.2)).

5. По выборке остатков вычислить значение статистики критерия (по (3.3), (3.4) или (3.5)).

6. Повторив пункты 1-5 N раз, получим эмпирическое распределение статистики $G_N(y | H_0)$.

Выбор объема N обусловлен желаемой точностью моделирования (подробнее о выборе N см. в [23]).

Описанный алгоритм позволяет корректно оценить достигнутый уровень значимости при проверке гипотезы о виде параметрических моделей пропорциональных интенсивностей или AFT-модели.

Исследования показали, что рассматриваемые критерии на основе выборок остатков обладают высокой мощностью при проверке гипотез о виде базового распределения отказов. Однако при проверке гипотезы о виде регрессионной зависимости высокая мощность данных критериев достигается лишь при очень больших объемах выборок [102].

3.2 Критерии согласия на основе оценки Берана

Основная идея предлагаемого критерия проверки гипотезы о виде параметрической регрессионной модели надежности на основе оценки Берана заключается в том, чтобы оценить расстояние между теоретической условной функцией надёжности, соответствующей проверяемой гипотезе, и непараметрической оценкой Берана при различных значениях ковариаты. Идея схожа с классическим подходом к проверке гипотез о виде моделей с помощью непараметрических критериев согласия. Главное отличие разработанного подхода заключается в том, что не возникает необходимости перехода к выборке остатков. Отсутствие такого перехода делает предлагаемый метод более гибким и универсальным.

В [59] для случайного плана эксперимента и фиксированного значения x показано, что при $n \rightarrow \infty$:

$$\sup_{t \geq 0} |\tilde{S}_b(t | x) - S_x(t)| = O\left(\sqrt{\frac{\ln n}{nb_n}}\right), \quad (3.7)$$

если $\frac{nb_n^5}{\ln n} = O(1)$.

Проверим справедливость выражения (3.7) с помощью методов компьютерного моделирования. Для этого можно аппроксимировать расстояние между оценкой Берана и истинной функцией надежности следующей зависимостью от объема выборки:

$$D_n = \sup_{t \geq 0} |\tilde{S}_{b_n}(t | x) - S_x(t)| = \alpha_0 + \alpha_1 \left(\frac{\ln n}{nb_n}\right)^{0.5} + \varepsilon, \quad (3.8)$$

где ε – случайная ошибка модели, которая удовлетворяет условиям: $M(\varepsilon) = 0$, $M(\varepsilon\varepsilon^T) = \sigma^2 < \infty$, α_0 , α_1 – параметры, которые следует оценить.

Чтобы оценить параметры α_0 и α_1 , смоделируем $N = 2000$ выборок размером от 20 до 500 наблюдений при случайном плане, в котором опорные

точки плана выбирались из равномерного на $[0,1]$ распределения. В качестве истинной модели рассмотрим модель Кокса:

$$S_x(t) = (S_0(t))^{r(x;\beta)}$$

с функцией от ковариат вида $r(x;\beta) = \ln(1 + e^{\beta x})$, $\beta = 2$ и логнормальным базовым распределением с функцией плотности:

$$f_0(t) = \frac{1}{\sqrt{2\pi\theta_1 t}} \exp\left(-\frac{1}{2\theta_1^2} \ln^2\left(\frac{t}{\theta_2}\right)\right)$$

с параметрами $\theta_1 = 21.5$, $\theta_2 = 1.6$. Значения ковариаты моделировались в соответствии с равномерным распределением на интервале $[0, 1]$.

По сгенерированным выборкам считалось расстояние D_n между истинной условной функцией надёжности и оценкой Берана при $x = 0.5$. В результате получена выборка вида $\{(D_{n_1}, n_1, b_{n_1}), \dots, (D_{n_l}, n_l, b_{n_l})\}$, по которой методом наименьших квадратов оценивались параметры α_0 и α_1 , входящие в (3.8). На рисунке 3.1 в зависимости от объёма выборки показаны средние значения статистики D_n и полученная аппроксимация зависимости (3.8) с оценками параметров.

Как видно из рисунка 3.1, функция (3.8) достаточно точно описывает скорость сходимости оценки, при этом коэффициент детерминации $R^2 = 0.988$. Таким образом, результат, полученный в [59], о том, что оценка Берана сходится к истинной условной функции надёжности со скоростью $O\left(\sqrt{\frac{\ln n}{nb_n}}\right)$ при случайном плане, подтверждается результатами компьютерного моделирования.

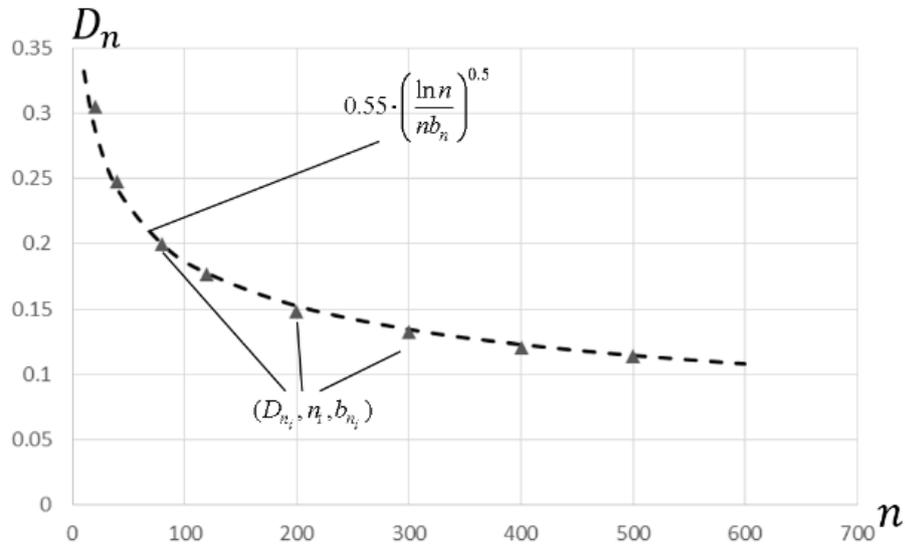


Рисунок 3.1 – Аппроксимация статистики (3.8) при случайном плане эксперимента для различных объёмов выборки

Одной из наиболее популярных статистик, применяемых для оценки расстояния между функциями надёжности, является статистика типа Колмогорова. В случае проверки гипотезы (1.11) статистика принимает вид:

$$\sup_{i=1..n, t>0} \left| S_{x_i}(t; \theta, \beta) - \tilde{S}_{b_n}(t | x_i) \right|, \quad (3.9)$$

где $S_{x_i}(t; \theta, \beta)$ – условная функция надёжности, соответствующая проверяемой гипотезе H_0 .

Используя результаты исследования сходимости оценки Берана к истинной функции надёжности, можно построить критерий согласия на основе следующей статистики:

$$S_B = \left(\frac{nb_n}{\ln n} \right)^{0.5} \sup_{i=1..n, t>0} \left| \tilde{S}_{b_n}(t | x) - S_x(t) \right|, \quad (3.10)$$

распределение которой быстро сходится к некоторому предельному закону.

Вместе с тем, статистика (3.10) имеет два существенных недостатка. Во-первых, она применима только для случайного плана. В противном случае

необходимо учитывать количество групп, соответствующих опорным точкам плана, и количество элементов в группе. В качестве демонстрации зависимости распределения статистики (3.10) от количества групп на рисунке 3.2 показаны распределения статистики S_B при проверке гипотезы относительно модели Кокса (объем выборки $n = 120$, количество групп $m = 4, 10, 20, 40, 60$).

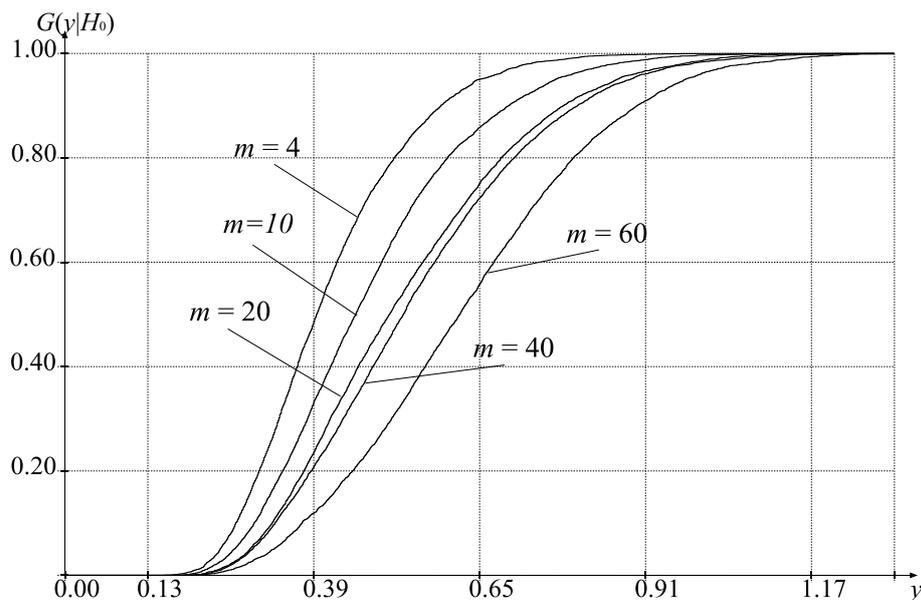


Рисунок 3.2 – Распределения статистики (3.10) при различном числе групп плана эксперимента

Как видно из рисунка 3.2, распределение статистики (3.10) существенно зависит от числа групп, тестируемых под разными воздействиями. Другими словами, в случае дискретного плана эксперимента, распределения статистики (3.10) при различных значениях ковариаты с ростом объемов выборок стремятся к различным предельным распределениям.

Вторым недостатком статистики (3.10) в контексте построения критерия согласия является то, что она учитывает только одно значение ковариаты. Другими словами, при построении статистики не используются все данные, представленные в выборке, и, следовательно, теряется часть информации об исследуемом явлении. Кроме того, остается открытым вопрос о том, как

выбрать значение ковариаты, для которого необходимо рассчитать статистику. Мощность критерия, построенного по одному значению ковариаты, будет заведомо ниже, чем мощность критерия, построенного с учётом всех значений ковариат.

Для того чтобы учесть все значения ковариаты в выборке, в данном диссертационном исследовании предлагается использовать статистику вида:

$$S_B^1 = \sup_{i=1..m} \left[\left(\frac{nb_n}{\ln n} \right)^{0.5} \sup_{t>0} |S_{x_i}(t; \theta, \beta) - \tilde{S}_{b_n}(t | x_i)| \right], \quad (3.11)$$

в которой берется наибольшее отклонение оценки Берана от предполагаемой функции надежности, либо среднее отклонение вида:

$$S_B^2 = \left(\frac{1}{m} \right) \sum_{i=1}^m \left(\frac{nb_n}{\ln n} \right)^{0.5} \sup_{t>0} |S_{x_i}(t; \theta, \beta) - \tilde{S}_{b_n}(t | x_i)|, \quad (3.12)$$

где $\tilde{S}_{b_n}(t | x_i)$ – оценка Берана при значении ковариаты x_i .

Предлагаемые статистики позволяют использовать особенности каждого из видов плана эксперимента. Статистика (3.11) предназначена для плана эксперимента с большим количеством элементов в группе, так как в этом случае информации о каждом значении ковариаты в плане эксперимента достаточно для того, чтобы использовать наибольшее отклонение. При случайном же плане, либо при планах, в которых количество наблюдений для каждого значения ковариаты невелико, следует использовать статистику (3.12).

Гипотеза о виде параметрической регрессионной модели надёжности (1.11) отвергается при больших значениях статистик S_B^1 и S_B^2 . Применение предложенных критериев предполагает реализацию интерактивного режима для моделирования распределений статистик $G_N(y | H_0)$ в ходе проводимого

анализа, чтобы использовать $G_N(y|H_0)$ при формировании статистического вывода о результатах проверки гипотезы.

Таким образом, для проверки гипотезы о виде параметрической регрессионной модели надёжности (1.11) с помощью предложенных критериев необходимо:

1. В соответствии с построенной моделью $H(S_0(t; \hat{\theta}), x, \hat{\beta})$, где $\hat{\theta}, \hat{\beta}$ – ОМП параметров модели по исходной выборке, смоделировать выборку отказов $Y = \{(Y_1, x_1), (Y_2, x_2), \dots, (Y_n, x_n)\}$.
2. По полученной выборке оценить параметры модели методом максимального правдоподобия.
3. Вычислить значение статистики (3.11) или (3.12).
4. Повторить пункты 1-3 N раз, получив в результате эмпирическое распределение статистики критерия $G_N(y|H_0)$.

Проверяемая гипотеза H_0 отклоняется, если достигнутый уровень значимости $\alpha_n = 1 - G_N(S_B | H_0)$ меньше заданного α , где S_B – значение статистики соответствующего критерия согласия, полученное по исходной выборке.

Использование описанного алгоритма проверки гипотез о виде параметрических регрессионных моделей надёжности позволяет проверить предположение как о виде параметризации базового распределения отказов, так и о виде регрессионной зависимости функции надёжности от ковариат.

3.3 Исследование распределений статистик и мощности критериев на основе оценки Берана

Здесь приводятся исследование поведения распределений статистик предложенных критериев в зависимости от объёма выборки и количества

групп. В качестве истинной модели, соответствующей проверяемой гипотезе H_0 , рассматривалась модель ускоренных испытаний:

$$S_x(t) = S_0\left(\frac{t}{r(x;\beta)}\right)$$

с функцией от ковариат вида $r(x;\beta) = \ln(1 + e^{\beta x})$ и базовой функцией надёжности, соответствующей экспоненциальному распределению. На рисунках 3.3 и 3.4 приведены распределения статистик (3.11) и (3.12), при объёмах выборок $n = 20, 40, 80, 120, 200, 300$ для количества групп $m=10$.

Как видно из рисунков 3.3 и 3.4, распределения обеих статистик зависят от объёма выборки и с увеличением n сдвигаются в область меньших значений статистики. Что вполне логично, так как с увеличением объёма выборки и при фиксированном количестве групп, информации о каждой группе наблюдений с одинаковым значением ковариаты становится больше, а, следовательно, и точность оценки как для одной группы, так и в целом для всей выборки увеличивается.

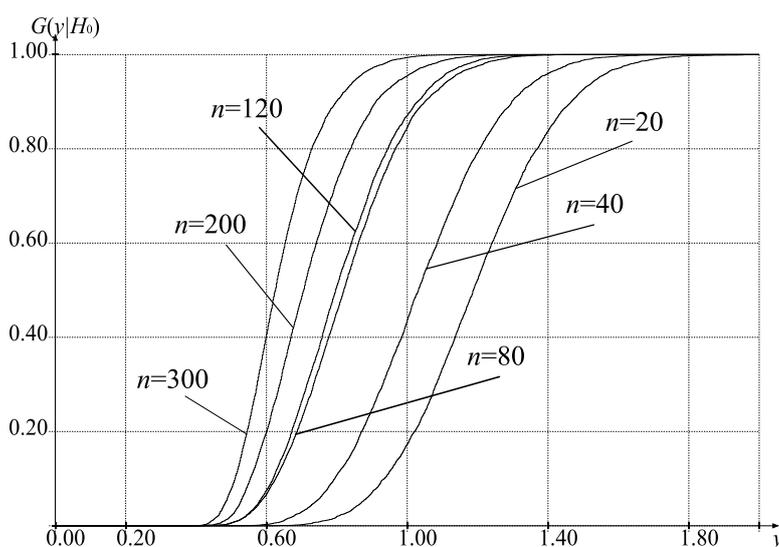


Рисунок 3.3 – Распределения статистики (3.11) при различных объёмах выборки для количества групп $m=10$

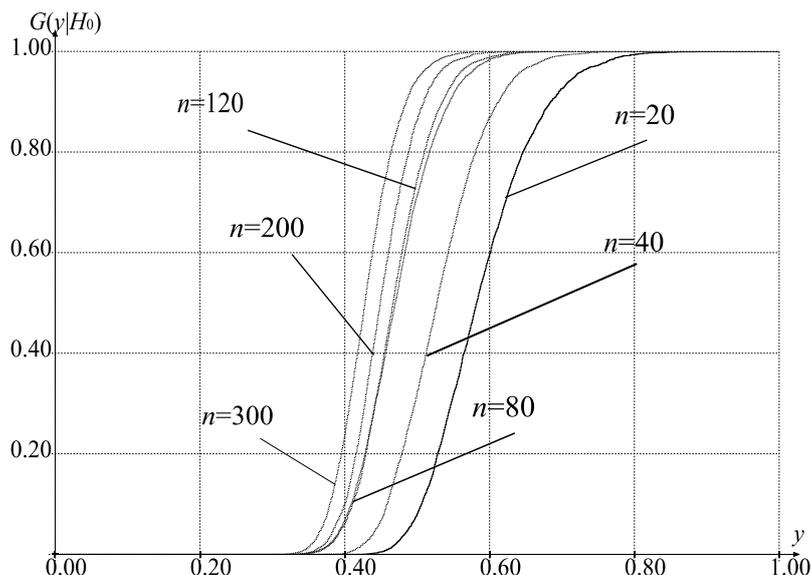


Рисунок 3.4 – Распределения статистики (3.12) при различных объёмах выборки для количества групп $m=10$

Ниже приведены результаты исследования поведения распределений статистик (3.11) и (3.12) в зависимости от числа групп.

На рисунках 3.5 и 3.6 представлены распределения исследуемых статистик для фиксированного объёма выборок $n=120$ с различным числом групп: $m=4,10,20,30,40,60$. Количество наблюдений в каждой группе одинаково.

Как показано на рисунке 3.5, с увеличением количества групп распределения статистики (3.11) оказываются правее. Однако распределения статистики (3.12) с ростом количества групп смещаются влево, как видно на рисунке 3.6. Это связано с тем, что статистика (3.11) соответствует максимальному расстоянию между теоретической функцией надёжности и наихудшей оценкой Берана. Таким образом, с увеличением количества различных значений ковариаты информация, связанная с одним значением уменьшается. В то же время, статистика (3.12) соответствует усреднённому значению максимального расстояния по всем значениям ковариаты. Поэтому с

увеличением количества различных значений ковариаты вклад наихудшей оценки уменьшается, и усреднение происходит в большем количестве групп.

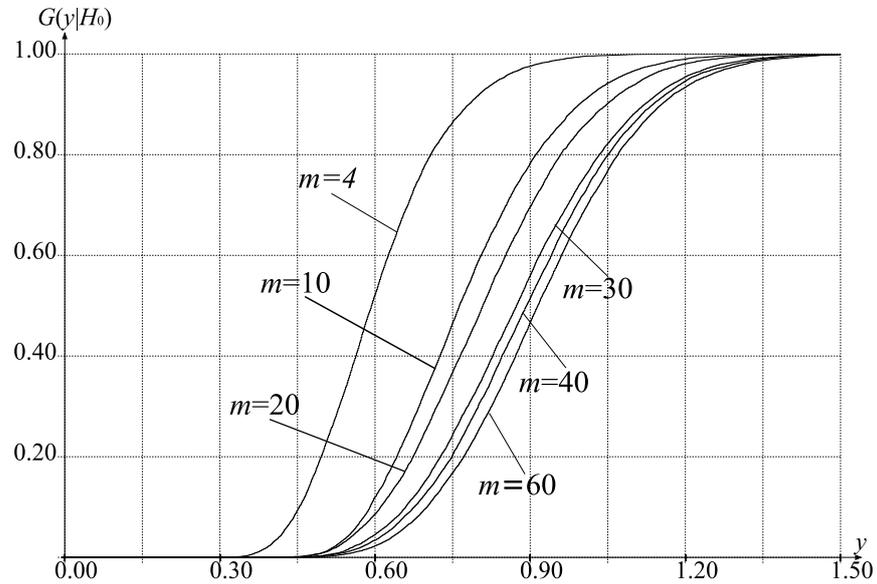


Рисунок 3.5 – Распределения статистики (3.11) для различного количества групп при $n = 120$

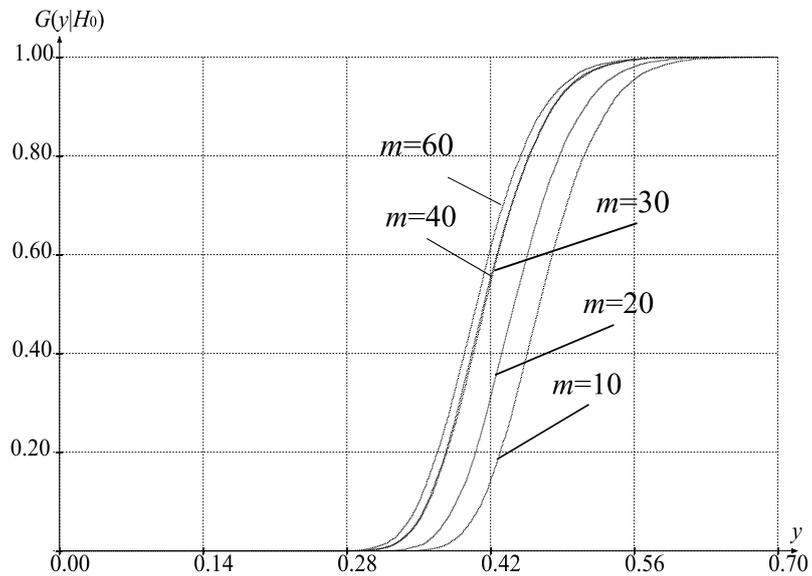


Рисунок 3.6 – Распределения статистики (3.12) для различного количества групп при объёме выборки $n = 120$

Далее приведены результаты исследования мощности предложенного критерия и сравнение с мощностью критериев на основе остатков. Для этого проводились те же эксперименты, что и в [102], и сравнивались результаты, полученные в случае полных данных для критерия Колмогорова, применяемого к выборкам остатков [9], и для критерия хи-квадрат в случае модели Кокса [102].

В качестве проверяемой гипотезы H_0 рассматривалась модель Кокса с экспоненциальным базовым распределением с параметром масштаба $\theta_1 = 21.6147$ и параметром $\beta = 0.2$. Моделируемые выборки случайных величин объемом $n = 100$ содержат бинарную ковариату. Количество моделируемых выборок $N = 16600$. В качестве конкурирующих гипотез рассматривались два типа обобщённых моделей с базовым экспоненциальным законом:

H_1^1 : модель Ксая $\Lambda_x(t; \beta, \gamma) = \exp(\beta^T \cdot x) \{ \Lambda_0(t) \}^{\exp(\gamma^T \cdot x)}$ с параметрами $\beta = 0.2, \gamma = 0.5$;

H_1^2 : SCE модель $\Lambda_x(t; \beta, \gamma) = \left(1 + \exp((\beta + \gamma)^T \cdot x) \cdot \Lambda_0(t) \right)^{\exp(-\gamma^T \cdot x)} - 1$ с параметрами $\beta = 0.2, \gamma = 0.8$.

Оценки мощности, полученные при различных уровнях значимости, представлены в таблице 3.1.

Как следует из результатов таблицы 3.1, предложенный критерий согласия на основе оценки Берана способен различить гипотезы о виде регрессионной зависимости с большей мощностью, чем критерий Колмогорова, основанный на остатках. Кроме того, критерий со статистикой (3.11) обладает большей мощностью, чем критерий со статистикой (3.12). Это объясняется тем, что в данном эксперименте количество наблюдений в группе достаточно велико (в каждой группе 50 наблюдений), а статистика (3.11) ориентирована на такие случаи.

Таблица 3.1 – Оценки мощности критериев согласия

Конкурирующая гипотеза	n	α	Предложенный критерий (3.11)	Предложенный критерий (3.12)	Критерий типа Колмогорова на основе остатков
H_1^1	50	0.15	0.77	0.70	0.64
		0.10	0.67	0.58	0.52
		0.05	0.55	0.45	0.36
	100	0.15	0.97	0.93	0.84
		0.10	0.94	0.90	0.75
		0.05	0.89	0.81	0.63
H_1^2	50	0.15	0.60	0.55	0.48
		0.10	0.52	0.47	0.38
		0.05	0.43	0.37	0.24
	100	0.15	0.86	0.82	0.72
		0.10	0.81	0.77	0.64
		0.05	0.73	0.67	0.54

Ниже мощность предложенного критерия согласия на основе оценки Берана анализируется для случая проверки гипотез относительно параметрической модели с определенным базовым распределением моментов отказов.

В первом случае в качестве гипотезы H_0 рассматривалась параметрическая модель Кокса с логарифмически линейной формой функции от ковариат, регрессионным параметром 0.2 и некоторой базовой функцией распределения F_0 . В качестве H_1 рассматривалась параметрическая модель с такой же функцией от ковариат, но отличной параметризацией базового закона распределения F_1 .

Ковариата принимала возможные значения 0,1,2 или 3, количество наблюдений в каждой группе равнялось 25, $n = 100$.

На рисунке 3.7 приведены функции надёжности при значении ковариаты 1 для параметрических моделей пропорциональных интенсивностей (PH) со следующими базовыми распределениями отказов:

- экспоненциальное распределение (Exp) с параметром $\theta_1 = 21.6147$;
- распределение Вейбулла (W) с параметрами $\theta_1 = 22.0, \theta_2 = 1.1$;
- гамма-распределение (Gm) с параметрами $\theta_1 = 22.0, \theta_2 = 0.7$;
- логнормальное распределение (Ln) с параметрами $\theta_1 = 20.0, \theta_2 = 1.15$.

Функции надёжности параметрических моделей пропорциональных интенсивностей с данными базовыми распределениями представлены на рисунке 3.7.

Как можно видеть из рисунка 3.7, некоторые из выбранных моделей достаточно близки, например, модель с экспоненциальным распределением и модель с распределением Вейбулла. Также видно, что модель с логнормальным распределением сильно отличается от модели с гамма-распределением.

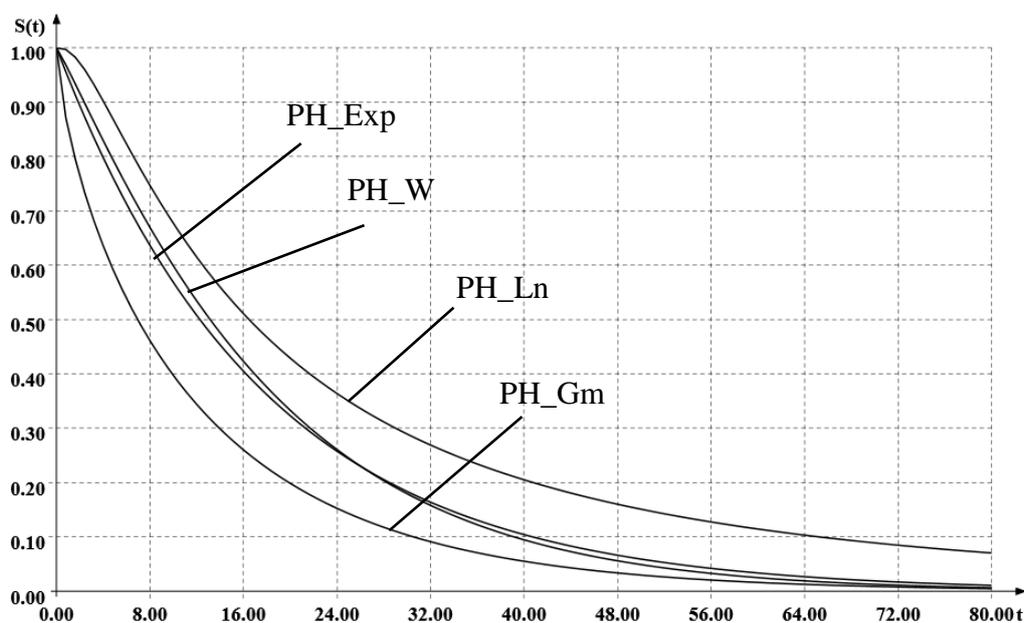


Рисунок 3.7 – Функции надёжности параметрических моделей пропорциональных интенсивностей с разными базовыми распределениями

В таблице 3.2 приведены значения оценок мощности для предложенного критерия в случае проверки гипотезы о согласии с параметрической моделью Кокса относительно конкурирующих гипотез, соответствующих той же параметрической модели Кокса, но с другим базовым распределением отказов. Также в таблице приведены значения мощности для критерия, основанного на остатках и статистики Колмогорова.

Таблица 3.2 – Оценки мощности предложенных критериев при проверке гипотезы о согласии с параметрической моделью Кокса с определенным базовым распределением отказов, $\alpha = 0.1$

	H_0	H_1			
		Exp	W	Gm	Ln
Предложенный критерий (3.11)	Exp	*	0.12	0.6	0.89
	W	0.10	*	0.10	0.78
	Gm	0.10	0.10	*	0.65
	Ln	0.21	0.28	0.19	*
Предложенный критерий (3.12)	Exp	*	0.10	0.78	0.88
	W	0.10	*	0.10	0.86
	Gm	0.10	0.10	*	0.80
	Ln	0.40	0.50	0.39	*
Критерий типа Колмогорова на основе остатков	Exp	*	0.29	0.73	0.86
	W	0.10	*	0.13	0.48
	Gm	0.10	0.10	*	0.55
	Ln	0.79	0.81	0.68	*

Как видно из таблицы 3.2, на этот раз предложенные критерии согласия в большинстве случаев уступают критерию, основанному на остатках. Это вполне логично, так как критерии, основанные на остатках, разрабатывались специально для проверки такого рода гипотез.

3.4 Выводы

В настоящей главе диссертации предложены новые критерии согласия для проверки гипотезы о виде параметрической регрессионной модели надёжности, основанные на использовании оценки Берана. Применение предложенных критериев предполагает использование распределений статистик $G_N(y|H_0)$, получаемых методами статистического моделирования, в том числе, в ходе проводимого анализа (в интерактивном режиме).

В результате исследования распределений статистик и мощности предложенных критериев согласия показано, что:

1. В случае конкурирующих гипотез, соответствующих другому виду регрессионной зависимости функции надёжности от ковариаты, предложенные критерии превосходят по мощности критерий Колмогорова, применяемый к выборкам остатков.

2. В случае конкурирующих гипотез, соответствующих другому базовому распределению отказов, предложенные критерии уступают по мощности критерию Колмогорова, применяемому к выборкам остатков.

ГЛАВА 4 Описание разработанного программного обеспечения и примеры его применения для анализа реальных данных

В данной главе представлено описание разработанного программного обеспечения, позволяющего осуществлять построение оценки Берана. Описываются основные этапы статистического анализа результатов эксперимента по определению коэффициента газообразования, проведённого в [68]. Приводится решение задачи статистического анализа продолжительности сотрудничества ООО «ДГ-Софт» с рекламодателями.

4.1 Описание разработанного программного обеспечения

В настоящее время среди известных и доступных автору программных систем статистического анализа данных ни одна не позволяет строить непараметрическую оценку Берана с выбором параметра размытости. Поэтому было принято решение о реализации возможности построения оценки Берана в программной системе «LiTiS» [110,111], которая разрабатывается совместно с к.т.н. Е.В. Чимитовой, А.В. Румянцевым, Н.С. Галановой и к.т.н. М.А. Семёновой.

Программная система «LiTiS» предназначена для анализа данных типа времени жизни, в ней реализованы модули оценивания параметров, проверки гипотез о виде распределения, построения моделей пропорциональных интенсивностей и ускоренных испытаний. Предпосылки к созданию данной системы описаны в [102].

В очередную версию программной системы «LiTiS 1.1» автором диссертации была добавлена возможность построения оценки Берана с использованием предлагаемого алгоритма выбора оптимального значения параметра размытости. Расширенная версия программной системы была зарегистрирована как «LiTiS 1.2» [112].

Для работы с оценкой Берана во вкладку «Предварительный анализ» был добавлен пункт меню «Оценка Берана», как показано на рисунке 4.1.

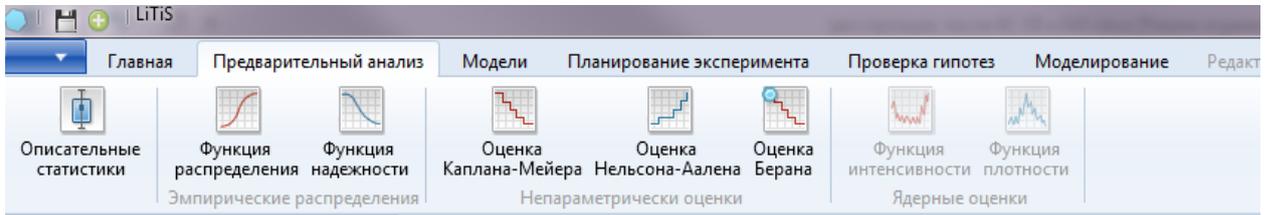


Рисунок 4.1 – Вкладка «Предварительный анализ» с кнопкой «Оценка Берана»

При нажатии на кнопку «Оценка Берана» появляется диалоговое окно, в котором необходимо выбрать значения ковариаты.

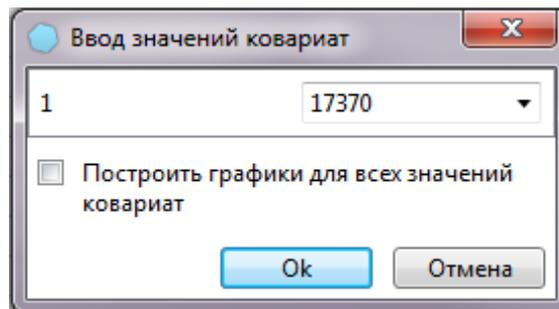


Рисунок 4.2 – Окно выбора значений ковариат

После выбора одного из возможных значений ковариаты (рисунок 4.2) можно построить оценку Берана для этой ковариаты, нажав кнопку «Ок». В результате построения оценки создается объект типа «График» и выводится на экран в основном окне программной системы, как показано на рисунке 4.3.

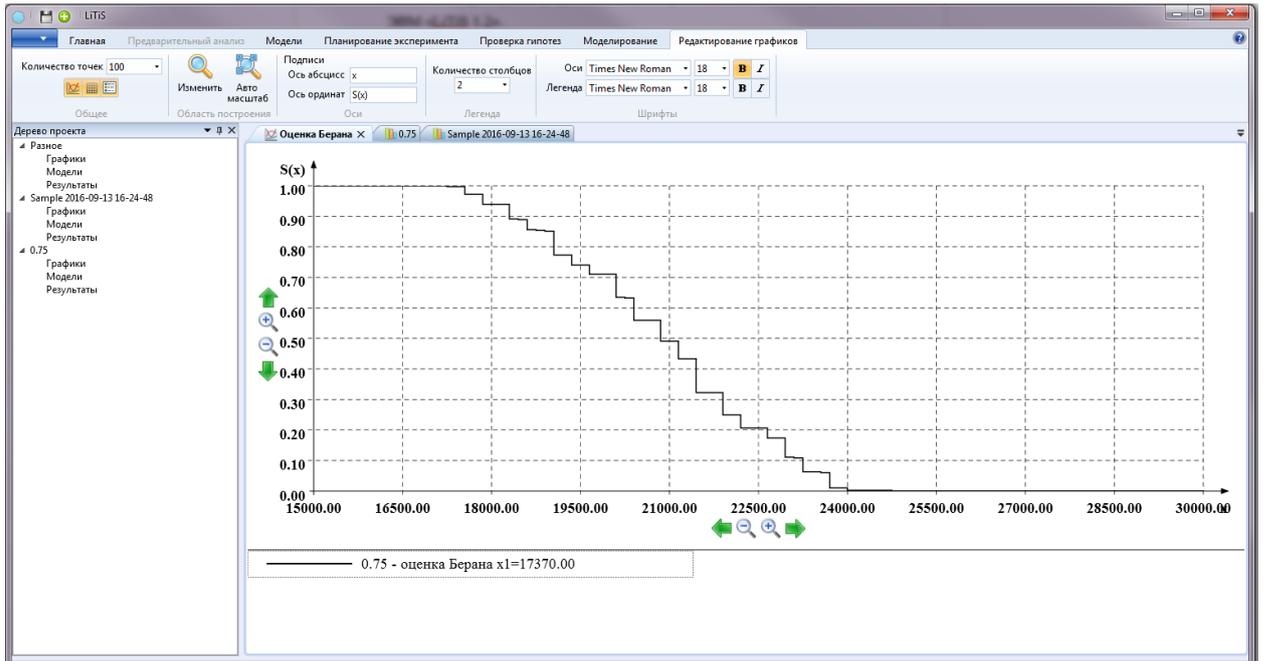


Рисунок 4.3 – Пример построения оценки Берана для заданного значения ковариаты

Если в окне ввода значений ковариаты поставить галочку в пункте «Построить графики для всех значений», то на результирующем графике будут построены оценки Берана для каждого заданного значения ковариаты (см. рисунок 4.4).

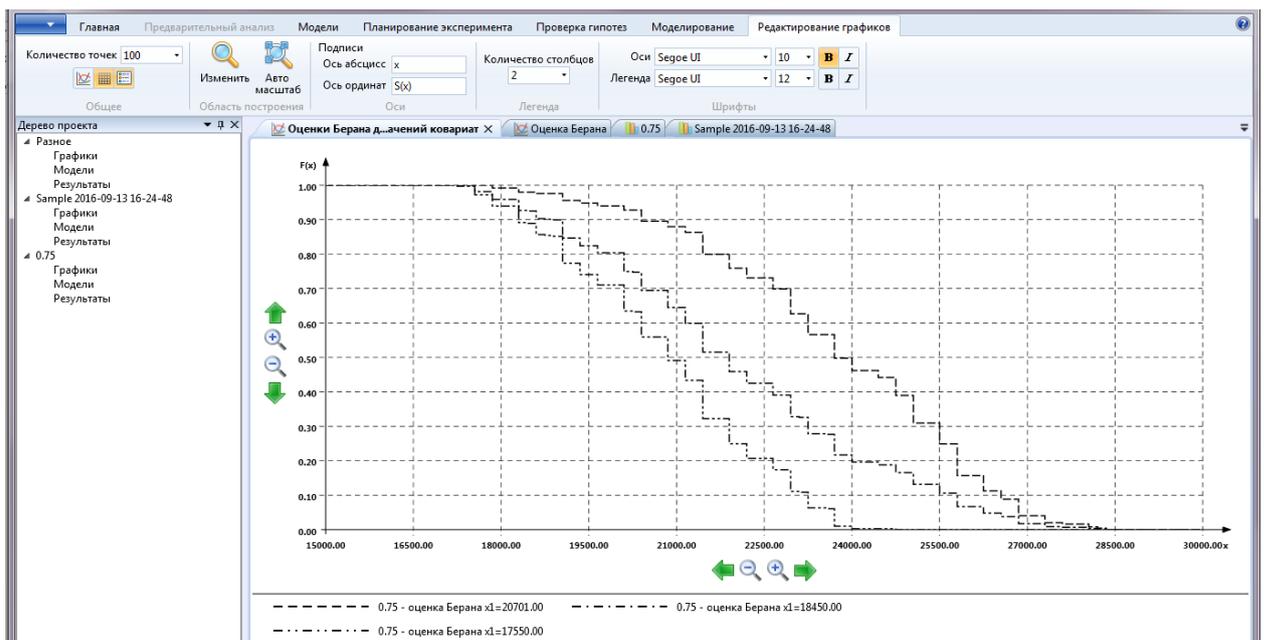


Рисунок 4.4 – Пример построения оценок Берана для всех значений ковариаты

Таким образом, пользователь имеет возможность построить оценки Берана для интересующих значений ковариаты или для всех опорных точек плана эксперимента, добавить на график теоретические функции надёжности и визуально оценить близость предполагаемых параметрических моделей надёжности к оценкам.

4.2 Анализ данных о частичных разрядах, возникающих в жидких диэлектриках

4.2.1 Постановка задачи

Одной из проблем современной энергетики являются аварии и нештатные ситуации на крупных объектах. На данный момент наблюдается рост таких аварий, что обычно связывают с тенденцией старения парка электрооборудования в России [72]. Замена всех устаревших объектов выглядит не только бессмысленной, но и экономически и технически не осуществимой. Диагностика и выявление наиболее аварийного оборудования является важной и актуальной задачей. Диагностика оборудования обычно значительно дешевле, чем замена. При этом на первый план выходят методы диагностики, которые обеспечивают контроль текущего состояния оборудования на месте его установки, под рабочим напряжением и, желательно, в процессе нормальной эксплуатации. Основным вопросом, на который должна ответить диагностическая система, является возможность или невозможность дальнейшей безопасной эксплуатации оборудования.

Наиболее информативными и часто применяемыми методами диагностики развивающихся повреждений на ранней стадии в настоящее время являются: регистрация частичных разрядов (ЧР) в бумажно-маслянистой изоляции, физико-химический анализ изоляционных жидкостей и газов.

В настоящее время существует несколько методик по определению коэффициента газообразования в изоляционных жидкостях. Принципиальным

отличием методик является различие в типах разрядов. При этом наиболее принципиальным моментом является определение энергозатраты в жидкость, приходящееся на газообразование, которое будет зависеть от типа разряда.

Известно, что при эксплуатации электрооборудования о степени износа судят по интенсивности и энергозатраты ЧР, от которых зависит количество образовавшихся газов. По концентрации газов и скорости её нарастания можно оценить степень развития дефектов. Существует несколько видов частичных разрядов [68]. Во-первых, это разряд в резко неоднородном поле, когда ЧР занимает малую область в той части изоляции, где существует резко неоднородное сильное поле. Во-вторых, разряд в газовой полости, либо в общем виде, в части промежутка с малой электрической прочностью.

С целью исследования энергозатраты в жидкость А.Л. Бычковым [68] были проведены экспериментальные исследования по определению коэффициента газообразования. В основе методики определения газостойкости лежит измерение характеристик частичных разрядов в диэлектрических жидкостях, с последующим вычислением коэффициента газообразования, который обратно пропорционален стойкости к разложению. Для проведения экспериментов и регистрации данных, выбрана экспериментальная ячейка с системой электродов «острие-плоскость». В таком случае принято считать, что «кажущийся заряд» примерно равен истинному.

Экспериментальная установка представляла собой металлический заземлённый сейф. В левой стенке сейфа вмонтирован высоковольтный маслонеполненный ввод, напряжение с которого через фильтрующий резистор подаётся на экспериментальную ячейку. Над ячейкой располагается ёмкость связи с измерительной схемой, экранированной медным корпусом. Для регистрации ЧР применялся цифровой осциллограф.

Для экспериментального определения коэффициента газообразования проводились следующие операции: дегазация трансформаторного масла, подготовка экспериментальной установки и заполнение ячейки

дегазированным трансформаторным маслом, определение оптимальной величины действующего значения напряжения (подбирается из максимально возможной скорости записи импульсов), определение концентрации газообразных продуктов в трансформаторном масле.

Таким образом, после дегазации трансформаторного масла, заполнялась ячейка, и отбирались первые исходные пробы. Далее, с помощью схемы (рисунок 4.5) проводилась регистрация импульса ЧР, мгновенного значения напряжения в момент ЧР, действующего значения напряжения.

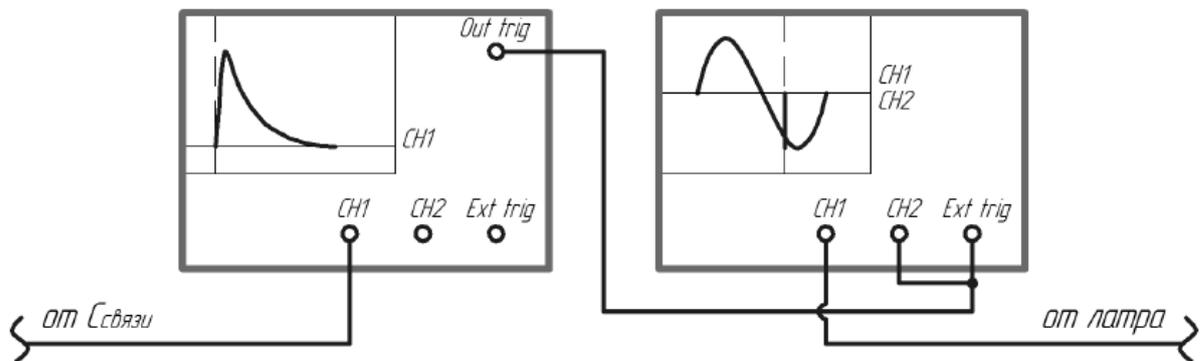


Рисунок 4.5 – Схема регистрации характеристик ЧР

В дальнейшем на основе измеренных параметров определялись: фазовое распределение импульсов ЧР, энергия единичного ЧР и суммарная энергия в ЧР, полученная за весь эксперимент. Также стоит добавить, что эксперименты проводились сериями в зависимости от действующего напряжения. Всего эксперимент длился 4-6 часов.

В результате была получена выборка частичных разрядов объемом $n = 626$, имеющая следующий вид:

$$\left\{ \left(U_i^{PD}, U_i, U_i^W, \varphi_i, Q_i, W_i \right), i = \overline{1, n} \right\},$$

где i – порядковый номер ЧР;

U_i^{PD} – амплитудное значение импульса ЧР;

U_i – мгновенное значение напряжения в момент ЧР;

U_i^W – действующее напряжение;

φ_i – фазовое распределение импульса ЧР;

Q_i – кажущийся заряд ЧР;

W_i – энергия единичного ЧР.

Наибольший интерес вызывает зависимость мгновенного напряжения (U_i) от действующего напряжения (U_i^W). В данном случае значение мгновенного напряжения можно считать случайной величиной, зависящей от действующего напряжения, которое можно считать ковариатой, тогда полученные значения можно представить в виде выборки:

$$\mathbf{X}_n = (U_1, U_1^W), (U_2, U_2^W), \dots, (U_n, U_n^W). \quad (4.1)$$

Кроме того, в качестве случайной величины можно рассмотреть фазовое распределение импульса ЧР φ_i , которое вычисляется как:

$$\varphi_i = \arcsin\left(\frac{U_i}{\sqrt{2} \cdot U_i^W}\right).$$

Из данного соотношения видно, что фазовое распределение импульса ЧР сочетает в себе как мгновенное значение напряжения в момент ЧР, так и действующее напряжение. Выборка в этом случае имеет вид:

$$\mathbf{Y}_n = (\varphi_1, U_1^W), (\varphi_2, U_2^W), \dots, (\varphi_n, U_n^W). \quad (4.2)$$

Таким образом, необходимо построить параметрическую вероятностную модель для описания условной функции распределения наблюдаемого напряжения при возникновении ЧР в зависимости от величины действующего напряжения U^W .

4.2.2 Предварительный анализ

На рисунках 4.6 и 4.7 представлены графики эмпирических функций распределения наблюдаемого напряжения и фазового распределения импульса

ЧР при различных значениях действующего напряжения, полученные по выборкам X_n и Y_n , соответственно.

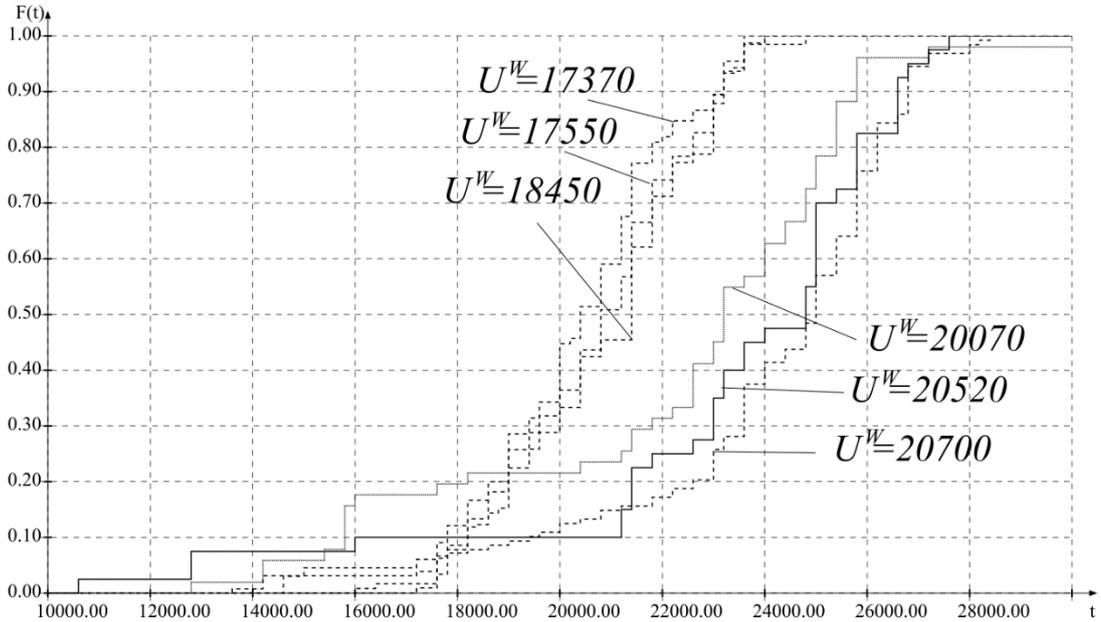


Рисунок 4.6 – Эмпирические функции распределения мгновенного напряжения

U_i при различных U^W

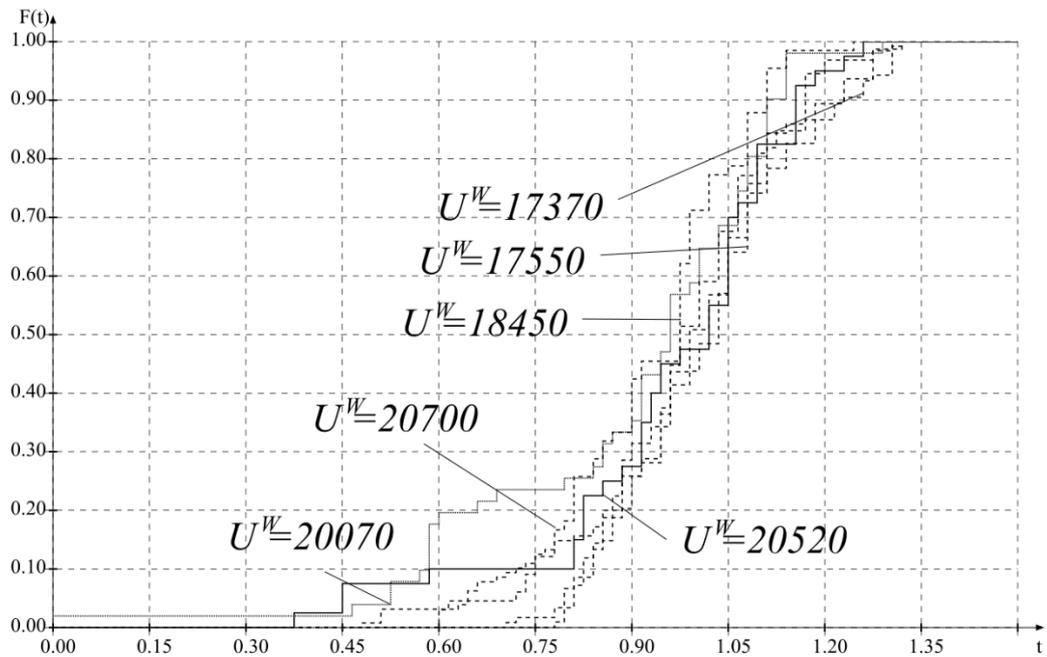


Рисунок 4.7 – Эмпирические функции распределения фазового распределения

импульса ЧР при различных U^W

Из рисунка 4.6 видно, что с увеличением действующего напряжения значения U_i стремятся в область больших значений. Однако, при рассмотрении фазового распределения импульса ЧР (рисунок 4.7) такого не происходит, и все функции распределения находятся примерно в одной окрестности. Таким образом, рассмотрение фазового отклонения ЧР, фактически, является способом нормировки, в которой элиминируется (исключается) влияние U^W .

Из рисунков 4.6 и 4.7 видно, что при большем значении действующего напряжения, форма распределения мгновенного напряжения и углов начинает меняться. При $U^W = 18450$ и выше наблюдается смесь двух различных распределений. При этом количество наблюдений, подчиняющихся другому распределению, увеличивается с ростом действующего напряжения: при $U^W = 18450$ в области левого хвоста, где наблюдается другое распределение, 6 наблюдений из 66; при $U^W = 20070$ таких наблюдений уже 11 из 50. Основное предположение заключается в том, что при больших значениях U^W начинают возникать два вида ЧР: не только ЧР в жидкости, но и ЧР в пузырьках газа.

Из рисунка 4.7 видно, что при $U^W < 18450$ ЧР со значением фазового распределения импульса менее 0.75 не возникают. Исходя из этого, можно сделать вывод, что ЧР, возникающие при значении фазового распределения импульса ЧР менее 0.75, можно отнести к ЧР в газе, а больше чем 0.75 – к ЧР в жидкости.

Сформулированная гипотеза подтверждается тем фактом, что ЧР в пузырьках газа имеют меньшее значение мгновенного напряжения, однако возникают при большем действующем напряжении. Также следует добавить, что расчёт энергозатрат от ЧР такого типа должен рассчитываться иначе. Таким образом, необходимо ввести поправку при расчёте коэффициента газообразования. Процент таких разрядов растёт с увеличением действующего напряжения и меняется в интервале от 9 до 24%.

Для построения вероятностной модели зависимости функции распределения мгновенного напряжения ЧР в жидкости от значения действующего напряжения были исключены из рассмотрения наблюдения, соответствующие ЧР в пузырьках газа и имеющих значение фазового напряжения импульса ЧР менее 0.75.

На рисунках 4.8 и 4.9 представлены графики эмпирических функций распределения мгновенных напряжений и углов (фазового распределения импульса ЧР), соответственно, полученные после исключения наблюдений, соответствующие ЧР в пузырьках газа.

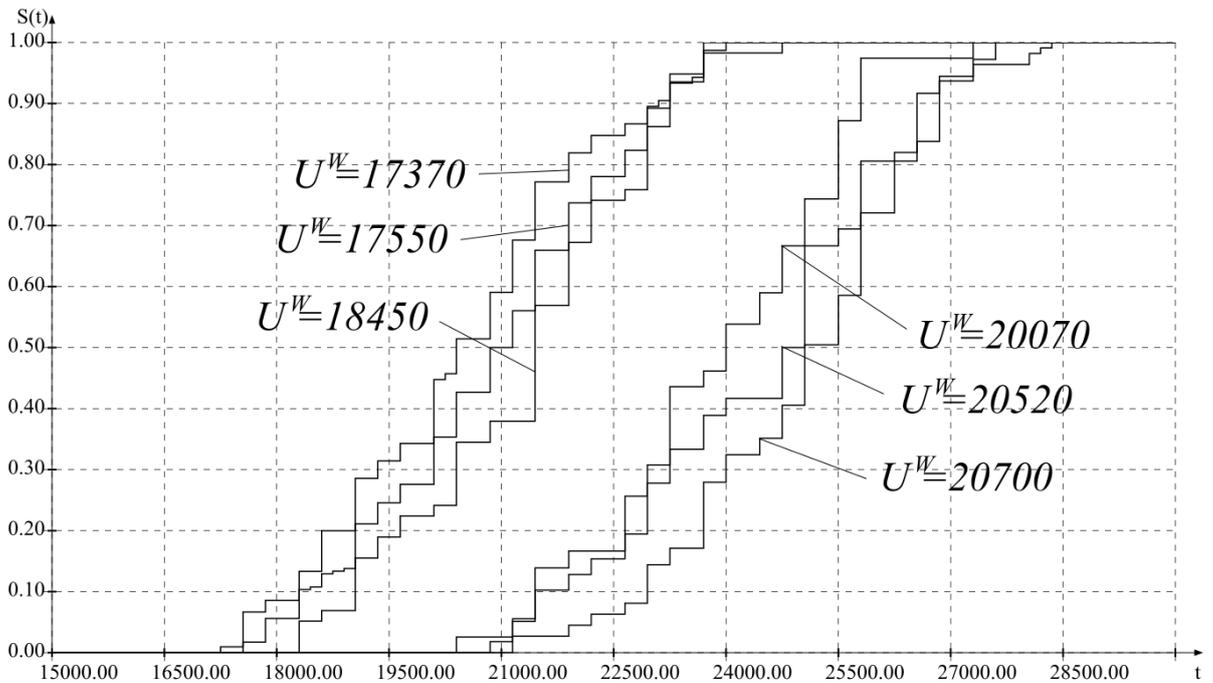


Рисунок 4.8 – Эмпирические функции распределения мгновенного напряжения U_i при различных U^W с усеченными левыми хвостами

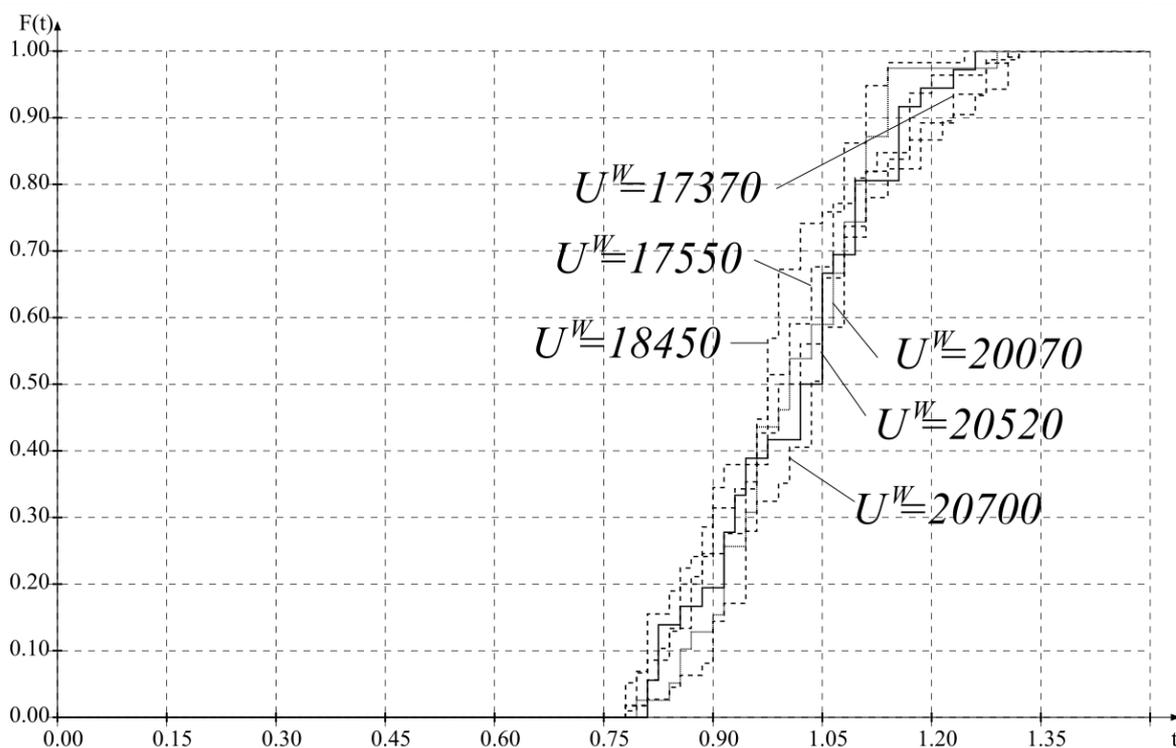


Рисунок 4.9 – Эмпирические функции распределения фазового распределения импульса ЧР при различных U^W с усеченными левыми хвостами

Из рисунка 4.8 видно, что с увеличением действующего напряжения мгновенное напряжение пропорционально увеличивается. Из рисунка 4.9 видно, что все функции распределения сгруппированы в одной области, что вполне логично и объяснимо, учитывая, что на графике фазового распределения импульса ЧР не должно проявляться влияние действующего напряжения.

4.2.3 Построение модели зависимости распределения мгновенного напряжения от значения действующего

После проведенной очистки данных стало возможным подобрать параметрическую вероятностную модель для описания функции надёжности мгновенного напряжения ЧР в зависимости от действующего напряжения.

Поскольку распределение мгновенного напряжения, как правило, хорошо описывается распределением Вейбулла [105], и при этом с изменением действующего напряжения меняется масштаб распределения, но не меняется его форма (см. рисунок 4.8), то логично было попробовать описать такую зависимость с помощью модели ускоренных испытаний вида

$$S_x(t) = S_0 \left(\frac{t}{r(x; \beta)} \right),$$

где в качестве функции от ковариат $r(x; \beta)$ рассмотрим функцию вида

$$r(x; \beta) = e^{\beta_0 + \beta_1 \ln x}.$$

Неизвестные параметры модели можно оценить с помощью метода максимального правдоподобия. Статистическая гипотеза о виде модели ускоренных испытаний проверялась с использованием предложенного в главе 3 критерия согласия на основе оценки Берана со статистикой (3.11). Проверяемая гипотеза о виде модели не отклоняется, так как достигнутый уровень значимости $p\text{-value} = 0.87$.

Таким образом, вероятность появления ЧР с заданным интервалом значений мгновенного напряжения U , в зависимости от величины действующего напряжения U^W можно вычислять на основе полученной модели:

$$S_x(t) = P\{U > t | x = U^W\} = \exp \left[- \left(\frac{t}{\exp(9.0207 + 5.4734 \cdot 10^{-5} x)} \right)^{14.3848} \right]. \quad (4.3)$$

На рисунке 4.10 представлены графики оценки Берана и соответствующих распределений (4.3) при различных действующих напряжениях U .

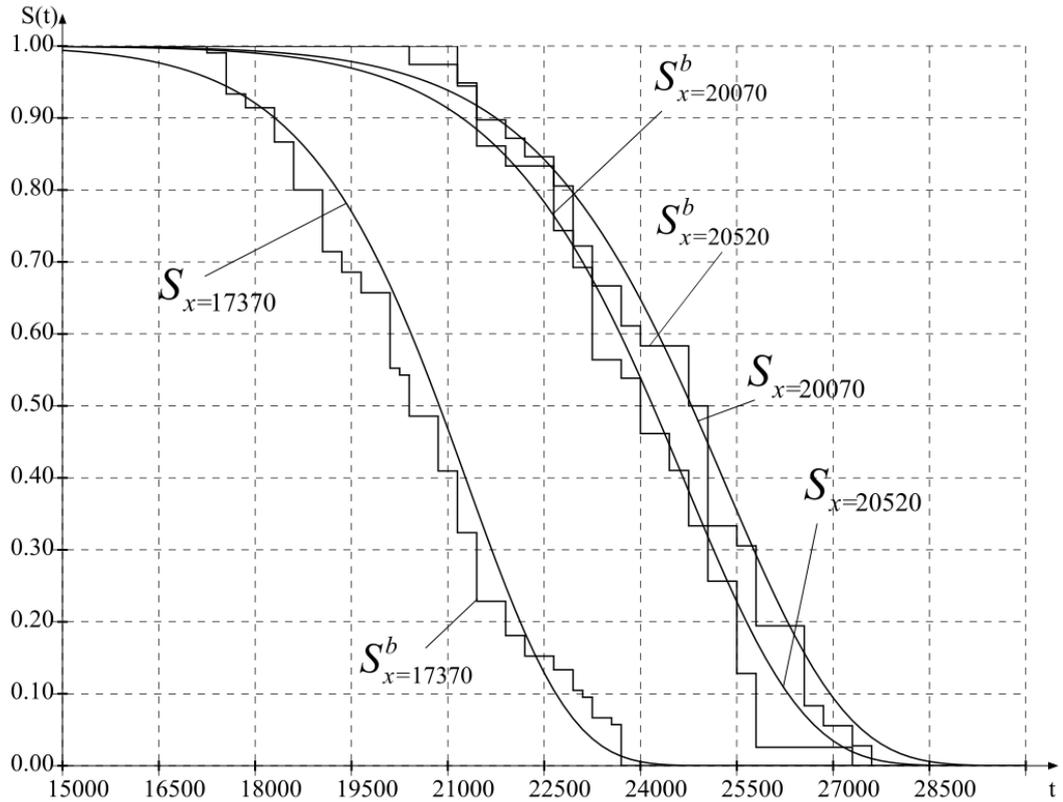


Рисунок 4.10 – Функции надёжности и оценки Берана при различных действующих напряжениях U

Из рисунка 4.10 видно, что подобранная модель достаточно хорошо согласуется с оценками Берана, построенными по выборкам при различных значениях действующего напряжения.

В таблице 4.1 представлены вычисленные на основе построенной модели вероятности возникновения ЧР со значением мгновенного напряжения, больше заданного для некоторых значений действующего напряжения. В таблице приведены значения для действующего напряжения $U^W = 17370, 20070, 20520$, однако подобные таблицы могут быть построены для любого значения действующего напряжения.

Таблица 4.1. Вероятности возникновения ЧР с заданными значениями мгновенного и действующего напряжения

U	U^W		
	17370	20070	20520
>18000	0.921	0.991	0.994
>21000	0.468	0.920	0.938
>24000	0.006	0.539	0.648
>27000	0.001	0.035	0.094
>30000	0	0.001	0.001

Таким образом, на основе полученной модели можно вычислить вероятность большего энерговклада в жидкость, и уровень её деградации.

Построение математической модели, способной описать распределение частичных разрядов в зависимости от действующего напряжения, позволяет моделировать процесс деградации изоляционных жидкостей, делать прогноз срока службы изоляционных жидкостей при различных напряжениях (даже для таких значений действующего напряжения, которые не наблюдались экспериментально). Обладая таким инструментом, можно экономить ресурсы для диагностики электрического оборудования.

4.3 Анализ вероятности продления контракта рекламодателем

4.3.1 Введение в предметную область

Компания ООО «ДГ-Софт» предоставляет бесплатный справочный и картографический сервис. Данный продукт называется «2ГИС». Монетизация (прибыль) компании происходит в большей степени за счёт продажи рекламных позиций (как в справочнике, так и на карте). Таким образом,

экономическое благополучие компании зависит от количества фирм-рекламодателей (РД). При работе с РД выделяются два основных направления: привлечение новых РД и удержание текущих РД. На данный момент задача удержания текущих РД стоит более остро, так как именно эти фирмы приносят большую часть прибыли компании.

Основным показателем, характеризующим эффективность работы с текущими РД является доля фирм, которые решили продолжить размещение рекламы после первого периода размещения. Процент таких фирм зависит от города: в некоторых городах процент оставшихся РД после первого периода размещения рекламы составляет чуть более 20%, тогда как приемлемым результатом является 40-56%.

Одним из направлений развития компании является решение задачи повышения процента продлений контракта с РД в «отстающих» городах, таких как Москва, Санкт-Петербург, Казань и других, в которых процент продлений контракта составляет менее 30%. Было выяснено, что в таких городах одним из факторов, влияющих на факт отказа от продления контракта, является количество купленных рекламных позиций в первый период размещения.

На рисунке 4.11 показаны примеры некоторых рекламных позиций, которые может купить РД: 1 – приоритет высшего порядка (поднятие в списке выдачи), 2 – указание логотипа фирмы в справочнике, 3 – увеличенный указатель (пин) на карте, 4 – медийный баннер на карте.

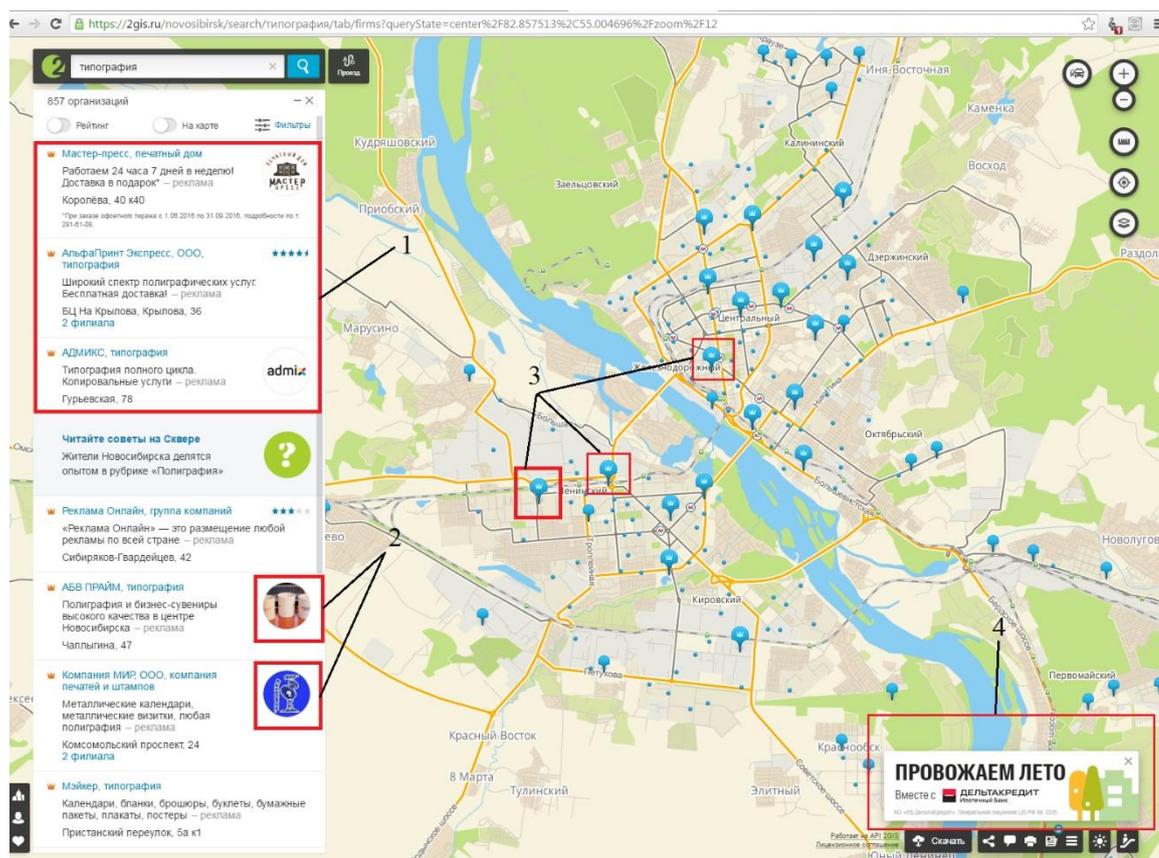


Рисунок 4.11 – Примеры рекламных позиций, которые может купить РД

В качестве времени наработки до отказа примем количество месяцев, в течении которых фирма размещала рекламу в 2ГИС до момента расторжения контракта. Вероятность безотказной работы (функция надёжности) зависит от количества рекламных позиций, поэтому в качестве ковариаты использовалось начальное количество рекламных позиций, купленных фирмой в первый месяц размещения.

Основным методом удержания РД является работа с этими фирмами менеджеров по продажам. Менеджер может совершать три вида действий по работе с клиентами: звонить в фирмы РД, назначать встречи, писать письма. Количество менеджеров в городе ограничено, соответственно и усилия, направляемые на удержания РД, крайне ограничены. На данный момент менеджеры по продажам осуществляют свои действия равномерно по всем фирмам. В среднем по одному звонку в один-два месяца на одну фирму.

Отправка писем и встречи происходят ещё реже. Поэтому необходимо понимать, в какой момент времени, какая группа РД (с каким количеством позиций) является наименее надёжной, и именно на неё направлять внимание менеджеров. Для анализа надёжности фирм необходимо построить оценку Берана.

Оценка надёжности фирм должна строиться в рамках одного города, так как в различных городах логика поведения РД сильно отличается из-за размера аудитории и качества работы менеджеров по продажам.

4.3.2 Построение оценки Берана для анализа надёжности фирм-рекламодателей в городе Казань

В исходную выборку попали данные о фирмах, купивших рекламу в «2ГИС» в апреле 2014 года в городе Казань. Процент продлений в этом городе составлял 28%. Особенность этих данных заключается в том, что часть фирм не расторгла контракт с 2ГИС в этот период, поэтому такие фирмы будут рассматриваться как цензурированные справа наблюдения. Выборка в этом случае будет иметь вид:

$$(Y_1, x_1, \delta_1), (Y_2, x_2, \delta_2), \dots, (Y_n, x_n, \delta_n),$$

где Y_i – количество месяцев до расторжения контракта i -й фирмой ($\delta_i = 1$), либо до момента цензурирования ($\delta_i = 0$);

x_i – начальное количество купленных позиций;

δ_i – индикатор цензурирования.

Объём выборки составил $n = 746$, из которых 155 наблюдений являются цензурированными справа (это фирмы, которые не прерывали контракт вплоть до сентября 2016 года).

На рисунке 4.12 представлены оценки Берана при $x = 1, 4, 10$, полученные при использовании предложенного алгоритма для выбора оптимального значения параметра размытости.

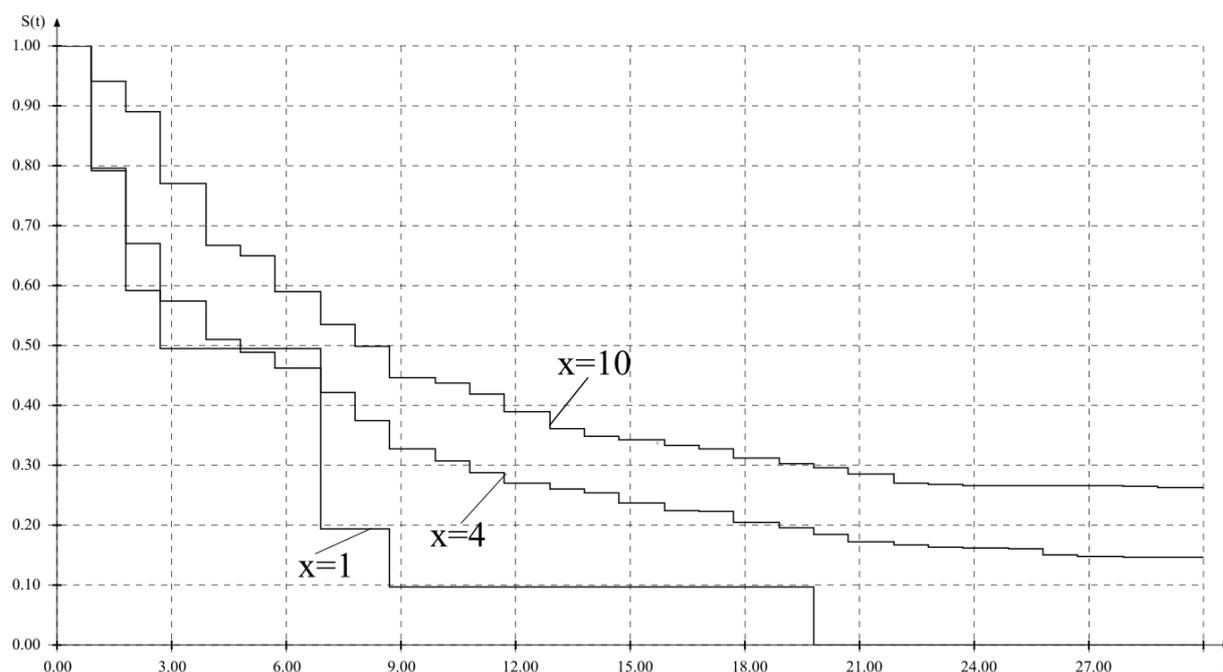


Рисунок 4.12 – Оценки Берана для фирм, которые купили 1, 4 или 10 позиций

Из рисунка 4.12 видно, что чем больше позиций купила фирма, тем больше вероятность её долгосрочного сотрудничества с «2ГИС». Это связано с тем, что эффект от рекламы при большем количестве позиций выше, чем при малом. Также видно, что для фирм с малым числом позиций критическим является седьмой месяц размещения, после которого произошло множество прерываний контрактов. Следовательно, стоит уделить повышенное внимание группам с малым числом позиций (1, 2) в период с шестого по девятый месяц размещения.

После девятого месяца размещения рекламы характер функции надёжности меняется как для группы с четырьмя позициям, так и для группы с десятью позициями: график функции становится более пологим. Другими словами, если РД с большим числом позиций «дожил» до девятого месяца, то вероятность того, что он расторгнет контракт в дальнейшем, падает. Поэтому на такие фирмы не стоит тратить ресурсы по удержанию клиентов.

Следует отметить, что для всех рассмотренных функций надёжности наблюдается резкое падение в первые три месяца, это объясняется тем, что фирмы не получили ожидаемый эффект от размещения рекламы и сразу прекратили сотрудничество. Следует добавить, что все фирмы, рассматриваемые как цензурированные справа наблюдения, являются фирмами с большим числом рекламных позиций.

Таким образом, можно сформулировать следующие рекомендации по работе менеджеров в городе Казань. Необходимо уделять равномерное влияние на РД только в первые три месяца. В дальнейшем необходимо дифференцированно проявлять активность в воздействии на разных РД. В период с шестого по девятый месяц необходимо проявлять большую активность по отношению к РД с малым числом купленных позиций, тогда как крупным РД можно уделять меньше времени после года размещения. Кроме того, следует стараться продавать больше рекламных позиций, так как при меньшем количестве рекламных позиций РД эффект от рекламы менее заметен в первые три месяца, которые важны для дальнейшего продления рекламы.

4.4 Выводы

На основе результатов исследований, изложенных в главах 2 и 3 настоящей диссертационной работы, разработан модуль программной системы статистического анализа данных «LiTiS». Версия программной системы с

реализованной оценкой Берана и методом выбора оптимального значения параметра размытости зарегистрирована как программа для ЭВМ [112].

Разработанные алгоритмы и программное обеспечение использовались при исследовании результатов экспериментов по определению коэффициента газообразования. Построена статистическая модель, описывающая зависимость распределения мгновенных напряжений от действующего напряжения. На основе полученной модели вычислены оценки вероятности возникновения частичных разрядов для некоторых значений действующего напряжения.

В результате статистического анализа продолжительности размещения рекламы фирмами в системе 2ГИС, проведенного с использованием разработанного программного обеспечения, сформулированы рекомендации по работе менеджеров по продажам компании «ДГ-Софт».

Заключение

В соответствии с поставленной целью диссертационного исследования получены следующие основные результаты.

1. Предложен адаптивный алгоритм выбора оптимального параметра размытости для построения непараметрической оценки Берана регрессионной модели надежности, который учитывает количество опорных точек плана эксперимента, вид ядерной функции, степень влияния ковариаты на функцию надёжности и значение ковариаты, для которой строится оценка.
2. Методами статистического моделирования исследованы свойства оценки Берана при использовании оптимальных значений параметра размытости, получаемых предложенным адаптивным алгоритмом. Показано, что по сравнению с ранее известными методами выбора параметра размытости применение его оптимальных значений позволяет получать более точные оценки Берана.
3. Предложены новые критерии согласия для проверки гипотезы о виде параметрической регрессионной модели надёжности, основанные на использовании оценки Берана. Методами статистического моделирования показано, что мощность предложенных критериев относительно конкурирующих гипотез, соответствующих другому виду регрессионной зависимости, превосходят мощность критерия типа Колмогорова, применяемого к выборкам остатков.
4. На основе результатов исследований разработан модуль программной системы статистического анализа данных типа времени жизни «LiTiS». Версия данной системы, в которой реализованы предложенный алгоритм выбора оптимального параметра размытости для оценки Берана и критерии согласия

зарегистрирована в виде объекта интеллектуальной собственности как программа для ЭВМ [112].

Результаты проведенных исследований и разработанное программное обеспечение были внедрены в практику деятельности ООО «ДГ-Софт» и филиала ПАО «Электросетьсервис ЕНЭС», а также нашли практическое применение в учебном процессе на факультете прикладной математики и информатики ФГБОУ ВО «Новосибирский государственный технический университет», что подтверждается соответствующими актами о внедрении.

СПИСОК ЛИТЕРАТУРЫ

1. Akritas, M.G. Nearest neighbor estimation of a bivariate distribution under random censoring / M.G. Akritas // Ann. Statist. – 1994. – Vol. 22. – P. 1299-1327.
2. Akushkina K.A. Models of statistical distributions of nonparametric goodness-of-fit tests in testing composite hypotheses of the generalized Weibull distribution / K.A. Akushkina, S.B. Lemeshko, B.Yu. Lemeshko // Proceedings Third International Conference on Accelerated Life Testing, Reliability-based Analysis and Design. 19-21 May 2010, Clermont-Ferrand, France. P. 125–132.
3. Anderson T.W. A test of goodness of fit / T.W. Anderson, D.A. Darling // Journal of the American Statistical Association, V. 29, 1954. –P. 765-769.
4. Bagdonavicius V. Accelerated life models : modeling and statistical analysis / V. Bagdonavicius, M. Nikulin // Boca Raton, Florida: Chapman & Hall/CRC.– 2002.– P. 334.
5. Bagdonavicius V. Accelerated Life Models. Modeling and Statistical Analysis / V. Bagdonavicius, M. Nikulin // Chapman&Hall/CRC. – 2002. – P.360.
6. Bagdonavicius V. Accelerated life testing / V. Bagdonavicius, M. Nikulin // John Wiley & Sons, Inc. – 2006.
7. Bagdonavicius V. Goodness-of-fit criteria for the Cox model from left truncated and right censored data / V. Bagdonavicius, R. Levulienė, M.S. Nikulin // Journal of Mathematical Sciences, No. 167(4).– Jun 2010. – P. 436-443.
8. Bagdonavicius V. Statistical modeling in survival analysis and its influence on the duration analysis / V. Bagdonavicius, M.A. Hafdi, M. Nikulin // In: Advances in Survival Analysis, Handbook of Statistics. Amsterdam: Elsevier. – P. 411–429.
9. Balakrishnan N. Testing goodness-of-fit of parametric AFT and PH models with residuals / N. Balakrishnan, E. Chimitova, N. Galanova, M. Vedernikova // Communications in Statistics - Simulation and Computation. – 2013. – Vol. 42. No. 6. – P.1352-1367.

10. Beran R. Nonparametric regression with randomly censored survival data. / R. Beran // Technical report. Department of Statistics, University of California, Berkeley. – 1981.
11. Chimitova E. A method for selection of the optimal bandwidth parameter for Beran's nonparametric estimator / E. Chimitova, V. Demin // Topics in statistical simulation: research papers from the 7 intern. workshop on statistical simulation. New York, Heidelberg, Dordrecht, London: Springer. – Vol. 114–2014. – P. 139-147.
12. Chimitova E. Application of classical Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests for censored samples / E. Chimitova, H. Liero, M. Vedernikova // Proceedings of the International Workshop “Applied Methods of Statistical Analysis. Simulations and Statistical Inference” – AMSA'2011, Novosibirsk, Russia, 20-22 September, 2011. – Novosibirsk: Publishing house of NSTU, 2011. – P. 176-185
13. Cox, D.R. Regression models and life tables (with Discussion) / D.R. Cox // Journal of the Royal Statistical Society. Ser. B. – 1972. – Vol. 34. – P. 187-220.
14. Crawford D.E. Analysis of Incomplete Life Test Data on Motorettes / D.E. Crawford // Insulation/Circuits. 1970. 16. P.43–48.
15. Dabrowska D. M. Nonparametric quantile regression with censored data / D. M. Dabrowska // Sankhya Ser. A. 54. – 1992. – P. 252-259.
16. Demin V. A goodness-of-fit test based on the Beran estimator / V. Demin, E. Chimitova // 11 International forum on strategic technology (IFOST 2016) : proc., Novosibirsk, 1–3 June 2016. – Novosibirsk : NSTU, 2016. – Pt. 1. – P. 483-487.
17. Demin, V. An adaptive method for selecting an optimal bandwidth parameter in nonparametric estimate of the conditional reliability function / V. Demin, E. Chimitova // Applied Methods of Statistical Analysis. Nonparametric Approach - AMSA'2015, Novosibirsk, Russia, 14-19 September, 2015: Proceedings of the International Workshop. - Novosibirsk: NSTU publisher, 2015. – P. 212-219.

18. Demin, V. Selection of the optimal smoothing parameter for the nonparametric estimation of the regression reliability model / V. Demin, E. Chimitova // Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control – AMSA'2013, Novosibirsk, 25–27 Sept. 2013: proc. of the intern. workshop. – Novosibirsk: NSTU publ., 2013. – P. 83-91.
19. Fuks I. Smooth estimation of multivariate reliability function / I. Fuks, G. Koshkin // Proceedings of the International Workshop "Applied methods of statistical analysis. Nonparametric approach – AMSA'2015", Novosibirsk, 14–19 Sept. 2015 – Novosibirsk & Belokurikha: NSTU publ., 2015. – P. 18-29.
20. Gang L. A bootstrap approach to nonparametric regression for right censored data / L. Gang, D. Somnath // technical report. – No 99-8. – 1999. – P. 6-10.
21. Gasser T. The choice of weights in kernel regression estimation / T. Gasser, J. Engel // *Biometrika*. – 1990. – P. 77, 377-381.
22. Gonzalez M., Cadarso S. Asymptotic properties of a generalized Kaplan-Meier estimator with some application / M.W. Gonzalez, S.C. Cadarso // *J. Nonparametric Statistics*, 4. – 1994. – P. 65-78.
23. Harrel F.E. Regression modeling with applications to linear models, logistic / Harrel F.E. // New York: Springer. – 2001.
24. Hjort N. L. Goodness of fit tests in models for life history data based on cumulative hazard rates / N. L. Hjort // Norwegian Computing Centre and University of Oslo / *the Annals of Statistics*. – Norway, 1990. – V. 18, No. 3. – P. 1221-1258.
25. Hjort N.L. On inference in parametric survival data / N.L. Hjort // *International Statistical Review*. – Vol. 60. – No. 3. – 1992. – P. 355-387.
26. Hsieh F. On heteroscedastic hazards regression models: theory and application / Hsieh F. // *Journal of the Royal Statistical Society*. – No. 63. – 2001. – P. 63-79.
27. Huber C. Mathematical Methods in Survival Analysis, Reliability and Quality of Life / C. Huber, N. Limnius, M. Nikulin // New Jersey: Wiley-ISTE. – 2008. – P. 420.
28. Kalbeisch J.D. The statistical analysis of failure time data / J.D. Kalbeisch, R.L. Prentice // John Wiley and Sons, Inc., New York. – 2002. – P.462.

29. Kalbfleisch, J.D. The statistical analysis of failure time data // J.D. Kalbfleisch, R.L. Prentice. – New York : John Wiley and Sons, Inc. – 1980. – P. 439.
30. Kalos, M.H. Monte Carlo methods. / M.H. Kalos, P.A. Whitlock. – New York: John Wiley and Sons, 1986.
31. Kaplan E.L. Nonparametric estimation from incomplete observations / E.L. Kaplan, P. Meier // Journal of the American Statistical Association. – 1958. Vol. 53. – P. 457-481.
32. Klein J.P. Survival analysis: techniques for censored and truncated data / J.P. Klein, M.L. Moeschberger // Springer, New York. 2003. 536 p.
33. Koziol J.A., Green S.B. A Cramer-von Mises statistic for randomly censored data / J.A. Koziol, S.B. Green // Biometrika, V. 63. – No. 3. – 1976. – P. 465-474.
34. Lawless J.F. Statistical models and methods for lifetime data / J.F. Lawless // John Wiley and Sons, Inc., Hoboken, New Jersey. – 2002. – P. 664.
35. Lee E. Statistical methods for survival data analysis / E. Lee, J. Wang // 3rd ed. Hoboken, New Jersey: John Wiley & Sons, Inc. – 2003. – P. 534.
36. Lee E.T., Wang J.W. Statistical Methods for Survival Data Analysis / E.T. Lee, J.W. Wang // John Wiley & Sons Inc., Hoboken, New Jersey. – 2003. – P. 513.
37. Lemeshko B.Yu. Construction of Statistic Distribution Models for Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses: The Computer Approach / B.Yu. Lemeshko, S.B. Lemeshko // Quality Technology & Quantitative Management, 2011. 8, № 4. . 359–373.
38. Lemeshko B.Yu. Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses / B.Yu. Lemeshko, S.B. Lemeshko, S.N. Postovalov // Communications in Statistics – Theory and Methods. – 2010. 39. – No 3. – P. 460–471.
39. Lemeshko B.Yu., Lemeshko S.B. Models of Statistic Distributions of Nonparametric Goodness-of-Fit Tests in Composite Hypotheses Testing for Double Exponential Law Cases / B.Yu. Lemeshko, S.B. Lemeshko // Communications in Statistics – Theory and Methods, 2011. 40, No 16. – P. 2879–2892.

40. Liero H. Testing the Hazard Rate, Part I / H. Liero // Germany: Department of Mathematics, University of Potsdam, 2003. – 19 c.
41. Martinussen T. Dynamic regression models for survival data / T. Martinussen, T.H. Scheike // In: Statistics for biology and health / Ed. by Gail M., Samet J.M., Tsiatis A., and Wong. Springer. – 2006. P. 173.
42. McKeague I. W., Utikal K. J. Inference for a nonlinear counting process regression model / I. W. McKeague, K. J. Utikal // Ann. Statist. – 18. – 1990. – P. 1172-1187.
43. Meeker W.Q., Escobar L.A. Statistical methods for reliability data / W.Q. Meeker, L.A. Escobar // John Wiley & Sons, 1998. – P.712.
44. Nair V. Plots and tests for goodness of fit with randomly censored data / V. Nair // Biometrika, Vol. 68, 1981. pp. 99-103.
45. Nelson W. Accelerated testing: Statistical models, test plans, and data analysis / W. Nelson // New York: John Wiley and Sons. – 2004. – P. 601.
46. Nikulin M. Accelerated Life Models / M. Nikulin // Encyclopedia of Statistics in Quality and Reliability. – 2007. – P. 503–506.
47. Nikulin M. Step Stress Accelerated Life Testing: Classical Methods / M. Nikulin // Encyclopedia of Statistics in Quality and Reliability. – 2007. – P. 57–62.
48. Nikulin M. Flexible regression models for carcinogenesis studies / M. Nikulin, H.I. Wu // Journal of Mathematical Sciences. – New York, 2007. – Vol. 145, № 2. – C. 4880-4893.
49. Nikulin M.S. Parametric and semiparametric models with applications to reliability, survival analysis, and quality of life / M.S. Nikulin, N. Balakrishnan, M. Mesbah, N. Limnios // Statistics for Industry and Technology / Ed. by Balakrishnan N. New York: Springer Science+Business Media. – 2004.
50. Panneton, F.O. Improved long-period generators based on linear recurrences modulo 2 / F.O. Panneton, P. L'Ecuyer, M. Matsumoto // ACM Transactions on Mathematical Software. – 2006. – Vol. 32. No. 1. – P. 1-16.

51. Pardo-Fernandez J.C. Goodness-of-fit tests for parametric models in censored regression / J.C. Pardo-Fernandez, I. Van Keilegom, W. Gonzalez-Manteiga // *The Canadian Journal of Statistics*. – 2007. – Vol. 35. No. 2. – P. 249-264.
52. Politis D. Adaptive estimation of density function derivative / D. Politis, V. Vasiliev, P. Tarassenko // *Proceedings of the International Workshop "Applied methods of statistical analysis. Nonparametric approach – AMSA'2015"*, Novosibirsk, 14–19 Sept. 2015 – Novosibirsk & Belokurikha: NSTU publ., 2015. – P. 56-64.
53. Prentice R.L. Exponential survivals with censoring and explanatory variables / R.L. Prentice // *Biometrika*. 1973. 60, № 2. P. 279–288.
54. Reineke D. Estimation of hazard, density and survival functions for randomly censored data / D. Reineke, J. Crown // *Journal of Applied Statistics*, Vol. 31, No. 10, 2004. – P. 1211-1225.
55. Rousseeuw P.J. Robust estimation in very small samples / P.J. Rousseeuw, S. Verboven // *Journal Computational Statistics & Data Analysis*. – 2002. – Vol. 40(4). – P. 741-758.
56. Semenova M. Parametric models in the analysis of patients with multiple myeloma / M. Semenova, A. Bitukov // *Proceedings of the International Workshop AMSA*. 2013. pp. 250-256.
57. Stablein D.M. A two sample test sensitive to crossing hazards in uncensored / D.M. Stablein, I.A. Koutrouvelis // *Biometrics*. – No. 41. – 1985. – P. 643-652.
58. Van Keilegom I. Estimation of the conditional distribution in regression with censored data: a comparative study / M.G. Akritas, N. Veraverbeke // *Computational Statistics & Data Analysis*. – 2001. – Vol. 35. – P. 487-500.
59. Van Keilegom I. Nonparametric estimation of the Conditional Distribution in Regression with Censored Data / I. Van Keilegom // *Limburgs Universitair Centrum*. – 1998. – P. 148.
60. Van Keilegom I. Nonparametric Estimation of the Conditional Distribution in Regression with Censored Data / I. Van Keilegom, N. Veraverbeke // *Dissertation*. 1998. - P. 7-51.

61. Van Keilegom I., Gonzalez-Manteiga W., Sellero C.S. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution / I. Van Keilegom, W. Gonzalez-Manteiga, C.S. Sellero // *Test.* – 2008. – Vol. 17. Issue 2. – P. 401-415.
62. Антонов А.В. Статистические модели в теории надежности / А.В. Антонов, М.С. Никулин // Учеб. пособие. М.: Абрис. – 2012. – 390 с.
63. Аронов И.З., Бурдасов Е.И. Оценка надежности по результатам сокращенных испытаний / И.З. Аронов, Е.И. Бурдасов // Москва: Изд-во стандартов, 1987.– 184 с.
64. Бессмельцев С.С. Бортезомиб (велкейд) в индукционной терапии множественной миеломы / С.С. Бессмельцев, Л.В. Стельмашенко, Е.В. Карягина, Н.В. Степанова, Е.Р. Мачюлайтене, Г.Н. Салогуб, Л.М. Матюхина, Низамутдинова А.С., Костина О.Я., Абдулкадыров К.М. // *Клиническая онкогематология. Фундаментальные исследования и клиническая практика*, Т. 1, № 4, 2008. С. 315-322.
65. Большев Л.Н. Таблицы математической статистики / Л.Н. Большев, Н.В. Смирнов // – М.: Наука, 1983. – 416 с.
66. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. / В.П. Боровиков // Учебное пособие для вузов. Москва: Горячая линия - Телеком, 2013. – 288 с.
67. Буртаев Ю.Ф. Статистический анализ надежности объектов по ограниченной информации / Ю.Ф. Буртаев, В.А. Острейковский // Москва: Энергоатомиздат, 1995. 240 с.
68. Бычков, А.Л. Исследование газообразования при частичных разрядах и совершенствование пробоотбора для газового анализа высоковольтного маслонеполненного электрооборудования: дис. ... канд. техн. наук: 05.14.12. / Бычков Александр Леонидович. – Новосибирск, 2014. – 156 с.
69. Ведерникова М.А. Тема: Отчёт о НИР// НГТУ, каф. ПМТ; исполн.: М.А. Вдерникова. – Новосибирск, 2011.– 44 с.
70. Галанова Н. С. Тема: Магистерская диссертация // Н.С. Галанова; НГТУ; науч. рук. Е.В. Чимитова. – Новосибирск, 2011.-84

71. Галанова Н.С. Применение непараметрических критериев согласия к проверке адекватности моделей ускоренных испытаний / Н.С. Галанова, Б.Ю. Лемешко, Е.В. Чимитова // Автометрия. – 2012. – № 6. – С.53-68.
72. Дарьян, Л.А. Научные основы физико-химической диагностики высоковольтного маслонаполненного электрооборудования с изоляцией конденсаторного типа: дис. ... д-ра техн. наук: 05.14.12 / Дарьян Леонид Альбертович. – Новосибирск, 2009. – 437 с.
73. Демин В. А. Разработка критерия согласия для регрессионной модели надежности на основе оценки Берана / Е. В. Чимитова, В. А. Демин, Т. А. Лисичкина // Актуальные проблемы электронного приборостроения (АПЭП–2014) = Actual problems of electronic instrument engineering (APEIE–2014) : тр. 12 междунар. конф., Новосибирск, 2–4 окт. 2014 г. : в 7 т. – Новосибирск : Изд-во НГТУ, 2014. – Т. 6. – С. 103-107.
74. Демин В.А. Адаптивный метод выбора параметра размытости для оценки Берана / В.А. Демин // Материалы всероссийской научной конференции молодых учёных «Наука. Технологии. Инновации». – Новосибирск, 2015. – С. 17-19.
75. Демин В.А. Выбор оптимального параметра сглаживания для непараметрической оценки регрессионной модели надежности / В.А. Демин, Е.В. Чимитова // Вестник ТГУ. Управление, вычислительная техника и информатика, №1(22). – 2013. – С. 59-65.
76. Демин В.А. Исследование метода выбора оптимального параметра сглаживания при непараметрическом оценивании регрессионных моделей надежности / В.А. Демин, Е.В. Чимитова, В.Ю. Щеколдин // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. - 2014. - № 2 (27). - С.10-18.
77. Демин В.А. Непараметрическая оценка регрессионной модели надёжности. Материалы конференций / В.А. Демин, Е.В, Чимитова // Обработка информационных сигналов и математическое моделирование. – Новосибирск. – 2012

78. Демин В.А. Непараметрические оценки и критерий согласия для регрессионных моделей / В.А. Демин, Е.В. Чимитова // Материалы всероссийской научной конференции молодых учёных «Наука. Технологии. Инновации». – Новосибирск, 2011. – С. 93-95.
79. Демин В.А. Разработка и исследование критериев согласия для параметрических регрессионных моделей надёжности на основе оценки Берана / В.А. Демин // Доклады Академии наук высшей школы Российской Федерации. - 2016. - № 2 (31). - С. 43-56.
80. Князева М.С. Критерии проверки гипотез о согласии на основе оценки Нельсона-Аалена по цензурированным данным / М.С. Князева, Е.В. Чимитова // Материалы российской НТК «Обработка информационных сигналов и математическое моделирование». – Новосибирск, 2012. – С. 36-39.
81. Кокс Д.Р. Анализ данных типа времени жизни / Д.Р. Кокс, Д. Оукс // М.: Финансы и статистика, 1988. 191 с.
82. Кошкин Г.А. Основы страховой математики / Г.А. Кошкин // Учебное пособие/ Томск: Томский государственный университет, 2002. – 116 с.
83. Лемешко Б.Ю. К оцениванию параметров надёжности по цензурированным выборкам / Б.Ю. Лемешко, С.Я. Гильдебрант, С.Н. Постовалов // Заводская лаборатория. Диагностика материалов. 2001. – 67, № 1. – С. 52–64.
84. Лемешко Б.Ю. Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. I / Б.Ю. Лемешко, С.Б. Лемешко // Измерительная техника. 2009. № 6. С. 3–11.
85. Лемешко Б.Ю. Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. II / Б.Ю. Лемешко, С.Б. Лемешко // Измерительная техника. 2009. № 8. С. 17–26.
86. Лемешко Б.Ю. Моделирование распределений статистик непараметрических критериев согласия при проверке сложных гипотез

- относительно обратного гауссовского закона / Б.Ю. Лемешко, С.Б. Лемешко, М.С. Никулин, Н. Сааидиа // Автоматика и телемеханика, 2010. № 7. С. 83–102.
87. Лемешко Б.Ю. Об оценивании параметров распределений и проверке гипотез по цензурированным выборкам / Б.Ю. Лемешко // Методы менеджмента качества. 2001. – № 4. – С.32-38.
88. Лемешко, Б.Ю. Непараметрические критерии при проверке сложных гипотез о согласии с распределениями экспоненциального семейства / Б.Ю. Лемешко, А.А. Маклаков // Автометрия. – 2004. – № 3. – С. 3-20.
89. Лемешко, Б.Ю. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов / Б.Ю. Лемешко, С.Н. Постовалов // Заводская лаборатория. Диагностика материалов. – 1998. – Т. 64. № 3. – С. 61-72.
90. Лемешко, Б.Ю. Применение непараметрических критериев согласия при проверке сложных гипотез / Б.Ю. Лемешко, С.Н. Постовалов // Автометрия. – 2001. – № 2. – С. 88-102.
91. Лемешко, Б.Ю. Проверка простых и сложных гипотез о согласии по цензурированным выборкам / Б.Ю. Лемешко, Е.В. Чимитова, Т.А. Плешкова // Научный вестник НГТУ. – 2010. – № 4(41). – С. 13-28.
92. Лемешко, Б.Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография. // Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. Новосибирск: Изд-во НГТУ, 2011. – 888 с. (серия "Монографии НГТУ").
93. Лемешко, Б.Ю. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие. / Б.Ю. Лемешко, С.Н. Постовалов. – Новосибирск: Изд-во НГТУ, 2004. – 119 с.
94. Медведев А.В. Основы теории адаптивных систем: монография / А.В. Медведев. – Красноярск: Изд-во СибГАУ, 2015. – 526 с.
95. Осипов Д.С. Совершенствование СМК машиностроительного предприятия с использованием методов экспертно-функционального анализа / Д.С.

- Осипов, С.В. Овчинников, В.В. Майстренко, И.Ю. Мезин // Методы менеджмента качества. – No. 9, 2012. – С. 46-50.
96. Половко А. М., Гуров С.В. Основы теории надежности / А. М. Половко, С.В. Гуров // СПб: БХВ-Петербург. – 2006. – 704 с.
97. Птушкин В.В. Влияние низкомолекулярных гепаринов на выживаемость онкологических больных / В.В. Птушкин // Онкогематология, No 1, 2007. С. 57-60.
98. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. – М.: Изд-во стандартов. 2002. 64 с.
99. Расин Д. Непараметрическая эконометрика: вводный курс / Д.Расин // Квантиль, №.4. – 2008. – С.7-26.
100. Рыбалко В.В. Определение закона надежности высоконадежных и малосерийных объектов по случайно цензурированным выборкам / В.В. Рыбалко // Exponenta Pro: Методы. Алгоритмы. Программы. – 2003. – № 1(1). – С. 44-47.
101. Семёнова М. А. Критерии согласия в задачах проверки адекватности параметрических моделей надежности и выживаемости / М. А. Семёнова, В. А. Демин, Е. В. Чимитова // Материалы Российской научно-технической конференции «Обработка информационных сигналов и математическое моделирование». – Новосибирск, 2013. – С. 38-40.
102. Семёнова М.А. Разработка алгоритмического обеспечения и исследование обобщенных моделей пропорциональных интенсивностей: дис. ... канд. техн. наук: 05.13.17. / Семёнова Мария Александровна. – Новосибирск, 2015. – 154 с.
103. Скрипник, В.М. Анализ надежности технических систем по цензурированным выборкам. / В.М. Скрипник, А.Е. Назин, Ю.Г. Приходько, Ю.Н. Благовещенский. – М.: Радио и связь, 1988. – 183 с.
104. Хардле В. Прикладная непараметрическая регрессия / В. Хардле // М.: Мир. – 1993. – 349 с.

105. Хаушильд В. Статистика для электротехников в приложении к технике высоких напряжений / В. Хаушильд, В. Мош // Пер. с нем. – Л.: Энергоатомиздат. Ленингр. отделение, 1989. – 312 с.
106. Чимитова Е.В. Проверка адекватности параметрических регрессионных моделей надежности по усеченным слева и цензурированным справа данным / Е.В. Чимитова, М.А. Семёнова // Доклады Академии наук высшей школы Российской Федерации. - 2015. - № 1 (26). - С. 104-120.
107. Чимитова Е.В., Галанова Н.С. Оценивание параметров модели ускоренных испытаний при исследовании надежности систем // Информатика и проблемы телекоммуникаций: Материалы конференции. – Новосибирск, 2009. – Том 1. – С. 70-71.
108. Чимитова, Е.В. Выбор параметризации базовой функции распределения для АFT-модели с помощью методов компьютерного моделирования / Е.В. Чимитова, Н.С. Галанова // Материалы X международной конференции “Актуальные проблемы электронного приборостроения” АПЭП-2010. Т.6, Новосибирск, 2010. – С.31-35.
109. Чимитова, Е.В. Непараметрические критерии согласия в задачах проверки адекватности моделей надежности по цензурированным данным / Е.В. Чимитова, М.А. Ведерникова, Н.С. Галанова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2013. – №4(25). – С. 115-124.
110. Чимитова, Е.В. Система статистического анализа данных типа времени жизни «LiTiS 1.0» / Е.В. Чимитова, А.В. Румянцев, М.А. Ведерникова, Н.С. Галанова // М.: Роспатент. Свидетельство о государственной регистрации программы для ЭВМ №2012618138 от 7.09.2012.
111. Чимитова, Е.В. Система статистического анализа данных типа времени жизни «LiTiS 1.1» / Е.В. Чимитова, А.В. Румянцев, М.А. Ведерникова, Н.С. Галанова // М.: Роспатент. Свидетельство о государственной регистрации программы для ЭВМ №2012618143 от 7.09.2012.
112. Чимитова, Е.В. Система статистического анализа данных типа времени жизни «LiTiS 1.2» / Е.В. Чимитова, А.В. Румянцев, М.А. Семёнова, Н.С.

Галанова, В.А. Демин // М.: Роспатент. Свидетельство о государственной регистрации программы для ЭВМ №2015610901 от 20.01.2015.

113. Шпер, В.Л. Реферативный аналитический обзор наиболее значимых публикаций в отечественной и зарубежной периодике по вопросам оценки надежности продукции, в том числе об опыте предприятий / В.Л. Шпер // Reliability: Theory & Applications, 2006. – № 3. – С. 122-148.

ПРИЛОЖЕНИЕ А

Акты о внедрении результатов диссертационной работы

“У Т В Е Р Ж Д А Ю”



Заместитель главного инженера,
филиала ПАО «Электросетьсервис ЕНЭС»
СПБ «Электросетьремонт»

А.В. ШЕЛЕПОВ

АКТ

внедрения результатов диссертационной работы
Демина Виктора Андреевича

г. Новосибирск

20 июня 2016 г.

На предприятии создана комиссия в следующем составе:

1. Председатель: главный специалист, д.т.н., профессор Овсянников А.Г.
2. Член комиссии: начальник службы диагностики маслонаполненного оборудования, к.т.н. Живодерников С.В.
3. Член комиссии: ведущий инженер, к.т.н. Шиллер О.Ю.

Рассмотрены результаты внедрения результатов диссертационного исследования Демина В.А., представленного на соискание учёной степени кандидата технических наук.

Комиссия пришла к следующему выводу: использование предложенной в диссертации методики построения непараметрических оценок условных функций распределения и разработанного программного обеспечения позволило выявить закономерности возникновения частичных разрядов (ЧР) в жидких диэлектриках. Найдены условия возникновения ЧР в пузырьках газа, построено распределение значений напряжения возникновения ЧР в зависимости от действующего напряжения, что позволяет точнее рассчитать коэффициент газообразования и оценить надежность маслонаполненного электрооборудования высокого напряжения, использующих трансформаторные масла в качестве жидких диэлектриков.

Председатель комиссии

Член комиссии

Член комиссии

А.Г. Овсянников

С.В. Живодерников

О.Ю. Шиллер

“УТВЕРЖДАЮ”

Руководитель сервиса коммерции ООО «ДГ Софт»,

Цивинский С.В.

2016 г.



АКТ

о внедрении результатов диссертационной работы Демина Виктора Андреевича

Настоящим актом подтверждается, что результаты диссертационного исследования Демина В.А. внедрены в практику деятельности ООО «ДГ-Софт». В частности, адаптивный алгоритм выбора оптимального параметра размытости для непараметрической оценки функции надёжности и программная система статистического анализа данных типа времени жизни «LiTiS 1.2» использовались для решения задачи анализа длительности контрактов на размещение рекламы в зависимости от количества купленных позиций. В результате проведенного анализа скорректирован алгоритм работы менеджеров по продажам.

Руководитель группы разработки ПО

Д.Н. Ширшаков

Руководитель группы разработки ПО

А.С. Романчук

УТВЕРЖДАЮ

Проректор по учебной работе НГТУ
д.т.н., профессор

 С.В. Брованов
 “ 19 ” 2016 г.

АКТ

**о внедрении результатов диссертационной работы В.А. Демина
в учебный процесс факультета прикладной математики и информатики**

Результаты диссертационной работы Демина Виктора Андреевича, в частности, разработанные алгоритмы и программная система статистического анализа данных типа времени жизни «LiTiS 1.2», внедрены в учебный процесс на факультете прикладной математики и информатики (ФПМИ) ФГБОУ ВО «Новосибирский государственный технический университет» и используются при изучении дисциплины «Компьютерные технологии анализа данных и исследования статистических закономерностей» по направлению 01.04.02 – Прикладная математика и информатика, в рамках магистерской программы «Математическое моделирование детерминированных и стохастических процессов». Освоение магистрантами соответствующих разделов дисциплины способствует приобретению необходимых знаний и умений для применения на практике современных алгоритмов построения непараметрических моделей надёжности и выживаемости.

 Декан ФПМИ,
д.т.н., доцент



В.С. Тимофеев

 Заведующий кафедрой ТПИ,
д.т.н., доцент



В.М. Чубич