

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт космических и информационных технологий  
Кафедра систем искусственного интеллекта

УТВЕРЖДАЮ  
Заведующий кафедрой  
Г.М. Цибульский  
\_\_\_\_\_ 2016 г.



**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

Построение гарантированных оценок показателей функционирования  
сложных технических систем в условиях ограниченного объема информации

09.04.02 – Информационные системы и технологии  
09.04.02.01 – Информационно-управляющие системы

Научный  
руководитель

  
подпись, дата

доцент, канд. тех. наук


О.А. Попова

Выпускник

  
подпись, дата

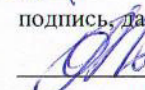
Д.Е. Пооль

Рецензент

  
подпись, дата

ст. науч. сотр., канд. физ.-мат. наук А.Н. Рогалев

Нормоконтролер

  
подпись, дата

М. А. Аникьева

Красноярск 2016

## СОДЕРЖАНИЕ

Введение.....	4
1 Анализ проблемы исследования.....	7
1.1 Проблема обеспечения надежности и гарантированного управления эксплуатацией технических объектов ответственного назначения.....	7
1.2 Мониторинг и прогнозирование состояний ТСОИ .....	12
1.3 Анализ методов к построению надежных оценок показателей функционирования сложных технических систем .....	21
1.3.1 Непараметрические оценки характеристик надежности .....	22
1.3.2 Интервальное оценивание характеристик надежности .....	23
1.4 Выводы по главе 1 .....	27
2 Численный вероятностный анализ .....	28
2.1 Элементы численного вероятностного анализа .....	28
2.2 Практическая реализация .....	31
2.2.1 Постановка задачи .....	31
2.2.2 Алгоритм решения задачи .....	32
2.3 Выводы по главе 2 .....	36
3 Методы сглаживания эмпирических функций распределения .....	38
3.1 Теория сплайнов .....	38
3.2 Порядковые статистики .....	51
3.2.1 Определениеслучайнойвыборки .....	51
3.2.2 Законраспределенияпорядковыхстатистик.....	52
3.2.3 Эмпирическая функцияраспределения.....	54
3.2.4 Группирование выборочных данных,гистограмма .....	57
3.2.5 Понятие доверительного интервала.....	59
3.2.6 Основные этапы построения доверительных интервалов.....	60
3.3 Надежные оценки эмпирических функций распределения .....	61
3.4 Выводы по главе 3 .....	64
4 Разработка модуля построения гарантированных оценок показателей надежности ТСОИ.....	66
4.1 Базовая методика оценки надёжности .....	67

4.2	Оценки интенсивности отказов.....	73
4.3	Модельный пример .....	74
4.4	Информационно – аналитический модуль построения гарантированных оценок показателей надежности.....	76
	Заключение .....	78
	Список использованных источников .....	80
	Приложение А .....	83
	Приложение Б.....	91

## ВВЕДЕНИЕ

Диссертация посвящена актуальной теме — надежности сложных технических систем. В работе рассмотрен анализ оценок показателей функционирования технических систем на основе информационно-аналитического подхода. В настоящее время особую значимость приобретают вопросы информационной поддержки надежности функционирования оборудования ответственного назначения в условиях ограниченного объема информации. Важную роль играют подходы к повышению достоверности оценок показателей надежности и методы вычислений оценки вероятности безотказной работы.

Отказы сложных технических систем (ТС) ответственного назначения могут стать источником техногенных чрезвычайных ситуаций. Экономические, экологические и социальные последствия подобных происшествий обуславливают необходимость совершенствования научных основ анализа техногенных рисков и обеспечения техногенной безопасности.

Как правило, объекты ответственного назначения относятся к числу уникальных систем. Под уникальными ТС понимают системы, изготавливаемые в небольшом числе экземпляров, эксплуатируемые в отличающихся условиях и реализующие экстремальные технологические операции [15]. Отказы уникальных ТС во многом обусловлены процессами деградации их свойств, приводящими к выработке ресурса и потере работоспособности.

Предотвращение аварий ТС связано с задачей понижения возможности отказов и снижения техногенных рисков, которая, в свою очередь, в значительной степени зависит от прогнозирования динамики изменения технического состояния исследуемых технических объектов.

В условиях уникальных ТС число испытаний ограничено, а имеющаяся информация не достаточна для получения гарантированных оценок показателей надежности. Оценка надежности многих видов

оборудования осуществляется индивидуально для каждого экземпляра по результатам периодических обследований.

Для таких систем необходимо применять методы, которые позволяют получать гарантированные оценки в условиях недостаточности информации и малых выборок.

**Целью магистерской диссертации** является повышение эффективности анализа надежности сложных технических систем ответственного назначения на основе построения гарантированных оценок показателей функционирования в условиях неопределенности и ограниченного объема информации.

**Объектом исследования** является надежность ТСОИ на основе информационно – аналитического подхода.

**Предметом исследования** являются численные методы представления, обработки, моделирования, анализа данных в условиях неопределенности и ограниченного объема информации.

Основные задачи диссертационного исследования:

- 1) на основе изучения научных публикаций провести проблемный анализ по теме исследования;
- 2) изучить теоретические основы численного вероятностного анализа и применить его методы для построения численных оценок показателей надежности;
- 3) разработать алгоритм построения эмпирической функции распределения с использованием сглаживающего сплайна;
- 4) на тестовых примерах провести анализ численных оценок показателей надежности в условиях ограниченного объема информации;
- 5) разработать модуль для построения гарантированных оценок, показателей надежности в условиях малых выборок.

Диссертация состоит из введения, четырех разделов и заключения. В первой главе рассматривается проблема исследования, исследуется ее актуальность, дается анализ публикаций и определяется круг решаемых

задач. Во втором разделе рассматриваются теоретические основы численного вероятностного анализа, методы которого предлагаются для построения численных процедур обработки, представления и анализа данных в условиях ограниченного объема информации. Рассматриваются численные примеры. В третьем разделе изучаются методы сглаживания эмпирических функций распределения. Четвертый раздел посвящен вопросам надежности, изучаются базовая методика оценки показателей безотказной работы сложного технического оборудования, и приводится подход для построения гарантированных оценок показателей надежности.

## **1 Анализ проблемы исследования**

### **1.1 Проблема обеспечения надежности и гарантированного управления эксплуатацией технических объектов ответственного назначения**

Обеспечение надежности сложных технических систем представляет собой один из элементов общей системы оценки качества технических систем ответственного назначения (ТСОН). Анализ требований к показателям надежности современных отечественных ТСОН показывает, что требования к таким показателям, как долговечность (срок активного существования или ресурс), параметр безотказности (вероятность безотказной работы (ВБР)), показатель интенсивности отказов и другие непрерывно ужесточаются. Процесс повышения качества таких систем предполагает наличие определенной методологии оценки показателей надежности состояний системы на всех этапах ее жизненного цикла, который представляет собой набор последовательных взаимосвязанных процессов изменения состояния на этапах ее проектирования и эксплуатации.

Наличие неопределенности при решении многих практических задач во многом изменяет способы обработки, численного моделирования и анализа данных. Для улучшения качества результатов проводимых исследований уже на первых этапах предобработки данных меняется принцип представления исходных данных, требуются новые способы и методы их преобразования. Причины, по которым в исходных данных возникает неопределенность, могут быть различными. При этом, как правило, неопределенность делят на два вида: элиторную и эпистимическую. Первая подразумевает под собой случайность и позволяет неоднократно использовать один и тот же набор выборки для оценки плотностей вероятности случайных величин. В отличие от первого типа, второй тип таким качеством не обладает. К последнему типу можно отнести и интервальную неопределенность. Для интервальной неопределенности, представленной интервальными числами, был разработан интервальный анализ еще в середине прошлого века. На данный момент

интервальный анализ занимает достаточно значимую область в вычислительной математике. Однако, не смотря на достаточно большое количество исследований, интервальные оценки могут давать слишком размытые интервалы. Об этом, в том числе упоминается в работах [7, 9]. В этих работах особо подчеркивается, что в условиях различных видов неопределенности актуализируется проблема повышения надежности вычислительных алгоритмов и методов численного моделирования и анализа.

В современных условиях функционирования сложных технических комплексов, таких как электростанции, крупные предприятия, транспортные средства, трубопроводы и т.п. оборудование зачастую работает за пределами первоначально предназначенного для них срока службы. Другими словами, они уже выработали свой ресурс. Однако, несмотря на это, мнение специалистов по вопросам эксплуатации, надежности и безопасности потенциально опасных и уникальных объектов остается незамеченным. Частично, причиной игнорирования экспертного мнения, является то, что техническими вопросами на всех уровнях управления, в том числе в сфере эксплуатации и безопасности, занимаются менеджеры, которые в большинстве своем не имеют должного уровня квалификации и профессионально опыта.

Задача обеспечения адекватного функционирования ТСОИ в последнее время трансформируется из инженерно-технической в научно-техническую, поскольку в процессе их работы привносится гораздо больший, чем ранее, фактор неопределенности, что объясняется постепенным старением оборудования и отсутствием достаточных средств для его техобслуживания либо замены.

Наиболее значимым процессом использования в работе систем является техническое обслуживание, под которым понимают все действия, предпринимаемые с целью сохранения или восстановления работоспособности эксплуатируемых систем.



Абрамов, О. В. в одной из своих публикаций, посвященных авариям сложных технических систем, делит техобслуживание на три вида:

- техническое обслуживание, определяемое отказом
- техобслуживание через определенные интервалы времени (обслуживание по наработке)
- техническое обслуживание по состоянию.

Техобслуживание, определяемое отказом, происходит без существенных затрат на профилактику и контроль. Соответственно, в этом случае система эксплуатируется до отказа. Такую стратегию эксплуатации можно считать приемлемой только в тех редких случаях, когда соответствующее оборудование имеет 100%-й резерв или является не первостепенным для использования в целевом назначении. При техобслуживании, определяемом отказом, оборудование выходит из-под контроля обслуживающих его рабочих. Поэтому, каждый последующий выход из строя оборудования происходит неожиданно. Естественно, это сказывается на планировании и зачастую делает его просто невозможным.

Учитывая все перечисленные недостатки, всегда считалось, что этот вид техобслуживания приемлем лишь для тех объектов, которые не выполняют ответственные функции, и внезапный выход из строя которых не может глобально сказаться на функционировании других объектов. Например, к таким устройствам можно отнести различную бытовую технику. К сожалению, единственное преимущество этого подхода, заключающегося в простоте и дешевизне, привели в последнее время к его повсеместному использованию, в том числе и в случае объектов ТСОИ, таких, например, как судовое оборудование, системы электроснабжения, трубопроводы и др.

Регулярное техобслуживание, проводимое через фиксированные промежутки времени (техобслуживание по наработке), является наиболее эффективным и используется чаще всего в современной практике. Такая стратегия подразумевает некий план, в соответствии с которым объект по истечении заданного срока, установленного на основе предыдущего опыта

функционирования, выводится из эксплуатации, проходит полное обслуживание, разбирается, а выработавшие свой срок эксплуатации узлы заменяются.

Преимущество такой подхода определяется в первую очередь простотой планирования мероприятий по техобслуживанию, а также тем, что в процессе функционирования системы нет необходимости в постоянном замере каких-либо ее параметров. Эта стратегия имеет достаточно хорошо проработанную теоретическую базу, но поскольку режимы обслуживания определяются для однотипных технических систем исключительно по априорной информации (без учета их реального состояния), то эта стратегия несет за собой достаточно ощутимые временные и экономические потери.

Подход, ориентированный на время эксплуатации, зачастую становится весьма дорогим и далеко не самым оптимальным. Обосновать это можно тем, что мероприятия по техобслуживанию (например: остановка, разборка, замена, сборка, проверка) проводятся, как правило, либо преждевременно, либо – слишком поздно.

В последнее время, такая стратегия, как стратегия планирования эксплуатации технических объектов в зависимости от их действительного состояния привлекает в последнее время все большее внимание специалистов. Такой подход называют еще индивидуальным, поскольку он ориентируется исключительно на реальное состояние и учитывает особенности именно данной конкретной системы, а не опыт эксплуатации аналогичных систем и статистические данные, для корректного использования которых необходимы статистическая однородность и существенный объем используемой информации.

Основные преимущества от перехода к индивидуальной стратегии назначения интервалов техобслуживания обуславливается главным образом следующими факторами:

а) возможностью в максимальной степени использовать ресурс использования каждого отдельного объекта, что достигается уменьшением числа преждевременных вмешательств в его работу;

б) возможностью предотвращения отказов, что обеспечивается своевременным проведением профилактических (предупредительных) мероприятий.

Следует понимать, что индивидуальное планирование эксплуатации возможно только при условии получения исчерпывающей информации о фактическом состоянии каждого объекта, т.е. реализация индивидуальной стратегии эксплуатации требует постоянного контроля и сбора информации о состоянии объекта. Предполагается, что реальное техническое состояние объекта можно оценить по результатам контроля (измерения) его параметров, а прогнозирование их изменений позволяет эксплуатировать объект до появления признаков опасного снижения надежности, исключив при этом преждевременные замены узлов и агрегатов, а также выполнение других трудоемких и экономически невыгодных работ, имеющих кроме того сомнительную полезность для надежности функционирования.

Основные сложности при решении задачи прогнозирования для применения стратегии эксплуатации по состоянию связаны с тем, что прогноз приходится составлять для каждого объекта индивидуально, при достаточно малых объемах исходной информации (по небольшому набору результатов контроля параметров функционирования системы) и в обязательном наличии различного рода помех (ошибок контроля), статистические свойства которых доподлинно неизвестны. В таких условиях традиционные методы математической статистики и теории вероятности теряют свои привлекательные свойства, а их использование для прогнозирования приводит к большим погрешностям и низкой достоверности прогноза [10].

## 1.2 Мониторинг и прогнозирование состояний ТСОИ

Прогнозирование состояния систем и надежности объектов занимает важное место при стратегии индивидуального планирования эксплуатации. Навык предсказания предполагаемого момента отказа особенно актуально для объектов ответственного назначения, выход из строя которых связан с большими экономическими потерями или катастрофическими последствиями. Предупреждение отказов для объектов такого рода становится первостепенной задачей. Однако особые условия эксплуатации большинства из них не дает возможность решить эту задачу даже при постоянном контроле их параметров, так как в течение определенных интервалов времени функционирования техническое обслуживание невозможно в принципе (самолетов – во время полета, надводных и подводных судов – во время рейса, оборудования связи – в течение сеанса связи или непосредственной передачи данных и т.д.).

Для большинства технических систем постоянный контроль осуществить не имеется возможным, а при периодическом контроле каждая оценка их действительного состояния зачастую связана с ощутимыми материальными потерями. В подобных случаях прогнозирование дает возможность решать задачу назначения наиболее подходящих моментов контроля, в промежутках между которыми не произойдет выхода из строя с гарантированной вероятностью.

Главные препятствия при решении задачи прогнозирования для использования стратегии эксплуатации по состоянию связаны с тем, что прогноз приходится осуществлять для каждого объекта индивидуально, при малых объемах исходной информации (по малому объему результатов контроля) и в присутствии погрешностей (ошибок контроля), статистические свойства которых априорно не известны. В подобных условиях традиционные методы математической статистики и теории вероятности

теряют свои лучшие качества, а их использование для прогнозирования приводит к ощутимым погрешностям и низкой достоверности прогноза.

Тем не менее, имеются некоторые методы и подходы к решению задачи индивидуального прогнозирования и планирования эксплуатации при дефиците и неточности исходных данных, позволяющие получать в таких условиях весьма неплохие результаты. К числу подобных способов относится метод индивидуального гарантированного прогноза (МГП). Его идея базируется на том, что из множества возможных реализаций случайного процесса ухудшения свойств параметров (состояния) исследуемого технического объекта (системы), которые не противоречат результатам контроля, выбираются наихудшие. Под наихудшими понимаются такие, которые располагаются за пределами области наблюдения (контроля). Реализации такого рода можно называть экстремальными [12].

Исходя из исследования в работе [12], если в качестве модели случайного процесса изменения параметров состояния исследуемого технического объекта может быть принята структура, представляющая собой полином Чебышева со случайными коэффициентами, то наихудшими реализациями в этом случае представят собой экстремальные полиномы Карлина [25]. Аппарат экстремальных полиномов Карлина дает возможность построить так называемый конус прогноза, образованный двумя наихудшими реализациями нижней  $x(t)_-$  и верхней  $x(t)_+$ , в границах которого на интервале прогноза будет гарантировано находиться истинная реализация случайного процесса изменения технического состояния исследуемой системы. В отличие от традиционных вероятностно-статистических методов прогноза МГП позволяет получать решение при отсутствии сведений о вероятностных характеристиках ошибок наблюдений и другого рода помех.

Таким образом, МГП дает возможность определить гарантированную область, в пределах которой с высокой долей вероятности будут находиться параметры состояния системы в любой заданный момент времени. Он

обладает необходимыми свойствами несмещенности, однозначности и оптимальности. Итоговые значения прогноза позволяют задавать целесообразные интервалы контроля состояния параметров системы, а также профилактических и ремонтных работ.

### **1.2.1 Прогнозирование с использованием теории нейронных сетей**

В существующих способах прогнозирования (в том числе и в методе МГП) обычно априорно известна структура случайного процесса изменения параметров состояния системы, либо делаются предположения о свойствах класса, которому принадлежит реальный процесс. Таким образом, выдвигается гипотеза, на основе которой принимается та или иная модель процесса изменения параметра, относительно которой и решается задача прогноза. Однако, зачастую априорной информации недостаточно для принятия той или иной гипотезы о подходящей модели процесса, а выборка апостериорных данных (результатов контроля) содержит в себе слишком мало информации для каких-то статистических выводов.

В таких условиях получение достаточно точных результатов прогноза дают активно развиваемые в последние годы методы теории искусственных нейронных сетей [3, 7].

Основными преимуществами искусственных нейронных сетей являются свойства обучения и обобщения. Свойство обучения означает, что после многократного ввода в сеть обучающих примеров, она способна давать правильные ответы на любые другие входные данные из обучающего множества. Кроме того, после завершения процесса обучения и тестирования сеть способна давать правильные ответы также и на любые другие входные данные, не являющиеся данным обучающего множества. Эта способность нейронных сетей называется свойством обобщения.

Пусть техническое состояние рассматриваемого нами объекта в любой заданный момент времени  $t$  определяется некоторым векторным параметром  $x(t)$ . Фактический процесс изменения параметра представляет собой

неизвестную функцию времени и задан только своими значениями, полученными в ходе наблюдения за процессом. Соответственно, возникает задача оценки значения процесса  $x(t)$  при  $t > t_0$ .

Пусть известны значения параметра  $x$  в промежуточные моменты времени:

$$\left( \begin{array}{l} x_1 = x(t_0), \\ x_2 = x(t_0 - \delta_1), \\ x_3 = x(t_0 - \delta_1 - \delta_2), \\ \dots \\ x_n = x(t_0 - \delta_1 - \delta_2 - \dots - \delta_n) \end{array} \right) \delta_i > 0. \quad (1.1)$$

Решением будем считать значения  $x(t_0 + \delta_0)$ ,  $x(t_0 + 2\delta_0)$  и т.д.

В качестве нейро-сетевой модели для прогноза выберем многослойную сеть, состоящую из неограниченного количества слоев нейронов. Нейроны каждого слоя соединяются с нейронами предыдущего и последующего «каждый с каждым», образуя плотную сеть. Первый слой называется входным, внутренние – скрытыми, а последний – выходным. Количество нейронов в каждом слое может быть любым. Как правило, во всех скрытых слоях содержится одинаковое количество нейронов. На данный момент доказана достаточность трехслойной нейронной сети для реализации любой функции [3]. Функционирование подобной сети с тремя слоями (одним входным, одним скрытым и одним выходным) можно описать следующим образом:

$$f(x) = F\left(\sum_i w_i^{(2)} \cdot F\left(\sum_j w_{ij}^{(1)} x_j\right)\right), \quad (1.2)$$

где  $F$  – функция активации нейронов;

$W_1 = \{w_{ij}^{(1)}\}$  ,  $W_2 = \{w_i^{(2)}\}$  – весовые коэффициенты между входным и скрытым слоем и между скрытым и выходным слоем соответственно.

Общий алгоритм решения задач с помощью многослойной нейронной сети.

1. Формализация входных и выходных параметров сети.
2. Выбор функции активации  $F$ .
3. Выбор числа слоев и нейронов в слое.
4. Присвоение начальных значений весовым коэффициентам и параметрам функции активации.
5. Обучение.
6. Подача на вход сети условия задачи (входного вектора  $x$ ), расчет результирующего выходного вектора  $y$ .

Формализацию входных и выходных параметров сети для задачи прогнозирования выполним так. Разложим результаты измерений технического состояния исследуемого объекта (1) на обучающие последовательности, состоящие из входного вектора и соответствующего ему выходного значения:

$$\begin{aligned}
 & (\{x_1, x_2, \dots, x_k\}, x_{k+1}); \\
 & (\{x_2, x_3, \dots, x_{k+1}\}, x_{k+2}); \\
 & \dots \\
 & (\{x_{n-k}, x_{n-k+1}, \dots, x_{n-1}\}, x_n)
 \end{aligned} \tag{1.3}$$

Выбор функции активации, как правило, зависит от выбранного метода обучения. Под обучением имеется в виду выбор наиболее подходящих параметров сети: весовых коэффициентов и функции активации. Универсального алгоритма обучения, подходящего для всех архитектур нейронных сетей, не существует. Способы обучения отличаются друг от друга методом настройки синаптических весов нейронов и способом связи обучаемой нейросети с окружающими объектами [7]. Выберем для обучения алгоритм обратного распространения. Для этого алгоритма важна непрерывность первой производной функции активации, поэтому самыми приемлемыми являются логистическая функция (1.4) и гиперболический тангенс (1.5):



$$F(S) = \frac{1}{1 + e^{-\alpha S}} \quad (1.4)$$

$$F(S) = \frac{e^{\frac{S}{\alpha}} - e^{-\frac{S}{\alpha}}}{e^{\frac{S}{\alpha}} + e^{-\frac{S}{\alpha}}} \quad (1.5)$$

где  $\alpha$  – крутизна сигмоиды.

Главный способ выбора слоев и нейронов – это непосредственный перебор различного количества слоев и нейронов в них и определение наилучшей модели структуры опытным путем [7]. Размерность входного вектора и количество нейронов в скрытых слоях для нашего случая определялось размерностью данных измерений технического состояния объекта так, чтобы количество связей между нейронами в сети было меньше числа обучающих пар.

Если выбрано множество обучающих пар и функция ошибки, то обучение сети превращается в задачу многомерной оптимизации, имеющую большую размерность. Поэтому, для обучения сети использовался стандартный алгоритм обратного распространения, когда сигналы ошибки рассматриваются в направлении, обратном прямому распространению сигналов в обычном режиме работы.

Минимизируемой целевой функцией ошибки нейронной сети, согласно этому алгоритму и применяемому в нем методу наименьших квадратов, являлась величина  $E$ :

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t), \quad (1.6)$$

$$\Delta w_{ij}^{(n)}(t) = -\eta \frac{\partial E}{\partial w_{ij}^{(n)}} + \alpha [\Delta w_{ij}^{(n)}(t-1)], \quad (1.7)$$

где  $n$  – номер слоя;  $\eta$  – коэффициент скорости обучения,  $0 < \eta < 1$ ,  $\alpha$  коэффициент импульса, который принимаем  $\alpha = 0,9$ .

Можно доказать, что

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot x_i^{(n)}, \quad (1.8)$$

где вспомогательная переменная определяется рекурсивно

$$\delta_j^{(n)} = \left[ \sum_k \delta_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{dS_j} \quad (1.9)$$

Для выходного слоя справедливо соотношение:

$$\delta_j^{(N)} = (y_i^{(N)} - d_i) \frac{dy_i}{dS_i} \quad (1.10)$$

Здесь 
$$S_j = \sum_{i=1}^m x_i w_{ij}, y_i = F(S_i)$$

$$\frac{dy_i}{dS_i} = \alpha \cdot y_i(1 - y_i) \quad (1.11)$$

если функция активации  $F$  – логистическая из (3);

$$\frac{dy_i}{dS_i} = \frac{1}{\alpha^2} \cdot (1 - y_i^2) \quad (1.12)$$

если  $F$  – гиперболический тангенс из (1.5).

Рассмотренный алгоритм действует итерациями, которые принято называть эпохами. В каждой эпохе на вход сети поочередно подаются все обучающие наблюдения, выходные значения сети сравниваются с целевыми значениями и вычисляется ошибка. Значения ошибки, а также градиента поверхности ошибок используются для корректировки весов, после чего все действия повторяются. Начальная конфигурация сети (значения весовых коэффициентов) выбирается случайным образом. Процесс обучения прекращается либо когда пройдено заданное количество эпох, либо когда погрешность достигнет определенного уровня малости, либо когда ошибка перестанет изменяться. Для того чтобы сформированная сеть обладала свойством не только запоминать данные, но и обобщать их, весь временной ряд был разделен на обучающее (90-80%) и тестовое множество (10-20%). В качестве точки остановки обучения использовался такой критерий:

распознано в обучающей выборке – 100% и распознано в тестовой выборке 96-100%.

Алгоритм обратного распространения, использующий градиентный метод, имеет ряд недостатков, главным из которых является то, что он не гарантирует нахождения глобального экстремума, однако по сравнению с методами глобальной оптимизации (сканирования или случайного поиска), которые решают эту проблему, он имеет довольно высокую скорость сходимости, что важно при обучении нейронной сети.

Для расчета результирующего прогнозного значения  $x_{n+1} = x(t_0 + \delta)$  на 13 вход сети подается вектор  $(x_{n-k+1}, x_{n-k+2}, \dots, x_n)$ . Затем компоненты входного вектора сдвигаются так, чтобы выходное значение стало одним из компонентов входного вектора (3). Таким образом, для того чтобы получить  $x_{n+2} = x(t_0 + 2\delta)$  на вход сети надо подать вектор  $(x_{n-k+2}, x_{n-k+3}, \dots, x_{n+1})$ .

### **1.2.2 Индивидуальное планирование эксплуатации объектов контроля**

Остановимся на схеме принятия решений в задаче управления эксплуатацией уникальных технических систем. Информацией для принятия решений являются результаты контроля и прогнозирования состояния исследуемой системы. Результаты контроля позволяют оценить техническое состояние и запас работоспособности в момент контроля, а результаты прогноза – динамику изменения состояния (запаса работоспособности) в будущие (упрежденные) моменты времени. Реализация схемы управления эксплуатацией заключается в построении правила решения, связывающего конкретное решение с данными контроля и результатами прогноза.

Простейшей моделью объекта контроля является однопараметрическая модель, в которой техническое состояние зависит от одного параметра. На параметр задан гарантийный допуск, характеризуемый границами  $[A, B]$ . Объект считается работоспособным, если параметр состояния находится внутри поля допуска, и неработоспособным – в противном случае. За

показатель качества однопараметрического объекта можно принять характеристическую функцию гарантийного поля допуска

$$W = L(x) = \begin{cases} 1, x[A, B] \\ 0, x[A, B] \end{cases} \quad (1.13)$$

Рассмотрим алгоритм управления эксплуатацией однопараметрического объекта с использованием результатов прогноза его технического состояния. Пусть работоспособность объекта определяется состоянием контролируемого дискретно выходного параметра  $x(t)$ . Условие работоспособности задано в виде

$$A \leq x(t) \leq B, \quad (1.14)$$

где  $A, B$  – соответственно нижняя и верхняя границы допустимых изменений параметра. Задача состоит в назначении таких моментов контроля и коррекций параметра, при которых гарантируется его нахождение в области допустимых значений (выполнение условий работоспособности) в течение времени эксплуатации  $[0, T]$ . При этом необходимо стремиться к тому, чтобы число контрольных замеров (и коррекций) было по возможности минимальным. Пусть по результатам  $p$  контрольных измерений построены экстремальные реализации  $x(t)_-$  и  $x(t)_+$ ,  $t > t_p$ . Пересечение экстремальных реализаций с границами области допустимых изменений параметра  $A$  и  $B$  определит моменты времени  $\tau_A$  и  $\tau_B$ , минимальный из которых целесообразно принять за момент очередного  $(p+1)$ -го контроля:

$$t_{(p+1)} = \min\{\tau_A, \tau_B\} \quad (1.15)$$

Моменты времени  $\tau_A, \tau_B$  находятся решением уравнений:

$$x(t)_- = A \text{ и } x(t)_+ = B \quad (1.16)$$

Очевидно, что в течение времени контролируемый параметр будет гарантированно находиться в области допустимых значений, поэтому до момента  $t(p+1)$  производить измерение или коррекцию параметра нет

необходимости. В момент  $t(p+1)$  выполняется очередное измерение параметра, результат которого используется для расчета параметров новых экстремальных реализаций. Определяется очередной промежуток времени, в течение которого параметр не выйдет за допустимые пределы. Если этот промежуток (назовем его интервалом гарантированной эксплуатации) окажется меньше некоторого наперед заданного минимально целесообразного времени эксплуатации, то в момент времени  $t(p+1)$  следует произвести профилактическую коррекцию параметра  $x(t)$ .

Многопараметрический объект, описываемый совокупностью независимых параметров, для которых определены гарантийные допуски, можно считать работоспособным, если все параметры находятся в пределах своих допусков, и неработоспособным, если хотя бы один параметр выходит за пределы своего поля допуска. За обобщенный показатель качества такого объекта можно принять величину, вычисляемую как произведение характеристических функций полей допусков всех параметров:

$$W = \prod_{i=1}^N l_i(U_i) \quad (1.17)$$

где  $N$  – общее количество параметров объекта, а функция  $l_i(U)$  определяется соотношением

$$l_i(U) = \begin{cases} 1, U_i \in [A_i, B_i] \\ 0, U_i \notin [A_i, B_i] \end{cases} \quad (1.18)$$

Схема принятия решения заключается в прогнозировании изменений каждого из параметров, сравнении результатов прогноза с границами своих полей допусков и назначении технического обслуживания по моменту первого выхода одного из параметров за пределы поля допуска [11].

### **1.3 Анализ методов к построению надежных оценок показателей функционирования сложных технических систем**

Оценку показателей надежности объектов осуществляют на основании информации, полученной в процессе специальным образом организованных испытаний. Наиболее объективной информацией для определения характеристик надежности элементов является эксплуатационная информация, потому что она отражает специфику функционирования объектов, комплекс воздействующих факторов и другие особенности эксплуатации.

Применяемые традиционно параметрические методы анализа статистической информации в последнее время подвергаются критике специалистов по математической статистике. Дело в том, что использование параметрических методов требует предположений о виде закона распределения наблюдаемых случайных величин. Как правило, нельзя указать каких-либо веских причин, по которым конкретное распределение результатов наблюдений должно входить в то или иное параметрическое семейство. В настоящее время для решения задач анализа статистической информации развиваются непараметрические методы, в частности, методы ядерного оценивания.

### 1.3.1 Непараметрические оценки характеристик надежности

Ядерные оценки впервые были введены в работах Парзена [6] и Розенблатта [7]. Приведем формулу для построения ядерной оценки плотности распределения непрерывной случайной величины. Пусть в результате наблюдения за объектом исследования получена выборка наработок до отказа  $\xi_1, \xi_2, \dots, \xi_n$ .

Ядерную оценку плотности распределения (ПР) определяют следующим образом:

$$f_n(t) = \frac{1}{n\sigma} \sum_{i=1}^n V\left(\frac{t - \xi_i}{\sigma}\right) = \frac{1}{n\sigma\sqrt{2\pi}} \sum_{i=1}^n \exp\left[-\left(\frac{t - \xi_i}{\sqrt{2}\sigma}\right)^2\right] \quad (1.19)$$

где  $V(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$  гауссовское ядро;  $n$  – объем выборки;  $\sigma$  –

параметр локальности (ширина окна ядерной оценки, управляемый параметр). В работах [1, 2, 3] получены формулы и проведены исследования ядерной оценки плотности распределения для различных схем формирования исходной информации.

Обратим внимание на одну особенность решения задач по определению характеристик надежности. Наблюдаемой случайной величиной в рассматриваемой задаче является время (наработка до отказа). Область определения данного параметра  $[0, \infty)$ . Следовательно, априори известно, что обрабатываемые случайные величины относятся к классу неотрицательных величин. Для положительно определенных случайных величин в качестве ядра предложено использовать гауссовское ядро с зеркальным отражением:

$$\hat{f}(t, \sigma) = \frac{1}{n\sigma} \sum_{i=1}^n \left[ V\left(\frac{t - \xi_i}{\sigma}\right) + V\left(\frac{t + \xi_i}{\sigma}\right) \right] \quad (1.20)$$

Таким образом, изложена процедура получения плотности распределения наблюдаемой случайной величины наработки до отказа. Определив данную характеристику, можно переходить к оценке других показателей надежности. Отметим основные из них. Вероятность безотказной работы (ВБР) – это вероятность того, что в пределах заданной наработки  $t$  отказ объекта не возникнет, то есть случайное время наработки до отказа  $\xi$  окажется не меньше  $t$ :

$$P(t) = P(\xi \geq t) = \bar{Q}(t) \quad (1.21)$$

### 1.3.2 Интервальное оценивание характеристик надежности

#### Доверительные границы для плотности распределения

Рассмотрим вначале построение доверительных границ для ядерной оценки плотности распределения, полученной по данным, содержащим

только полные наработки. Пусть у нас есть выборка  $\xi_1, \xi_2, \dots, \xi_n$ , представляющая собой результаты наблюдений за некоторым случайным параметром. Запишем выражение для ядерной оценки плотности распределения этого параметра

$$f_{\xi}(t) = \frac{1}{n\sigma} \sum_{i=1}^n V\left(\frac{t-\xi_i}{\sigma}\right) \quad (1.22)$$

$$\text{Обозначим } \eta_i = \frac{1}{\sigma} V\left(\frac{t-\xi_i}{\sigma}\right).$$

Случайные величины  $\eta_i$  являются независимыми и одинаково распределёнными (н.о.р.с.в.). Тогда, согласно центральной предельной теореме (ЦПТ), при больших  $n$  получим

$$f_{\xi}(t) = \frac{1}{n} \sum_{i=1}^n \eta_i \approx N\left(M_{\eta}, \sqrt{\frac{D_{\eta}}{n}}\right) \quad (1.23)$$

где  $M_{\eta}$ ,  $D_{\eta}$  – математическое ожидание и дисперсия случайной величины  $\eta$ . При небольших значениях числа  $n$  для построения доверительных границ можно воспользоваться распределением Стьюдента с  $n$  степенями свободы:

$$\frac{M_{\eta}(t) - f_{\xi}(t)}{\sqrt{D_{\eta}(t)/n}} \approx t_n \quad (1.24)$$

Тогда получим:

$$f_{\xi}(t) - t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{n}} \leq M_{\eta}(t) \leq f_{\xi}(t) + t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{n}} \quad (1.25)$$

Значения математического ожидания и дисперсии случайной величины  $\eta$  рассчитываются следующим образом:

$$M_{\eta}(t) = \frac{1}{\sigma} \int_{\Omega} V\left(\frac{t-\tau}{\sigma}\right) f_{\xi}(\tau) d\tau \quad (1.26)$$



$$D_{\eta}(t) = \frac{1}{\sigma^2} \int_{\Omega} V^2 \left( \frac{t-\tau}{\sigma} \right) f_{\xi}(\tau) d\tau - M_{\eta}^2(t) \quad (1.27)$$

### Доверительные интервалы для вероятности безотказной работы

Используя формулу (1.26), достаточно просто рассчитать доверительные интервалы для ВБР. Обозначим в (t)  $f_B(t)$  и в (t)  $f_H(t)$  верхнюю и нижнюю границы доверительного интервала для плотности распределения, которые определяем из выражений

$$f_H(t) = f_{\xi}(t) - t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{b}} \quad (1.28)$$

$$f_B(t) = f_{\xi}(t) + t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{b}} \quad (1.29)$$

Тогда доверительные границы для оценки ВБР можно записать в следующем виде:

$$1 - \int_0^t f_B(t) dx \leq P(t) \leq 1 - \int_0^t f_H(t) dx \quad (1.30)$$

### Доверительные границы для интенсивности отказов

Рассмотрим построение доверительных границ для ядерной оценки интенсивности отказа. Оценка интенсивности отказа будем рассчитывать в виде

$$\hat{\lambda}(t) = \frac{\hat{f}_{\xi,n}(t)}{1 - \hat{F}_{\xi,n}(t)} \quad (1.31)$$

где  $\hat{F}_{\xi,n}(t)$  и  $\hat{f}_{\xi,n}(t)$  ядерные оценки функции и плотности распределения соответственно. Как уже было показано, для оценки плотности распределения, построенной по полным наработкам, справедливо выражение (5). С другой стороны, последовательность случайных величин

$$1 - \hat{F}_{\xi,n}(t) \xrightarrow{n \rightarrow \infty} 1 - F_{уст}(t) \quad (1.32)$$

где  $F_{ист}(t)$  – истинное значение оцениваемой функции распределения случайной величины  $\xi$ . Согласно [7], получим при  $n \rightarrow \infty$

$$\frac{\hat{f}_{\xi,n}(t)}{1 - \hat{F}_{\xi,n}(t)} \approx N\left(\frac{M_{\eta}(t)}{1 - F(t)}, \sqrt{\frac{D_{\eta}(t)}{b} \frac{1}{1 - F(t)}}\right) \quad (1.33)$$

И для небольших  $n$ :

$$\lambda(t) - t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{b} \frac{1}{1 - F(t)}} \leq M_{\lambda}(t) = \left(\frac{M_{\eta}(t)}{1 - F(t)}\right) \leq \lambda(t) + t_{n,\alpha/2} \sqrt{\frac{D_{\eta}(t)}{b} \frac{1}{1 - F(t)}} \quad (1.34)$$

$M_{\eta}$  и  $D_{\eta}$  посчитаны при построении доверительных границ ядерной оценки плотности распределения (7).

### **Доверительные интервалы для среднего прямого и обратного остаточного времени**

Определение точности оценки математического ожидания прямого остаточного времени в сформулированных начальных условиях является достаточно сложной задачей. В случае непараметрического оценивания на данный момент не решено, каким образом можно получить оценку для дисперсии прямого остаточного времени.

Суть бутстреп-метода заключается в том, что теоретическое распределение генеральной совокупности заменяется выборочным. Далее происходит переход от одновыборочной схемы наблюдений к схеме многих выборок того же объема, которые извлекаются из первоначальной выборки и имеют распределение, совпадающее с выборочным. Эта процедура позволяет производить построение выборочного распределения оцениваемого параметра.

Рассмотрим применение бутстреп-метода для построения доверительных интервалов оценки прямого остаточного времени:

**Шаг 1.** Для конкретного оборудования имеем выборку наработок  $\xi_1$ ,  $\xi_2$ , ...,  $\xi_n$ . На основании этой выборки получаем оценку среднего прямого остаточного времени.

**Шаг 2.** Моделируем новую выборку, для чего из первоначальной выборки  $\xi_1$ ,  $\xi_2$ , ...,  $\xi_n$  извлекают случайную выборку объема  $n - 1$ . На этом шаге не получается перестановочного распределения, поскольку значения отбираются из первоначальной выборки с возвращением. Получаем оценку.

**Шаг 3.** Многократно повторяем шаг 2, получаем бутстреп-оценки.

**Шаг 4.** Для схемы непараметрического оценивания границы доверительных интервалов определяются следующим образом:

1. Выбираем уровень  $\alpha$ , соответствующий доверительной вероятности  $1 - 2\alpha$ ;
2. Для данного  $\alpha$  определяем границы интервалов, удовлетворяющих следующим соотношениям:

$$\alpha = \frac{d(V_i^* \leq V_H^*)}{r} \quad (1.35)$$

$$1 - \alpha = \frac{d(V_i^* \leq V_B^*)}{r} \quad (1.36)$$

где  $r$  – объем бутстреп-повторений;

$d(V_i^* \leq V_H^*)$  – количество оценок параметра  $V$  из числа бутстреп-повторений,  $\hat{V}_1^*, \hat{V}_2^*, \dots, \hat{V}_r^*$  принявших значения, меньшие чем  $V_B^*$  (или  $V_H^*$ ). В данном случае оценки  $V_B^*$  и  $V_H^*$  определяемые выражениями (8) и (9), будут характеризовать приближенный доверительный интервал, соответствующий доверительной вероятности  $1 - 2\alpha$ . Доверительную вероятность обычно принимают на уровне 0,9–0,95 [14].

## 1.4 Выводы по главе 1

В первой главе обоснована актуальность повышения качества построения гарантированных оценок показателей сложных технических систем в условиях малых выборок.

Для построения оценок надежности используются различные методы, однако учитывая малый объем данных, которые используются при анализе в следствии уникальности исследуемых систем, остро стоит проблема максимального извлечения знаний, для построения адекватных оценок.

Проанализированы существующие подходы и методы к построению надежных оценок показателей функционирования сложных технических систем. Рассмотрен непараметрический и интервальный методы, а также приведен пример по решению практических задач.

## **2 Численный вероятностный анализ**

### **2.1 Элементы численного вероятностного анализа**

Обозначим, через  $R$  – множество  $\{x\}$  случайных величин, заданных своими плотностями вероятности  $p_x$ , соответственно  $R^n$  – пространство случайных векторов.

Пусть имеется система непрерывных случайных величин  $(x_1, x_2, \dots, x_n)$  с плотностью распределения  $p(x_1, x_2, \dots, x_n)$ . Случайная величина  $y$  связана с  $(x_1, x_2, \dots, x_n)$  функциональной зависимостью

$$y = f(x_1, x_2, \dots, x_n). \quad (2.1)$$

Тогда функция распределения  $F_y$  для случайной величины  $y$

$$F_y(z) = P(y < z) = \int_{\Omega_z} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (2.2)$$

где  $\Omega_z = \{(x_1, x_2, \dots, x_n) | f(x_1, x_2, \dots, x_n) < z\}$ . В тех случаях, когда это возможно дифференцируя  $F_y$  можно получить плотность вероятности для  $y$

$$\frac{dF_y(z)}{dz} = f_y(z). \quad (2.3)$$

Плотность вероятности  $f_y$  будем называть *вероятностным расширением* функции  $f$ .

Полагая  $f(x_1, x_2) = x_1 * x_2, * \in \{+, -, \cdot, /\}$  можно получить аналитические формулы для определения плотности вероятности результатов арифметических действий над случайными величинами. Например, для нахождения плотности вероятности  $p_{x_1+x_2}$  суммы двух случайных величин  $x_1 + x_2$  используется соотношение [2,16]

$$p_{x_1+x_2}(x) = \int_{-\infty}^{\infty} p(x-v, v)dv = \int_{-\infty}^{\infty} p(v, x-v)dv. \quad (2.4)$$

для нахождения плотности вероятности  $p_{x_1/x_2}$  частного двух случайных величин  $x_1/x_2$

$$p_{x_1/x_2} = \int_0^{\infty} vp(xv, v)dv - \int_{-\infty}^0 vp(v, xv)dv. \quad (2.5)$$

Плотность вероятности  $p_{x_1x_2}$  произведения двух случайных величин  $x_1x_2$  определяется соотношением

$$p_{x_1x_2} = \int_0^{\infty} (1/v)p(x/v, v)dv - \int_{-\infty}^0 (1/v)p(v, x/v)dv. \quad (2.6)$$

### **Операции над плотностями вероятности случайных величин**

В качестве примера рассмотрим арифметические операции над гистограммными переменными. Реализация арифметических операций основана на работе с  $p(x, y)$  – совместной плотностью вероятности двух случайных величин  $x, y$ . Пусть  $p_z$  – гистограмма, приближающая плотность вероятности, арифметической операции над двумя случайными величинами  $x * y$ , где  $* \in \{+, -, \cdot, /\}$ . Тогда значение гистограммы на интервале  $[z_i, z_{i+1}]$  определяется по формуле [21]

$$p_i = \int_{\Omega_i} p(x, y) dx dy / (z_{i+1} - z_i), \quad (2.7)$$

где  $\Omega_i = \{(x, y) \mid z_i \leq x * y \leq z_{i+1}\}$ .

Рассмотренный выше подход обобщается на случай большего числа переменных. Пусть требуется найти гистограмму  $p_z$  суммы

$$z = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (2.8)$$

и пусть  $p(x_1, x_2, \dots, x_n)$  – плотность распределения вероятностей случайного вектора  $(x_1, x_2, \dots, x_n)$ . Тогда вероятность попадания  $z$  в интервал  $(z_i, z_{i+1})$  соответственно равна

$$P(z_i < z < z_{i+1}) = \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \quad (2.9)$$

где  $\Omega_i = \{(x_1, x_2, \dots, x_n) \mid z_i < a_1 x_1 + a_2 x_2 + \dots + a_n x_n < z_{i+1}\}$ ,  $p_{zi}$  имеет вид

$$p_{zi} = \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n / (z_{i+1} - z_i). \quad (2.10)$$

### Вероятностные расширения

Рассмотрим задачу определения закона распределения функции нескольких случайных аргументов.

Пусть  $f(x_1, \dots, x_n)$  – рациональная функция, тогда для вычисления гистограммы  $F$  заменим арифметические операции на гистограммные, а переменные  $x_1, x_2, \dots, x_n$  – их гистограммными значениями. Полученную гистограмму  $F$  будем называть –естественным гистограммным расширением [23].

**Теорема 1.**[23] Пусть  $f(x_1, \dots, x_n)$  – рациональная функция, каждая переменная которой встречается только один раз и  $x_1, \dots, x_n$  – независимые случайные величины. Тогда естественное гистограммное расширение аппроксимирует вероятностное расширение с точностью  $O(h)$ .

Для рациональной функции  $x(y+z) = xy + xz$  только первое представление попадает под условие теоремы 1. и, следовательно,

естественное гистограммное расширение будет аппроксимировать вероятностное с точностью  $O(h)$ .

## 2.2 Практическая реализация

### 2.2.1 Постановка задачи

Имеется случайная величина  $X$ , которая характеризуется функцией распределения  $F(x)$ , статистической характеристикой – дисперсией и математическим ожиданием. Предположим, что имеющаяся случайная величина имеет треугольное распределение.

1) Сгенерировать массив из случайных значений  $z_i$ , повторить генерацию 10000 раз и затем вычислить  $x_i$ , построив гистограмму отражающую частоту распределения последней величины на заданном интервале  $[0; 2]$ .

2) Повторить пункт 1, но генерацию осуществлять уже не случайно, а равномерно, сформировав так называемую «идеальную» выборку. Сравнить результаты этих двух этапов и сделать выводы.

3) Построить гистограмму распределения отдельных значений  $z_i$  на интервале  $[0; 1]$ , а также сгладить полученную гистограмму при помощи эрмитова сплайна.

### Треугольное распределение

Случайная величина  $\xi$  имеет треугольное распределение (распределение Симпсона) на отрезке  $[a, b]$  ( $a < b$ ), если функция плотности вероятности имеет вид:

$$f(x) = \begin{cases} \frac{2}{b-a} - \frac{2}{(b-a)^2} |a+b-2x| & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases} \quad (2.11)$$

Характеристическая функция (комплекснозначная функция вещественной переменной) треугольного распределения

$$\varphi(t) = \left[ \frac{2(e^{itb/2} - e^{ita/2})}{(b-a)it} \right]^2 \quad (2.12)$$

Свойства треугольного распределения: [9]

- Моменты:

$$E\xi^k = \frac{4}{(b-a)^2(k+1)(k+2)} \left[ a^{k+2} + b^{k+2} - 2 \left( \frac{a+b}{2} \right)^{k+2} \right] \quad (2.13)$$

- Дисперсия:

$$D\xi = \frac{(b-a)^2}{24} \quad (2.14)$$

- Коэффициент асимметрии:

$$\gamma_1 = 0;$$

- Коэффициент эксцесса:

$$\gamma_2 = -3/5.$$

Если  $\xi_1$  и  $\xi_2$  - независимые случайные величины, равномерно распределенные на отрезке  $[a/2, b/2]$ , то случайная величина  $\xi = \xi_1 + \xi_2$  имеет треугольное распределение. [8]

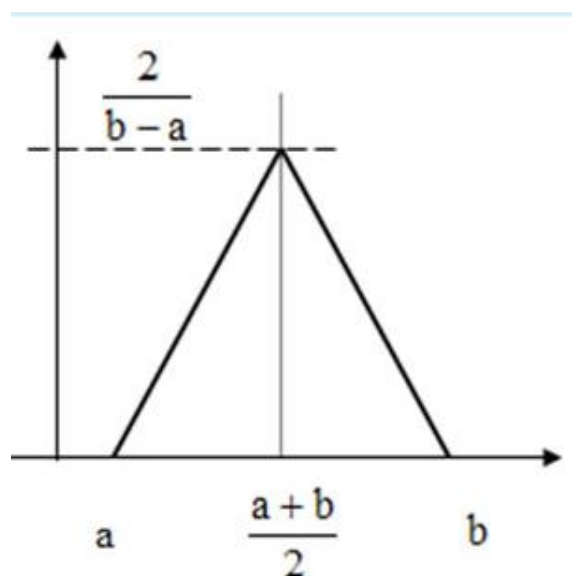


Рисунок 1 – Треугольное распределение

### 2.2.2 Алгоритм решения задачи



## Этап первый: построение гистограммы $x_i$

Решение задачи осуществлялось следующим образом:

- 1) Случайным образом формировался набор значений  $z_i$
- 2) Производилась сортировка массива  $z_i$
- 3) Для каждого значения массива  $z_i$  находилось соответствующее значение  $x_i$  по следующей формуле:

$$x_i = \begin{cases} \sqrt{2} * z_i, & \text{если } z_i \leq 0,5; \\ \frac{4 \pm \sqrt{16 - 8(1 + z_i)}}{2}, & \text{если } z_i > 0,5 \end{cases}$$

- 4) Пункты 2 и 3 повторялись  $N$  раз, после чего строилась гистограмма, отражающая частоту распределения  $x_i$  на заданном интервале

Итогом работы программы стала следующая гистограмма:

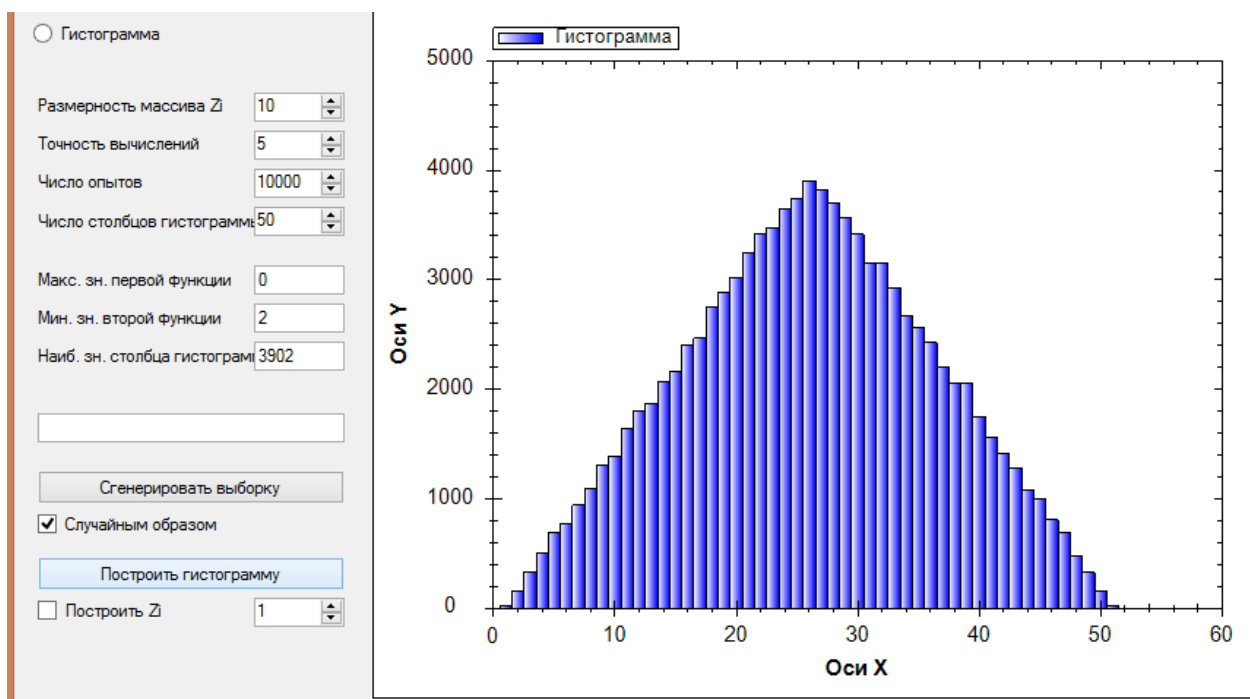


Рисунок 2 – Гистограмма, построенная на основе случайной выборки

Как мы видим, в результате образовалось так называемое треугольное распределение, однако также имеются и погрешности, что заметно в несколько неточном построении гистограммы – изменение высоты каждого следующего столбика отлично от предыдущего.

Для того чтобы избежать подобного рода погрешности, нужно делать выборку массива  $z_i$  не случайно, а равномерно, через равные интервалы. То есть, если мы строим выборку объемом в 10 значений на интервале (0; 1), то получим массив  $z_i = \{0.1, 0.2, 0.3, \dots, 0.9, 1\}$ . Итогом поправки получили другую гистограмму:

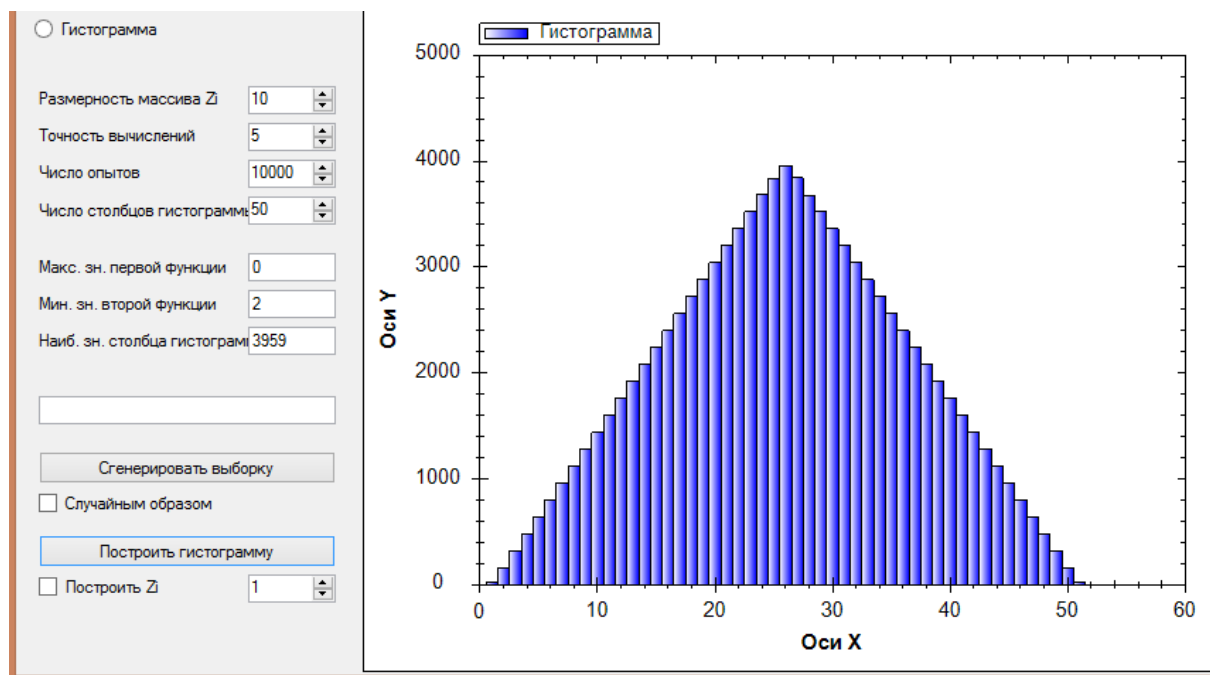


Рисунок 3 – Гистограмма, построенная на основе «идеальной» выборки

Теория подтвердилась – при наличии идеальной выборки, гистограмма получается симметричной, и высота столбцов меняется равномерно.

### Этап второй: построение гистограммы для $z_i$ поэлементно

Во втором этапе работы нашей задачей является проверка гипотезы о распределении элементов выборки  $z_i$ , когда она строится случайно. Гипотеза состоит в том, что при построении эмпирической функции, при выборке  $= 1/n$ , она не совсем точно совпадает с аналитической. Однако, если взять выборку  $= 1/n+1$ , то тогда мы получаем более явное совпадение. Соответственно, для проверки гипотезы требуется:

А) Построить для каждого элемента массива  $z_i$  гистограмму, отражающую частоту распределения этого элемента.

Б) Вычислить математическое ожидание для этих элементов и проверить будет ли оно равно теоретически выдвинутым значениям.

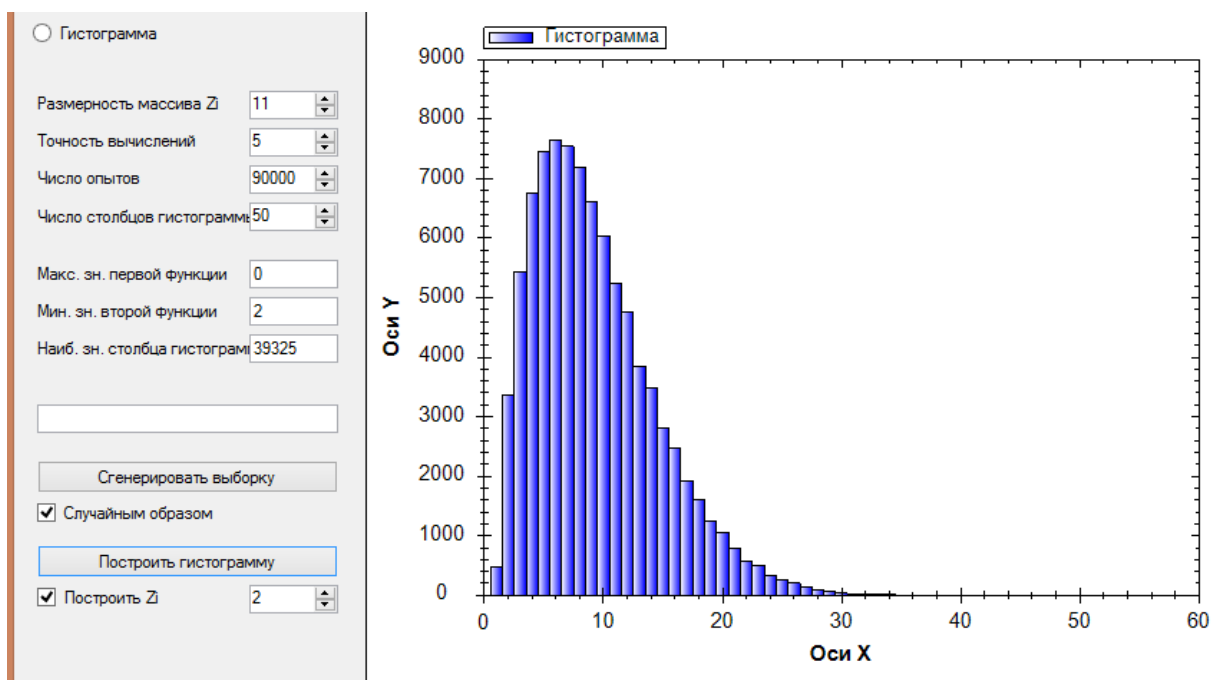


Рисунок 4 – Распределение 2-го элемента в массиве  $z_i$  объемом = 11 элементов

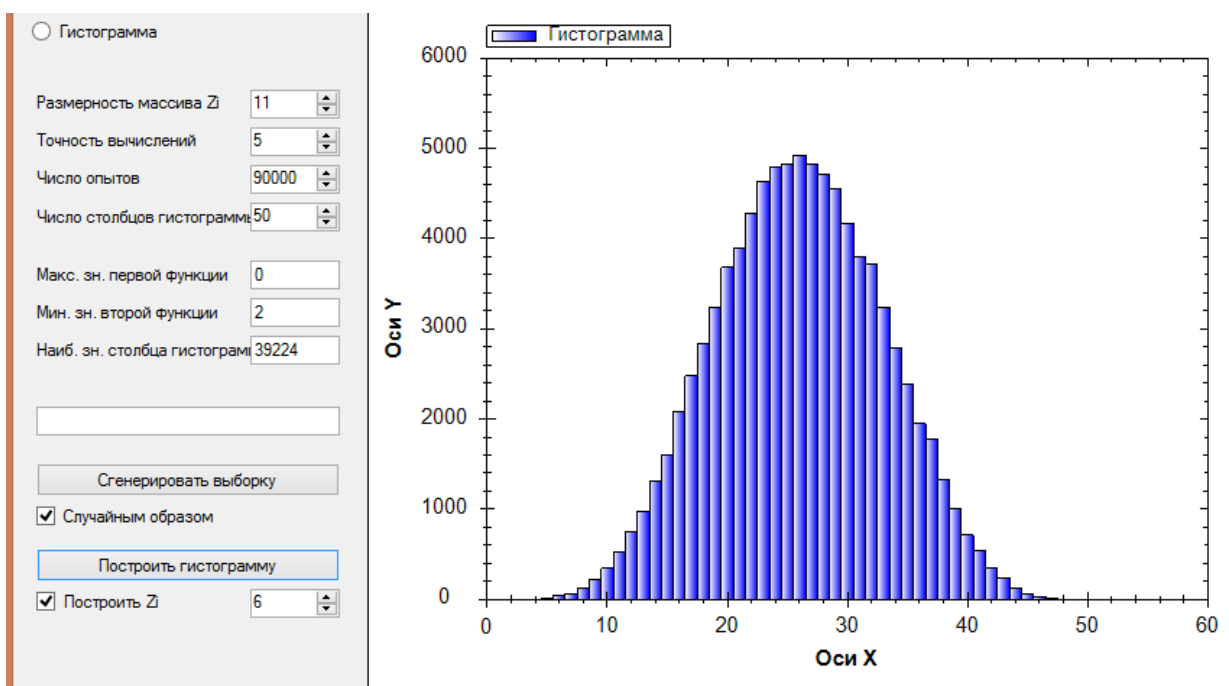


Рисунок 5 – Распределение 6-го элемента в массиве  $z_i$  объемом = 11 элементов

На данных рисунках отображается распределение случайной величины, для конкретной позиции массива  $Z_i$ . На рис.4 распределение построено для второго элемента, что подтверждается смещением распределения в левую

часть гистограммы. На рис.5 плотность распределения построена уже для 6-го элемента, что соответствует элементу, который находится точно в середине массива. И здесь уже частота выпадения случайных величин приходится на середину гистограммы, с плавным снижением частоты к крайним значениям.

### 2.3 Выводы по главе 2

В результате проделанной работы было выявлено, что при наличии так называемой «идеальной» выборки, когда избегается бросание по случайному закону с нормальным распределением, мы получаем более качественную гистограмму. Этот факт положительно сказывается на точности построения функции распределения случайной величины – в полученной гистограмме отсутствует «шум». Так же было выявлено, что при большем количестве бросаний гистограмма получается более сглаженной. Еще один важный момент – при увеличении количества столбцов гистограммы в случае, когда генерация случайных величин происходит случайным образом, гистограмма становится менее качественной, появляются выбросы и провалы (рис. 6-8).

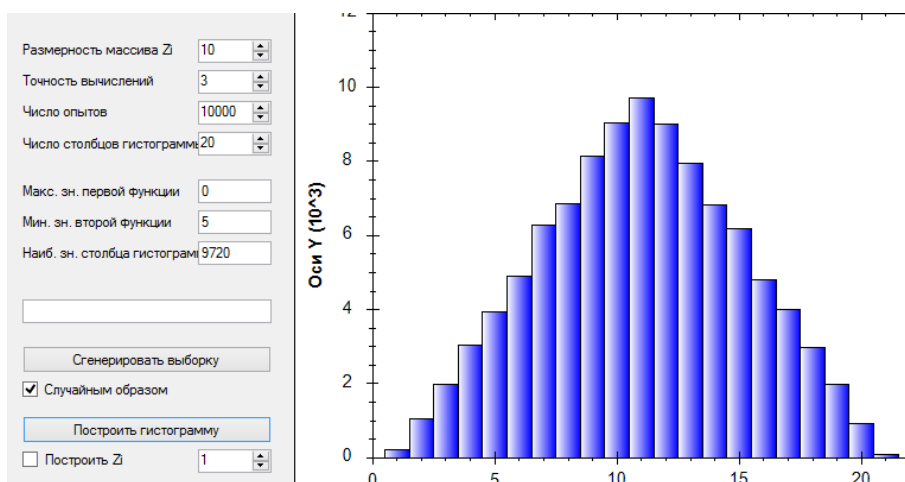


Рисунок 6 – Гистограмма с разрешением в 20 столбцов

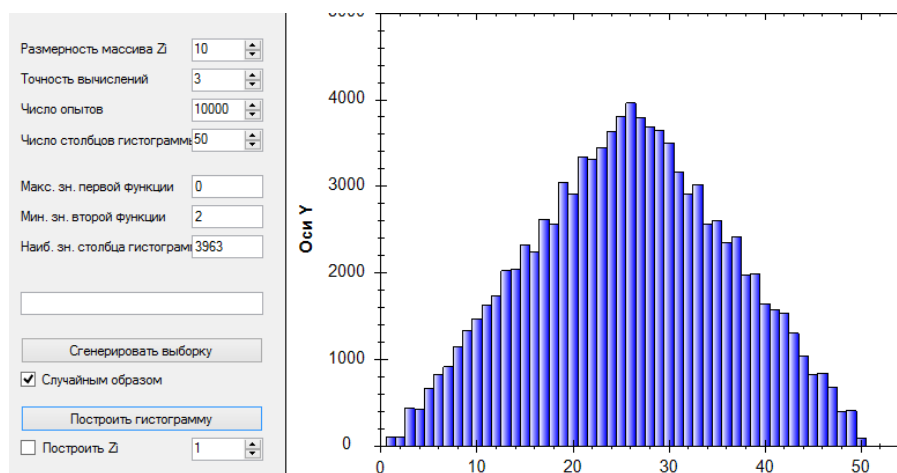


Рисунок 7 – Гистограмма с разрешением в 50 столбцов

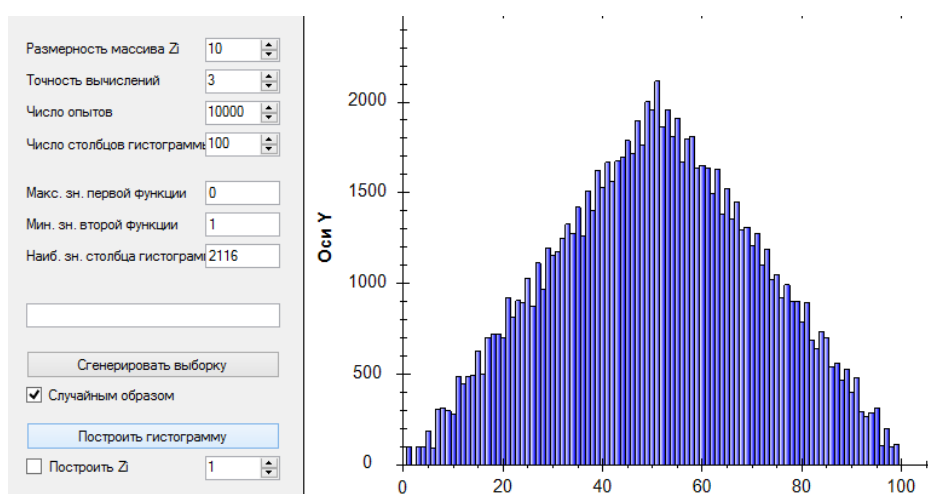


Рисунок 8 – Гистограмма с разрешением в 100 столбцов

Это можно объяснить тем, что в результате влияния погрешности измерений при достаточно малой ширине одного столбца, выборочные значения начинают попадать не в те интервалы, которым принадлежат истинные значения интересующей нас случайной величины. И в результате постепенное уменьшение ширины столбца приведет в конечном итоге к ухудшению получаемой гистограммы. Чтобы этого избежать, необходимо вводить некоторый критерий оптимальности, в основе которого, например, может лежать максимум получения полезной информации [3].

### 3 Методы сглаживания эмпирических функций распределения

#### 3.1 Теория сплайнов

Первоначально, сплайном (англ. *spline*) называли гибкую металлическую линейку – своеобразное лекало, применяемое чертёжниками для соединения точек на чертеже плавной кривой, другими словами – для графического исполнения интерполяции.

Более того, функция, описывающая деформацию гибкой линейки, фиксированной в некоторых точках, является сплайном. Соответственно, у нас есть физическая модель сплайн-функции (либо сплайн-функция является математической моделью гибкой линейки). Интуитивный метод к использованию кусочных функций в задачах приближения встречался в вычислительной математике в течение долгого времени. Физической моделью, представляющей собой механическую аналогию сплайна, является балка с несколькими точками опоры, на которую действуют исключительно внутренние силы, создающие деформацию смещения опор в заданных узлах. Математическая модель такой ситуации представляет собой дифференциальное уравнение деформации балки и по сути является многоточечной краевой задачей, впервые для решения которой был применён распространённый в то время сеточный метод, который получил решение именно в таком виде, называемом сегодня сплайном. Однако, советский учёный Николай Корнейчук говорит о том, что внедрение сплайнов в теорию приближения случилось из-за задачи интерполяции, во многом благодаря их привлекательным вычислительным и аппроксимативным свойствам. Сплайны позволяют приближать функции с достаточно высокой точностью, обладают универсальностью и обеспечивают лёгкость реализации вычислительных алгоритмов, полученных на их основе. Стоит отметить, что алгоритмы построения сплайнов аналогичен алгоритму метода конечных элементов, который в свою очередь является

главным промышленным методом анализа надежности систем автоматизированного проектирования (САПР).

Впервые, теория аппроксимации сплайнами и сам термин *сплайн* упоминается в статье Айзека Шонберга (англ. *Isaac Jacob Schoenberg*), датируемая 1946 годом. Пик ее развития пришелся на 50-70 годы. В данный момент, классической прикладной сферой применения интерполяционных сплайнов являются САПР. Но все же потенциальные возможности сплайнов гораздо шире, чем просто описание отдельных кривых. В реальных условиях огромное количество различных физических процессов по самой своей природе представляют собой сплайны. Например, в механике это деформация гибкой пластины или стержня, фиксированных в заданных точках; траектория передвижения объекта, если сила, воздействующая на него, изменяется ступенчато (траектория спутника с активными и инерционными интервалами движения, траектория полета самолёта при дискретном изменении тяги двигателей и изменении наклона крыла и т. д.). В термодинамике это теплообмен в системе, составленной из блоков с различной теплопередачей. В химии – диффузия через различные вещества. В электричестве – распространение электромагнитных полей через различного рода среды. Все вышесказанное говорит о том, что сплайн не выдуманная математическая абстракция, а реальный физический объект, который во многих случаях является решением дифференциальных уравнений, описывающих действительные физические процессы.

Теорию сплайнов начнём с описания алгебраического сплайна. Функция  $S(t)$ , определённая и непрерывная на отрезке  $[a, b]$ , называется полиномиальным сплайном порядка  $m$  с узлами  $x_j \in (a \leq x_0 < \dots < x_n \leq b)$ , если на каждом из отрезков  $[x_{j-1}, x_j)$ ,  $S(t)$  является алгебраическим полиномом в степени, не превышающей  $m$ , а в каждой из точек  $x_j$  некоторая производная  $S^{(w)}(t)$  может иметь разрыв. Если

функции  $S(t), S^{(i)}(t), \dots, S^{(m-k_i)}(t)$  в точках  $x_j$  постоянны, а производные  $S^{(m-k_i)}(t)$  в точках  $x_j$  терпят разрыв, то числа  $x_j$  называют дефектами сплайна. Множество  $\{x_0, x_1, \dots, x_n\}$  называют сеткой узлов сплайна, а точки  $x_j$  – узлами, точками соприкосновения или точками склейки сплайна.

Как следует из определения, для формирования сплайна, состоящего из  $n-1$  кусков, необходимо найти такие значения параметров для каждого куска – полинома степени  $m$ , которые зададут непрерывность в узлах, как самой функции, так и необходимых производных. Соответственно, всего следует обозначить  $(n-1) * m$  параметров. Учитывая условия интерполяции и непрерывности первых двух производных нахождение параметров сводится к решению системы, состоящей из  $n$  линейных уравнений. При этом обычно, значения коэффициентов для отрезков полиномов не рассчитываются.

Для определения интерполяционного сплайна с непрерывной первой производной необходимо вычислить значение первой производной в узлах. Метод определения производных в узлах сплайна определяет богатое разнообразие интерполяционных сплайнов. Зачастую, производные определяются не как константы, а как определенные зависимости от интерполируемой функции и сетки интерполяции.

Если значение первой производной в узлах вычислять исходя из условия непрерывности второй производной (решив систему, состоящую из  $n$  линейных уравнений), то сплайн будет содержать две непрерывные производные. Такой метод построения сплайна, как и сам сплайн называют глобальным, так как при вычислении каждого из его коэффициентов учитывается всё множество узлов интерполяции.

В остальных случаях, для вычисления отдельного коэффициента учитываются только ближайшие узлы интерполяции и такие методы построения, как и сами сплайны, называют локальными. Параметры



фрагмента локального сплайна можно вычислить независимо от других фрагментов.

Простым условием построения фрагмента локального сплайна является условие равенства полинома на концах отрезков соответствующим значениям интерполируемой функции.

$$P_j(t_j) = f(t_j), P_j(t_{j-1}) = f(t_{j-1}) \quad (3.1)$$

Для простейшего сплайна-ломаной – такого условия вполне достаточно. Два коэффициента прямой гарантированно вычисляются из двух уравнений. Такой сплайн является локальным. Для полиномов более высоких степеней необходимо добавить дополнительные условия так, чтобы общее количество уравнений было равно числу коэффициентов полинома. Например, для сплайна третьей степени таким условием будет равенство первой производной на концах отрезка значению, определяемому для соседних участков одинаковым образом (в формуле (3.2) через приближённое значение производной от функции).

$$P'_j(t_j) = f'(t_j), P'_j(t_{j-1}) = f'(t_{j-1}) \quad (3.2)$$

Система из 4-х уравнений

$$\left[ \begin{array}{l} P_j(t_j) = f(t_j) \\ P_j(t_{j-1}) = f(t_{j-1}) \\ P'_j(t_j) = f'(t_j) \\ P'_j(t_{j-1}) = f'(t_{j-1}) \end{array} \right] \quad (3.3)$$

дает возможность однозначно вычислить четыре коэффициента полинома. Для полинома пятой степени нужно добавить условие равенства второй производной на концах отрезка и т. д. Из всего вышесказанного должно быть ясно, почему сплайны зачастую строят из полиномов именно нечётных степеней (при этом с чётным числом коэффициентов).

Для полиномов чётных степеней при определении системы (3.3) остаётся неизвестной производная в одном из концов отрезка, поэтому и условие равенства производных (подразумевается гладкость кривой) не будет выполняться, поэтому для полинома второй степени невозможно достичь равенства первой производной в точках стыка, а для четвертой степени – второй производной и так далее. Для решения такой проблемы при построении сплайнов с чётными степенями вручную добавляют дополнительные граничные условия, чтобы сформировать систему уравнений, подобную (3.3). Если производные полинома сплайна вычисляются также, как соответствующие производные интерполируемой функции, сплайн называется является эрмитовым.

$$P_j^{(n)}(t_j) = f^n(t_j), P_j^{(n)}(t_{j-1}) = f^n(t_{j-1}) \quad (3.4)$$

Помимо рассмотренного способа построения сплайна также существуют локальные способы построения Бесселя и Акими, В – сплайны. В большинстве случаев, когда речь идет о сплайнах, то подразумеваются сплайны, построенные на основе алгебраических полиномов. Именно к таким сплайнам относится описанное выше определение, и именно сплайны такого рода являются наиболее изученными. Но при этом, сплайн может содержать в себе фрагменты функций любого класса. В работе [13] рассмотрено построение подобных сплайнов и исследуются их свойства. Однако, автор не даёт общего определения построенных сплайнов. Очевидно, что для любых классов функций, из которых состоит сплайн, приведённое в начале статьи определение не совсем подходит. Например, если сплайн состоит из отрезков экспоненты, то понятие дефекта сплайна теряет смысл. Хотя количество непрерывных производных останется важной характеристикой. Построение сплайна, фрагментами которого являются разрывные функции (рациональные функции, функции Паде), несколько выходит за рамки сплайновой идеи, поскольку одним из основных преимуществ сплайнов является их гладкость. Если произвольно расширять такие конструкции, то

стираются различия сплайнов от кусковых функций. Другим преимуществом сплайнов является эффективность вычислений. Чрезмерное усложнение фрагментов существенно снижает преимущество сплайнов перед классическими функциями.

Для сплайнов характерны следующие признаки: сплайн состоит из фрагментов – функций одного класса, которые отличаются только своими параметрами, на соседние фрагменты в точках стыковки накладываются определенные условия, которые сводятся к непрерывности значений и некоторых первых производных. Сплайны – направление прикладной математики, которое интенсивно развивается. В Интернете содержится обширная библиография по сплайнам (Splinebibliographydatabase (SBD)).

Как отмечалось выше, существует большое количество конструкций, которые называют сплайнами. Поэтому необходимо внести определенную классификацию в это многообразие, имея целью выделить те признаки, которые позволят выбрать сплайны годные для конкретной прикладной задачи.

**Назначение сплайнов.** По назначению можно выделить три основные группы сплайнов: «интерполяционные сплайны» или «функциональные сплайны» – проходящие точно через заданные точки, «сглаживающие сплайны» – проходящие через заданные точки с учётом погрешностей их определения; «корреляционные сплайны» – проходящие через корреляционное множество точек и отображающие его генеральную зависимость (тренд, регрессию). Интерполяционные и функциональные сплайны используют в задачах геометрического моделирования, например, задания обводов корпусов водных и воздушных судов. Сглаживающие сплайны используют чаще всего для описания зависимостей физических экспериментов с известной погрешностью измерений. Корреляционные сплайны используют в качестве нелинейных графиков регрессии, простейшими из которых можно считать описание зависимости ступенчатой и кусочно-линейной функцией (сплайнами нулевой и первой степени).

**Вид фрагментов сплайна.** То, что сплайн состоит из фрагментов одинакового вида, является одним из ключевых признаков, что отличает его от других кусковых функций. Однако существуют комбинированные сплайны, состоящие из фрагментов различных сплайнов.

Самые известные сплайны – состоящие из фрагментов – алгебраических полиномов не выше заданной степени. Как правило, это кубические полиномы, или полиномы нечётных степеней: первой, третьей (кубический), пятой степени. Более высокие степени применяют редко из-за усложнения расчетов и сложностей, описанных в предыдущем разделе. Основным их преимуществом является простота расчетов и анализа. Недостатком является то, что относительно мало реальных физических процессов соответствуют этой зависимости.

**Экспоненциальные сплайны.** Если гибкую металлическую линейку, зафиксированную в узлах, натянуть, то решением дифференциального уравнения будет не алгебраический полином, а экспонента. Поэтому такие сплайны называют также *напряжёнными*. Экспонента описывает многие физические процессы в динамических системах. Недостатком является трудоёмкость расчета.

По механической аналогии с металлической линейкой, представляющей собой расчётную модель балки, получают сплайны переменной жесткости, описанные в работах Снигирева В. Ф. и Павленко А. П. Первоначально такие сплайны называли вырождающимися или логарифмическими, так как решение исходного сплайнового дифференциального уравнения, представляющее собой фрагмент сплайна, будет содержать натуральные логарифмические функции. Жесткость в них может выступать как весовая, если она заранее задана, так и как управляющая функция, которая отыскивается из условий минимума функционала энергии оператора исходного сплайнового уравнения, аналогичного полной потенциальной энергии деформации линейки (балки). Функция жесткости позволяет управлять формой сплайна. В случае, когда

функция жесткости является управляющей функцией, то такие сплайны называют сплайнами минимальной жесткости.

Тригонометрическими являются сплайны, фрагменты которых описываются тригонометрическими полиномами. Имеют достаточно сложные расчётные выражения. Более пятидесяти различных по виду фрагментов сплайнов описаны в работах Б. А. Попова.

Также существуют рациональные сплайны и сплайны Паде. Их особенностью является возможность разрыва производных на фрагментах, при непрерывности в узлах. М. Ансерме строит фракциональные сплайны, где фрагменты заданы с помощью гамма-функции.

Целесообразность применения фрагментов определенного вида основана на конкретных условиях задачи и ограничениях реализации. Как правило, основное требование—это достижение заданной точности интерполяции при приемлемых затратах времени и ресурсов на реализацию. Удачный выбор фрагментов, который соответствует характеру процесса, позволяет сократить время вычислений и требуемый объём памяти.

**Число фрагментов.** Очевидно, что минимальное число фрагментов – один. Классическое определение сплайна ограничивает число фрагментов определенным числом на конечном отрезке. Однако можно строить сплайны и с бесконечным числом фрагментов, а реально эти методы и алгоритмы, которые не нуждаются в информации об определенном количестве фрагментов. Представителями этих сплайнов являются *кардинальные* сплайны, исследованные Шенбергом. Для построения сплайнов с неограниченным числом фрагментов лучше подходят локальные сплайны.

**Ширина фрагментов.** Следует выделить сплайны с равной шириной фрагментов. Это позволяет значительно упростить расчётные выражения, ускорить работу алгоритмов и снизить затраты на реализацию. Определенного упрощения можно достичь за счёт применения фрагментов с кратной шириной. Существуют сплайны с нулевой шириной фрагментов (Де

Бур). Это приводит к кратности узлов и возможности приближать сплайны с неразрывными фрагментами разрывных функций. Расчётные выражения получают в результате предельных переходов. Сплайны могут иметь также фрагменты с бесконечной шириной. Эти фрагменты должны быть крайними. Иногда это позволяет естественно задать краевые условия. Строго говоря, ширина фрагментов зависит от выбора параметра – аргумента сплайн-функции, а для этого требуется решать отдельную задачу параметризации. Идеальным выбором в качестве параметра является длина интерполируемой функции, которая не всегда известна, поэтому существует множество способов решения этой задачи. Наиболее распространен способ параметризации по хордам.

**Условия стыковки фрагментов.** Ещё один важный признак, что отличает сплайны. Когда идет речь о сплайнах, как правило, считают, что фрагменты стыкуются гладко. То есть обеспечивается непрерывность значений и первой производной. Понятие *дефекта сплайна* связано с числом непрерывных производных, которые имеет функция-фрагмент определенного вида и числом производных, непрерывность которых гарантирована в узлах. Экспонента, синусоида имеют бесконечное число производных. Для них это понятие не имеет смысла. Поэтому удобнее говорить прямо о числе производных, непрерывность которых гарантирована в узлах сплайна. Практически речь идет о непрерывности значений и первой, максимум второй производной. Разрыв второй и высших производных визуально не заметно, поэтому учитывается редко. Понятно, что первая производная в точках стыка может задаваться по-разному. Наиболее распространены два приёма. Значение первой производной выбирается так, чтобы обеспечить непрерывность второй (глобальные кубические сплайны минимального дефекта). Первая производная равняется первой производной интерполируемой функции (возможно приближенно) в эрмитовых сплайнах.

**Краевые условия.** Имеется 4 типа классических краевых условий и ряд неклассических. Если сплайны имеют ограниченное число фрагментов,

то, естественно, у них отсутствуют крайние фрагменты слева и справа, поэтому крайние узлы не с чем стыковать. Исключением являются лишь периодические сплайны, которые имеют естественное продолжение (3-й тип классических краевых условий). Иногда естественными называют краевые условия с нулевой производной, хотя никаких оснований считать их более естественными, чем другие, нет, но для кубического сплайна естественные (натуральные) краевые условия являются частным случаем 2-го типа классических краевых условий, задающего вторые производные на краях сплайна. В этом случае приравнивание вторых производных к нулю высвобождает края металлической линейки от нагружения изгибающим моментом, что естественным образом и происходило бы при прикладывании её к фиксированным (заданным) узлам в физическом пространстве. В 1-м типе классических краевых условий задают первые производные (касательные) на краях сплайна; во 2-м типе – задают вторые производные (кривизну); 3-й тип используется для интерполяции замкнутых или периодических линий и заключается в стыковке крайних фрагментов сплайна; 4-й тип используется когда на краях сплайна неизвестны ни первая, ни вторая производные и заключается в стыковке соседних пар крайних фрагментов (1-го со 2-м и последнего с предпоследним) по третьим производным, что на практике реализуется в проведении по узлам пар соседних крайних фрагментов функции, аналогичной одному фрагменту сплайна (у полиномиального сплайна – полинома той же степени, что и фрагмент сплайна). Используются различные комбинации краевых условий, которые сводятся к данным 4-м типам классических условий. В случае, если краевые условия нельзя свести к этим четырём типам, как, например, изменение на паре соседних крайних фрагментах сплайна его третьей производной по линейному (аффинному) закону, предложенное в работах Снигирева В. Ф., то такие условия называют неклассическим вариантом краевых условий. Далее приведены некоторые варианты, сводящиеся к классическим краевым условиям. Если сплайн имеет фрагменты одинаковой

ширины, считают недостающие фрагменты той же ширины. Другой вариант – это считать недостающие фрагменты продлёнными в бесконечность. Преимущество такого подхода в возможности экстраполяции. Можно считать ширину фрагментов нулевой. Расчётные выражения получают предельными переходами. Если взглянуть на краевые условия с точки зрения формирования сплайна из базисных функций, то они сводятся к продолжению соответствующих локальных базисных функций. Ширина соседних фрагментов влияет на их форму. А простое обрезание часто приводит к осцилляции и росту погрешности на краях. Важное значение краевые условия имеют при обработке изображений и в задачах с экстраполяцией.

**Дополнительные ограничения.** Они чаще всего касаются производных в узлах. Иногда они вытекают из физики процесса. Условия: неотъемлемость значений, равенство моментов, площадей, условия нормирования. Дополнительные условия иногда упрощают анализ свойств сплайнов, но могут серьёзно затруднять построение и затраты реализации.

Сетка точек интерполяции. Может существенно влиять на эффективность расчетов. Важны случаи равномерной сетки и равномерной сетки, с расстоянием между точками, кратным расстоянию между узлами сплайна. Нахождение сетки точек интерполяции (интерполяционных узлов) является задачей параметризации, о которой уже сказано в разделе «Ширина фрагментов».

**Локальные свойства базисных функций.** Сплайн можно представить как сумму взвешенных базисных сплайнов. Существенным является ширина этих базисных функций. Так, в глобальных сплайнах базисные сплайны ненулевые на всём отрезке интерполяции. Хотя стоит заметить, что с определенной точностью (достаточной для многих технических расчетов) их можно считать локальными. У локальных сплайнов ширина базисных функций невелика (четыре фрагмента у кубических эрмитовых сплайнов). Это существенно влияет на эффективность расчетов и затраты реализации.



**Форма представления.** Функции, задающие фрагменты сплайна, как правило, зависят от множества параметров, благодаря которым они меняют свою форму. Значения параметров на каждом из фрагментов индивидуальны. Эти параметры могут задавать конкретный сплайн. Для полиномиальных сплайнов это полиномиальные коэффициенты. Так, сплайн можно представить множеством параметров функций на каждом из фрагментов. Назовем это представление пофрагментным. Такое представление является наглядным, часто имеет явный физический смысл. Но число параметров является чрезмерным. Так, для кубического сплайна необходимо иметь  $4 * (r - 1)$  параметров ( $r$  – число узлов сплайна). Данное представление получается в результате неопределенного интегрирования фрагмента исходного сплайнового дифференциального уравнения и называется аналогичной кусочно-полиномиальной формой (pp-формой) по аналогии с полиномиальными сплайнами. Для явного выражения коэффициентов через уже известные значения координат узловых точек, применяют разложение аналогичной кусочно-полиномиальной формы на базисные функции путем подстановки её в краевые условия Эрмита (граничные условия фрагмента сплайна, условия интерполирования и опирания на производные). В результате получается базисная форма (B-форма) сплайна. Такое представление сплайна является значительно более компактным и записывается через базисные сплайн-функции в виде:

$$S(x) = \sum_{j=1}^r a_j B_j(x) \quad (3.5)$$

где  $B_j(x)$  – базисные сплайн-функции (как правило локальные),  $a_j$  – числовые коэффициенты, задающие вес базисных функций при формировании сплайна, физическим смыслом которых являются обобщённые (линейные и угловые) перемещения металлической линейки в узлах. Число параметров, задающих сплайн, равно числу узлов сплайна. Между параметрами функции на фрагменте и коэффициентами полинома-

сплайна существует зависимость, что позволяет с одними коэффициентами находить другие, хотя формулы могут иметь достаточно сложный вид.

Преобразование аналогичной кусочно-полиномиальной формы представления сплайна в базисную форму снижает порядок системы линейных алгебраических уравнений для нахождения неизвестных коэффициентов сплайна, так как они частично выражаются через уже известные параметры – координаты заданных точек (узлов), что позволяет значительно снизить вычислительные затраты за счет возможности применить экономичные методы решения, такие как метод алгебраической прогонки или разновидности метода Гаусса для разрежённых (ленточных) матриц с выбором ведущего элемента столбца.

**Содержание коэффициентов сплайна.** Как отмечалось в предыдущем пункте, содержание параметров сплайна при пофрагментном представлении определяется типом функции. При полиномиальном представлении следует выделить случай, когда коэффициенты имеют тот же физический смысл, что и входные данные. То есть, коэффициенты являются значениями сплайна в узлах. Такую форму называют Лагранжевой, по аналогии с полиномом Лагранжа. Следует заметить, что базисные сплайны этой формы равны единице в центральном узле и нулю во всех остальных.

Коэффициенты интерполяционных и функциональных сплайнов всегда содержат значения координат заданных точек, вытекающие из условий интерполирования. А также в зависимости от условий опоры на производные, содержат значения соответствующих производных на границах фрагмента сплайна (в узловых точках). Как правило, при записи таких условий фрагмент сплайна на его границах опирают на первые или вторые производные. Базирование фрагмента сплайна на первые производные явно отражает физический смысл, так как первые производные (касательные) – это угловые перемещения (повороты) металлической линейки относительно поперечной оси. Базирование сплайна на вторые производные применяют для упрощения вида расчётных выражений с целью уменьшения ошибок при

их ручной перезаписи, однако в некоторых случаях использование таких выражений в каких-либо дополнительных условиях может приводить к тривиальным решениям.

**Особые сплайны.** В ряде случаев рассматривают функции, которые находятся близко к границе между сплайнами и обычными функциями, а также сплайнами и кусковыми функциями. К примеру, это сплайны, состоящие из двух фрагментов. Имеют упрощенный вариант построения, но особое внимание следует уделять краевым условиям.

К особым сплайнам можно отнести многомерный ортогональный нормированный сплайн, описывающий нелинейную модель искусственного нейрона (сплайн-модель Хакимова), используемый для моделирования зависимости функции от совокупности множества аргументов.

## **3.2 Порядковые статистики**

### **3.2.1 Определения случайной выборки**

Пусть  $X$  – исследуемая случайная величина,  $F_X(x) = P(X < x)$  – ее функция распределения (вообще говоря, неизвестная). В ряде случаев может быть известен вид распределения случайной величины, а неизвестными являются один или несколько параметров, от которых зависит функция распределения. Ради краткости в записи  $F_X(x)$  индекс может в дальнейшем опускаться. Условимся также указывать, непрерывной или дискретной является исследуемая случайная величина.

Пусть проводится серия  $n$  независимых наблюдений (измерений) случайной величины  $X$  в одних и тех же условиях (эксперимент). В результате эксперимента получают  $n$  чисел – значений  $x_1, x_2, \dots, x_n$ , которые случайная величина  $X$  последовательно принимала в данной серии наблюдений. Эти числа будем считать значениями  $n$  одинаково распределенных независимых случайных величин  $X_1, \dots, X_n$ , каждая из

которых имеет функцию распределения  $F_X(x)$  – ту же, что исследуемая случайная величина  $X$ .

Конечную последовательность  $n$  независимых одинаково распределенных случайных величин будем называть случайной выборкой  $X_1, \dots, X_n$  (короче – выборкой) из распределения  $F_X(x)$ , а указанные числа  $x_1, x_2, \dots, x_n$ , полученные в данном эксперименте – реализацией выборки. Отметим, что множество всех возможных значений исследуемой случайной величины называют генеральной совокупностью.

На основе выборок строят оценки параметров распределения исследуемой случайной величины  $X$ , таких как математическое ожидание, стандартное отклонение и других, а также судят о виде функции распределения  $F_X(x)$ .

Понятно, что числа  $x_1, x_2, \dots, x_n$  можно также рассматривать как значение  $n$  – мерной случайной величины  $(X_1, \dots, X_n)$ , компоненты которой  $X_1, \dots, X_n$  независимы и одинаково распределены.

Всякую функцию выборки  $\varphi(X_1, \dots, X_n)$  называют статистикой. Статистика  $\varphi(X_1, \dots, X_n)$  – случайная величина, распределение которой зависит от распределения  $F_X(x)$ , из которого извлечена выборка, и от объема выборки  $n$ .

### 3.2.2 Закон распределения порядковых статистик

Пусть  $X_1, \dots, X_n$  – выборка объема  $n$  из распределения  $F_X(x)$ ;  $x_1, x_2, \dots, x_n$  – некоторая ее реализация.

Упорядочим числа  $x_1, x_2, \dots, x_n$  по возрастанию и обозначим их следующим образом:  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ , где  $x_{(1)} = \min(x_1, x_2, \dots, x_n)$ ,  $x_{(n)} = \max(x_1, x_2, \dots, x_n)$ .

Представим, что упорядочены все возможные реализации выборки  $X_1, \dots, X_n$  и введем новую случайную величину  $X_{(k)}$  – порядковую статистику порядка  $k$  ( $k = 1, 2, \dots, n$ ).

Множество возможных значений случайной величины  $X_{(k)}$  определим так: оно состоит из тех и только тех чисел  $x_{(k)}^i$ , которые оказываются на  $k$  – м

месте при упорядочении любой реализации  $x_1, x_2, \dots, x_n$  выборки  $X_1, \dots, X_n$  (индекс  $i = 1, 2, \dots$  – номер реализации).

Таким образом, по выборке  $X_1, \dots, X_n$  построена последовательность  $X(1), \dots, X(k), \dots, X(n)$ , называемая вариационным рядом. Элементы вариационного ряда – порядковые статистики удовлетворяют соотношениям:  $X(1) \leq \dots \leq X(k) \leq \dots \leq X(n)$ , при этом в любой реализации вариационного ряда числа  $x_{(1)}^i, \dots, x_{(k)}^i, \dots, x_{(n)}^i$  связаны неравенствами  $x_{(1)}^i \leq \dots \leq x_{(k)}^i \leq \dots \leq x_{(n)}^i$  (верхний индекс  $i$  – номер реализации,  $i = 1, 2, \dots$ ).

$$\begin{array}{cccc}
 x_{(1)}^1 & \dots & x_{(k)}^1 & \dots & x_{(n)}^1 \\
 \cdot & & \cdot & & \cdot \\
 \cdot & & \cdot & & \cdot \\
 \cdot & & \cdot & & \cdot \\
 x_{(1)}^i & \dots & x_{(k)}^i & \dots & x_{(n)}^i \\
 \cdot & & \cdot & & \cdot \\
 \cdot & & \cdot & & \cdot \\
 \cdot & & \cdot & & \cdot \\
 \hline
 X_{(1)} & \dots & X_{(k)} & \dots & X_{(n)}
 \end{array} \tag{3.6}$$

Найдем функцию распределения  $k$  – й порядковой статистики

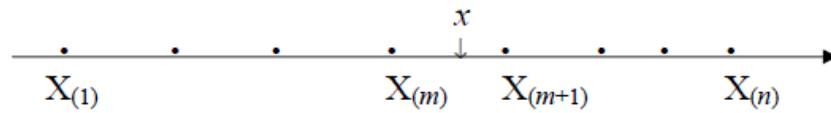
$$X_{(k)} : F_{X_{(k)}}(x) = P(X_{(k)} < x) \tag{3.7}$$

Эмпирической частотой  $N_n(x)$  назовем случайную величину, равную числу элементов выборки  $X_1, \dots, X_n$ , меньших  $x$  (иначе – числу элементов вариационного ряда  $X(1), \dots, X(n)$ , меньших  $x$ ). Ясно, что возможные значения эмпирической частоты  $N_n(x)$  – число осуществлений события  $(X < x)$  на выборке  $X_1, \dots, X_n$  объема  $n$  – это числа  $m = 0, 1, \dots, n$ . Действительно,

$$(N_n(x) = 0) = (x \leq X_{(1)}); \tag{3.8}$$

$$(N_n(x) = m) = (X_{(m)} < x \leq X_{(m+1)}) \forall m = 1, \dots, n-1; \tag{3.9}$$

$$(N_n(x) = n) = (x > X_{(n)}). \quad (3.10)$$



Извлечение выборки из распределения  $F_X(x)$  представляет собой серию  $n$  независимых испытаний –  $n$  наблюдений (регистраций значений) исследуемой случайной величины  $X$ . Для каждого из указанных испытаний вероятность события  $(X < x)$  равна  $P(X < x) = F_X(x)$ .

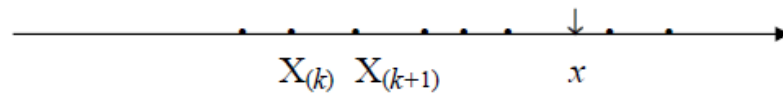
Отсюда следует, что случайная величина  $N_n(x)$  подчиняется биномиальному распределению:

$$P(N_n(x) = m) = C_n^m (F_X(x))^m (1 - F_X(x))^{n-m} \quad (3.11)$$

Заметим, что события  $(X_k(x) < x)$  и  $(N_n(x) \geq k)$  равносильны,

то есть  $X_k(x) < x = (N_n(x) \geq k) = \sum_{m=k}^n N_n(x) = m$ . Таким образом, получаем:

$$P(X_k < x) = F_{X_{(k)}}(x) = \sum_{m=k}^n C_n^m (F_X(x))^m (1 - F_X(x))^{n-m} \quad (3.12)$$



При  $k = 1$  и  $k = n$  имеем распределения экстремальных порядковых статистик:

$$\text{минимальной } X_{(1)} : F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n$$

$$\text{максимальной } X_{(n)} : F_{X_{(n)}}(x) = (F_X(x))^n$$

### 3.2.3 Эмпирическая функция распределения

Пусть  $X_1, \dots, X_n$  – выборка из распределения  $F_X(x)$ ,  $X(1), \dots, X(k), \dots, X(n)$  – вариационный ряд,  $N_n(x)$  – эмпирическая частота. Случайная величина  $F_n(x) = N_n(x)/n$ , называемая эмпирической функцией распределения –

относительная частота числа элементов выборки  $X_1, \dots, X_n$ , удовлетворяющих условию  $X_i < x$ .

Ясно, что множество возможных значений эмпирической функции распределения есть:  $0, 1/n, \dots, m/n, \dots, n/n$ .

События  $F_n(x) = m/n$  и  $N_n(x) = m$  – равносильны, эмпирическая частота  $N_n(x)$  распределена по биномиальному закону, поэтому

$$P(N_n(x) = m/n) = C_n^m (F_X(x))^m (1 - F_X(x))^{n-m} \quad (3.13)$$

( $m = 0, 1, \dots, n$ ) – закон распределения эмпирической функции распределения  $F_n(x)$ .

С помощью функции единичного скачка (функции Хевисайда) эмпирическая функция распределения может быть записана в виде:

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n e(x - X_{(k)}) \quad (3.14)$$

где  $e(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$  – функция Хевисайда.

Заметим, что для каждой реализации выборки эмпирическая функция распределения обладает всеми свойствами функции распределения. Действительно, пусть  $x_1, x_2, \dots, x_n$  – некоторая реализация выборки,  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  – соответствующая реализация вариационного ряда, где  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ .

Среди чисел  $x_1, x_2, \dots, x_n$  выберем только различные, упорядочим их и обозначим через  $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ , тогда

$$x_{(1)} = x_{i_1} < \dots < x_{i_k} = x_{(n)} \quad (3.15)$$

$x_{i_1}$	...	$x_{i_m}$	...	$x_{i_k}$
$n_1$	...	$n_m$	...	$n_k$
$n_1/n$	...	$n_m/n$	...	$n_k/n$

Здесь  $n_m$  – абсолютная,  $n_m/n$  – относительная частота элемента  $x_{i_m}$ , при этом, очевидно:  $\sum_{m=1}^k n_m = n$ ;  $\sum_{m=1}^k \frac{n_m}{n} = 1$ .

Введем случайную величину  $X^*$ , заданную рядом распределения

$X^*$	$x_{i_1}$	...	$x_{i_m}$	...	$x_{i_k}$
$P(X^* = x_{i_m})$	$n_1/n$	...	$n_m/n$	...	$n_k/n$

Заметим, что таким образом каждому элементу реализации выборки  $x_1, x_2, \dots, x_n$  приписана вероятность  $1/n$ .

Обозначим через  $F_n^*(x)$  реализацию случайной величины  $F_n(x)$ , отвечающую данной реализации выборки, тогда

$$F_n^*(x) = P(X^* < x) = \sum_{m: x_{i_m} < x} \frac{n_m}{n} \quad (3.16)$$

$F_n^*(x)$  – кусочно-постоянная (ступенчатая) функция, принимающая свои значения на отрезке  $[0; 1]$ . В каждой точке  $x$ , кроме точек  $x_{i_m}$ , функция  $F_n^*(x)$  непрерывна; в точках  $x_{i_m}$  – она непрерывна справа, величина скачка слева равна  $n_m/n$  ( $m = 1, 2, \dots, k$ ).

График эмпирической функции распределения для некоторой реализации выборки приведен ниже:

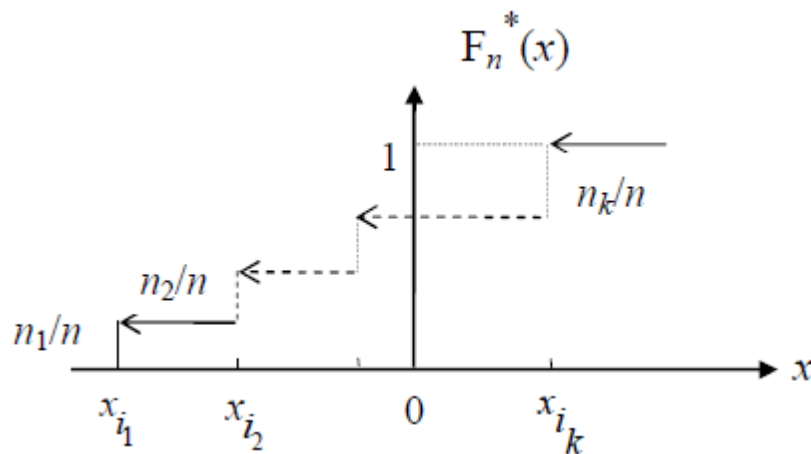


Рисунок 9 – Эмпирическая функция распределения



Эмпирическая функция распределения  $F_n(x)$  как относительная частота числа осуществлений на выборке события ( $X < x$ ) при любом  $x$  сходится по вероятности к вероятности этого события  $P(X < x) = F_X(x)$  – к теоретической функции распределения, вообще говоря, неизвестной:  $\forall x F_n(x) \xrightarrow[n \rightarrow +\infty]{P} F_X(x)$ , поэтому, если объем выборки  $n$  достаточно велик, то значение эмпирической функции распределения  $F_n(x)$  в каждой точке  $x$  оказывается близким к соответствующему значению теоретической функции распределения  $F_X(x)$ .

Доказано (теорема Гливенко), что отклонение эмпирической функции распределения  $F_n(x)$  – случайной величины – от теоретической функции распределения  $F_X(x)$  с вероятностью 1 сколь угодно мало при достаточно большом объеме выборки  $n$ :

$$\forall \varepsilon > 0 P(\sup_x |F_n(x) - F_X(x)| \leq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 1 \quad (3.17)$$

при этом  $F_n(x)$  служит равномерным приближением  $F_X(x)$  на всей числовой оси. Заметим, что разность  $(F_n(x) - F_X(x))$  асимптотически нормальна с нулевым математическим ожиданием.

### 3.2.4 Группирование выборочных данных, гистограмма

Эмпирическая функция распределения является характеристикой выборки, позволяющей наглядно представлять статистические данные и выдвигать предположения о виде неизвестной функции распределения исследуемой (наблюдаемой) случайной величины.

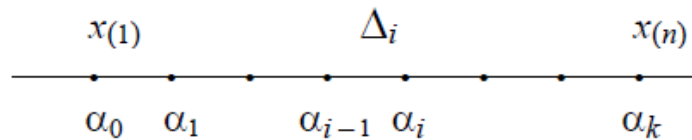
Другой способ представления статистического материала – это построение группированного статистического ряда и гистограммы.

Пусть исследуемая случайная величина  $X$  – непрерывна. Если выборка достаточно большая (обычно в статистике большими считают выборки объемом  $n \geq 100$ ), то ее реализацию  $(x_1, x_2, \dots, x_n)$  подвергают группировке следующим образом.

Отрезок  $[x_{(1)}; x_{(n)}]$ , где  $x_{(1)} = \min(x_1, x_2, \dots, x_n)$ ,  $x_{(n)} = \max(x_1, x_2, \dots, x_n)$ , содержащий все элементы выборки, разбивают на  $k$  равных интервалов  $\Delta_i$  (обычно  $5 \leq k \leq 15$ ):

$$\alpha_{(0)} = x_{(1)}, \alpha_{(k)} = x_{(n)}, \Delta_i = (\alpha_{i-1}; \alpha_i) \quad (3.18)$$

где  $|\Delta_i| = \frac{x_{(n)} - x_{(1)}}{k} = h$  – шаг разбиения.



Число  $n_i$  – частота,  $\frac{n_i}{n}$  – относительная частота числа элементов реализации выборки, попавших в  $i$  – й интервал ( $\sum_{i=1}^k n_i = n, \sum_{i=1}^k \frac{n_i}{n} = 1$ )

Наглядное графическое представление группированного статистического ряда дает гистограмма. Гистограммой называют ступенчатую фигуру, построенную следующим образом: на каждом интервале  $\Delta_i = (\alpha_{i-1}; \alpha_i)$ , как на основании длиной  $h = |\Delta_i|$ , строят прямоугольник с высотой, равной  $\frac{n_i}{nh}$ , так что площадь  $S_i$  каждого такого прямоугольника оказывается равной относительной частоте  $\frac{n_i}{n}$  числа элементов выборки, попавших в интервал  $\Delta_i$  ( $i = 1, \dots, k$ ).

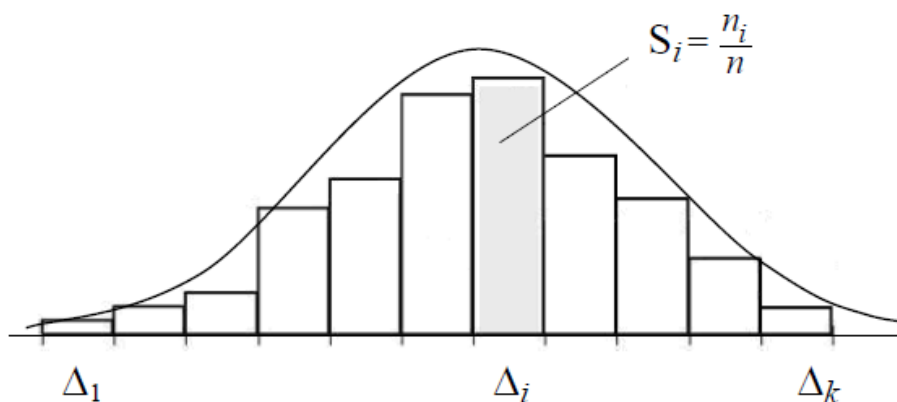


Рисунок 10 – Гистограмма

Относительная частота события по вероятности сходится к вероятности этого события, поэтому если длина интервалов разбиения  $h$  достаточно мала, то  $\frac{n_i}{n} \cong f_X(x)h \forall x \in \Delta_i$ . При больших  $n$  верхний контур гистограммы (ступенчатый график) служит приближением графика плотности вероятности  $f_X(x)$  (вообще говоря, неизвестной). Таким образом, разумно построенная гистограмма позволяет выдвинуть гипотезу о виде распределения исследуемой случайной величины  $X$ . Заметим, что слишком малое или слишком большое число интервалов разбиения  $k$  при построении гистограммы может привести к ее недостаточной информативности.

Число интервалов  $k$  при разбиении отрезка  $[x_{(1)}; x_{(n)}]$ , обычно определяют по формуле  $k = 1 + 3,32 \lg n$  (формула Старджесса), либо по формуле  $k = 1,72n^{1/3}$ .

### 3.2.5 Понятие доверительного интервала

Пусть  $\theta$  – некоторый неизвестный параметр распределения. По выборке  $X_1, \dots, X_n$  из данного распределения построим интервальную оценку параметра  $\theta$  распределения, то есть найдем такой интервал, внутри которого с заданной (высокой) вероятностью  $1 - \alpha$  находится истинное значение неизвестного параметра  $\theta$ . Указанную вероятность  $1 - \alpha$  называют доверительной вероятностью, а величину  $\alpha$  – уровнем значимости.

В качестве значений доверительной вероятности обычно выбирают величины 0,9, 0,95, 0,99, достаточно близкие к 1. В каждом конкретном случае выбор величины доверительной вероятности определяется спецификой решаемой практической задачи.

Итак, пусть  $X_1, \dots, X_n$  – выборка из данного распределения и задана величина доверительной вероятности  $1 - \alpha$ . Интервал  $(\vartheta_1; \vartheta_2)$  называют доверительным интервалом для параметра  $\theta$ , отвечающим доверительной вероятности  $1 - \alpha$ , если его границы  $\vartheta_1 = \vartheta_1(X_1, \dots, X_n)$  и  $\vartheta_2 = \vartheta_2(X_1, \dots, X_n)$ .

Заметим, что границы доверительного интервала  $\vartheta_1$  и  $\vartheta_2$  – случайные величины (функции выборки  $X_1, \dots, X_n$ ), параметр  $\theta$  – неслучайная величина, так что интервал  $(\vartheta_1; \vartheta_2)$  “накрывает” величину  $\theta$  с вероятностью  $1 - \alpha$  (соответственно, “не накрывает” с вероятностью  $\alpha$ ).

Длина интервала  $\vartheta_2 - \vartheta_1$  характеризует точность, а доверительная вероятность  $1 - \alpha$  – надежность интервальной оценки. Очевидно, что точность и надежность взаимосвязаны: увеличение надежности приводит к уменьшению точности – увеличению длины интервала  $(\vartheta_2 - \vartheta_1)$ . Выбирая величину доверительной вероятности  $1 - \alpha$ , принимают соглашение: считать события, вероятность которых  $P \geq 1 - \alpha$ , – практически достоверными, а события, вероятность которых  $P \leq \alpha$  – практически невозможными.

### 3.2.6 Основные этапы построения доверительных интервалов

Напомним, что границы доверительного интервала  $\vartheta_1(X_1, \dots, X_n)$  и  $\vartheta_2(X_1, \dots, X_n)$  – случайные величины (функции выборки  $X_1, \dots, X_n$ ). Результатом эксперимента (серии  $n$  независимых измерений данной случайной величины) является реализация выборки  $x_1, x_2, \dots, x_n$ . Соответственно, значения статистик  $\vartheta_1$  и  $\vartheta_2$  в точке  $(x_1, x_2, \dots, x_n)$  – это числа  $\vartheta_{1e} = \vartheta_1(x_1, x_2, \dots, x_n)$  и  $\vartheta_{2e} = \vartheta_2(x_1, x_2, \dots, x_n)$ .

Таким образом, будем различать, идет ли речь о доверительном интервале, границы которого по смыслу – случайные величины, или о реализации доверительного интервала, границами которого являются конкретные числа. Для построения реализации доверительного интервала на основе данной реализации выборки  $x_1, x_2, \dots, x_n$  выполняют следующие действия:

(1) формулируют предположения о распределении и о выборке  $X_1, \dots, X_n$  (допущения, принимаемые при построении априорной теоретической модели).

(2) строят доверительный интервал, для чего: а) выбирают значение доверительной вероятности  $1-\alpha$  (или уровня значимости  $\alpha$ ), то есть принимают соглашение: – вероятность практически достоверного события =  $1-\alpha$ ; – вероятность практически невозможного события =  $\alpha$ . б) записывают вероятностное равенство

(3) проводят эксперимент – получают конкретную реализацию выборки  $x_1, x_2, \dots, x_n$ .

(4) вычисляют значения  $\vartheta_{1e} = \vartheta_1(x_1, x_2, \dots, x_n)$  и  $\vartheta_{2e} = \vartheta_2(x_1, x_2, \dots, x_n)$ .

В результате перечисленных действий (1) – (4) получают реализацию доверительного интервала – числовой интервал  $(\vartheta_{1e}; \vartheta_{2e})$ .

### 3.3 Надежные оценки эмпирических функций распределения

Пусть  $\xi_1, \dots, \xi_n$  – выборка случайной величины  $X$  с функцией распределения  $F(t)$ ,  $t \in [a, b]$ . Эмпирическую функцию распределения определяет следующим образом

$$F_n(t) = \frac{m_t}{n} \quad (3.19)$$

где  $m_t$  число элементов  $\xi_i < t$ .

Пусть  $z_i = F(x_i)$ ,  $i = 1, \dots, n$ . Заметим что  $z_i, i = 1, \dots, n$  – равномерно распределенные случайные величины на  $[0, 1]$ . Если  $z_1 \leq z_2 \leq \dots \leq z_n$ , тогда  $z_k$  –  $k$ -я порядковая статистика и математическое ожидание  $M[z_k] = k / (n+1)$  [24]. Далее будем использовать точки  $(x_i, i / (n+1))$  для построения аппроксимации функции распределения  $F(t)$ . Пусть  $\omega = \{a = \xi_0 < \xi_1 < \dots < \xi_n < b = x_{n+1}\}$  – сетка. Построим на  $\omega$  кусочно-линейную функцию  $s$

$$s(x_i) = i / (n+1), \quad i = 1, \dots, n, \quad s(a) = 0, \quad s(b) = 1 \quad (3.20)$$

Заметим, если бы могли вместо математических ожиданий  $i/(n+1)$  использовать точные значения  $z_i$ , тогда погрешность кусочно-линейной функции  $s(x)$  на сетке  $\{x_i\}$  с  $h = \max(x_{i+1} - x_i)$  удовлетворяет оценке

$$\|F - s\| \leq h^2 \|F^{(2)}\| \quad (3.21)$$

Таким образом, даже при относительно небольших  $n$ , построенные оценки достаточно хорошо аппроксимируют функцию распределения  $F$ . Относительно  $z_i$  известно, что они образуют порядковые статистики.

Плотность вероятности  $k$ -й порядковой статистики

$$p_k(z) = \frac{n!}{(n-k)!(k-1)!} z^{k-1} (1-z)^{n-k} \quad (3.22)$$

Совместная плотность вероятности вектора  $(z_i, z_k)$  выражается следующим образом

$$p_{j,k}(z_j, z_k) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!} z_j^{j-1} (z_k - z_j)^{k-j-1} (1-z_k)^{n-k} \quad (3.23)$$

$$j < k, 0 \leq z_j \leq z_k \leq 1 \quad (3.24)$$

Каждому случайному вектору  $z_1 \leq z_2 \leq \dots \leq z_n$  соответствует кусочно-линейная функция  $s$ . Перебирая все возможные случайные вектора  $z_1 \leq z_2 \leq \dots \leq z_n$  получаем все множество кусочно-линейных функций  $\{s\}$ . Заметим, что  $\{s\}$  содержит интерполянт функции распределения  $F$ . Таким образом, используя для значения в узле  $\xi_k$  плотность вероятности  $k$ -й порядковой статистики множество  $\{s\}$  можно представить в виде случайной кусочно-линейной функции  $L$ . Соответственно  $L$  — надежная оценка эмпирической функции распределения.

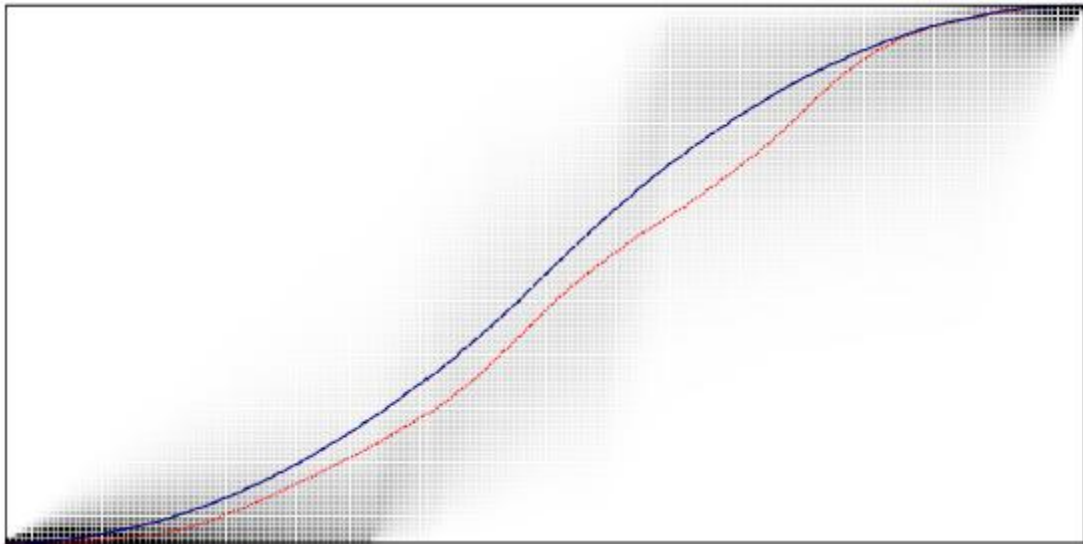


Рисунок 11 – Надежная оценка функции распределения

Рассмотрим аппроксимацию случайной кусочно-линейной функции  $L$  в виде гистограммной функции распределения (ГФР) [20]. Построим две сетки  $\omega_y = \{y_i = i/N_y, i = 0, 1, \dots, N_y\}$  и  $\omega_t = \{0 = t_0 < t_1 < \dots < t_{N_t}\}$ . Не ограничивая общности будем считать  $\omega \subset \omega_t$ . На сетке  $\omega$  построим гистограммную кусочно-линейную функцию. Для значений в узле  $\xi_k$  будем использовать плотность вероятности  $k$ -й порядковой статистики. Поскольку известны совместные плотности вероятности для порядковых статистик, то используя численные операции над плотностями случайных величин, построим на сетке  $\omega$  гистограммные значения  $h_i$  в узлах сетки  $\omega_t$ . Рассмотрим сетку  $\Omega_h = \omega_t \times \omega_y$ . Далее, используя гистограммы  $h_i$  несложно в центре каждой ячейке сетки  $\Omega_h$  восстановить среднее значение. Таким образом на сетке  $\Omega_h$  получаем кусочно-постоянную аппроксимацию, где  $(H_{ij})$  – матрица значений в ячейках сетки. Перепишем матрицу  $(H_{ij})$  в виде списка гистограммной функции распределения  $(h_j, p_j)$ , где  $h_j$  – гистограммы, образованные значениями  $H_{ij}$  при фиксированном  $j$ .

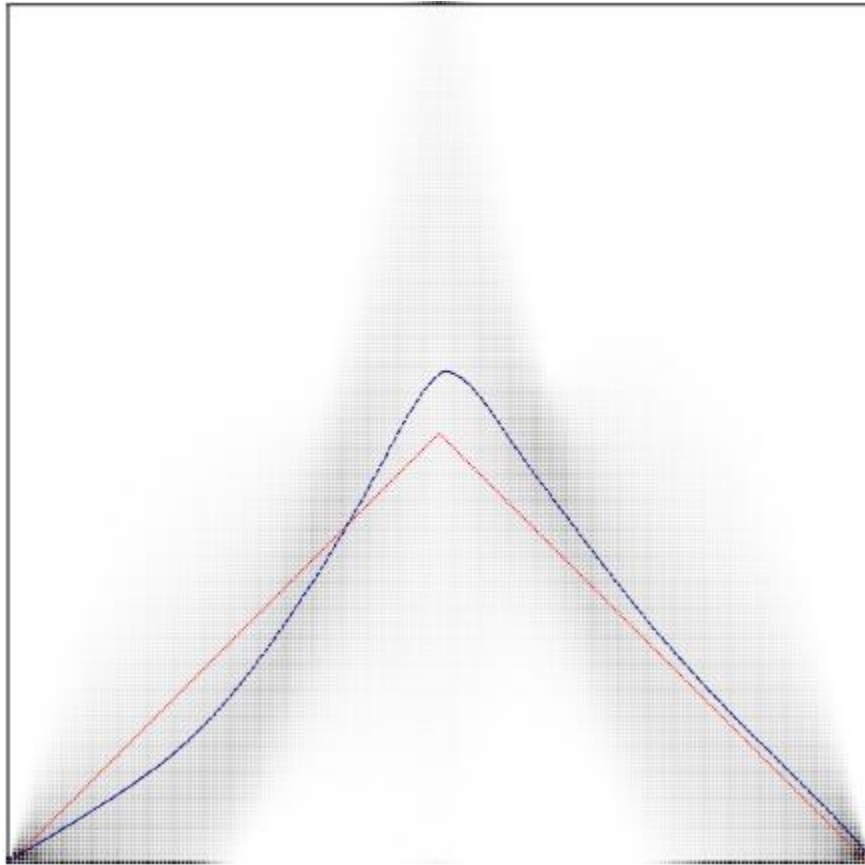


Рисунок 12 – Надежная оценка функции плотности вероятности

На рис. 12 представлена кусочно-постоянная аппроксимация надежной оценки эмпирической функции распределения, размерность выборки – 7. Оттенками серого представлены плотности вероятности, сплошной линией показана точная функция распределения, точечной линией – математические ожидания гистограмм  $h_j$ .

### 3.4 Выводы по главе 3

В третьей главе раскрыто понятие сплайна и порядковых статистик. Дается обоснование к выбору метода сглаживания эмпирической функции распределения при помощи сплайнов. Приводится результат сглаживания в виде гистограммных P-боксов. Анализируется результат и сравнивается с альтернативными методами построения гарантированных областей в условиях малых выборок.

Также приведены методы построения основных этапов построения доверительных интервалов и надежных оценок показателей функций



распределения, что показано на модельном примере, в котором были использованы тестовые данные.

#### **4 Разработка модуля построения гарантированных оценок показателей надежности ТСОИ**

Надёжность – это важный показатель любых технических систем. Её роль повышается, когда проектируется высокотехнологичное оборудование ответственного назначения [4,12]. Например, на этапе проектирования бортового оборудования космических аппаратов требуется гарантировать заданным техническим заданием срок его активного существования [5]. Такое оборудование не имеет серийного производства, и данные о потоке отказов достаточно ограничены. При этом повышение точности оценки *вероятности безотказной работы* (ВБР) – остаётся одной из важнейших проблем проектировщиков систем специального ответственного назначения.

Наиболее объективной информацией для определения характеристик надежности элементов является эксплуатационная информация, потому что она отражает специфику функционирования объектов, комплекс воздействующих факторов и другие особенности эксплуатации.

Применяемые традиционно параметрические методы анализа статистической информации в последнее время подвергаются критике специалистов по математической статистике. Дело в том, что использование параметрических методов требует предположений о виде закона распределения наблюдаемых случайных величин. Как правило, нельзя указать каких-либо веских причин, по которым конкретное распределение результатов наблюдений должно входить в то или иное параметрическое семейство. В настоящее время для решения задач анализа статистической информации развиваются непараметрические методы, в частности, методы ядерного оценивания.

В работе рассматриваются вопросы информационной поддержки надежности функционирования оборудования ответственного назначения в условиях ограниченного объема информации на основе непараметрического подхода с использованием численного вероятностного анализа. Предлагаются подходы к повышению достоверности оценок показателей

надежности и методы вычислений для оценки вероятности безотказной работы. Достоверность оценок обеспечивается построением вероятностных расширений для функции интенсивности отказов на основе применения процедур численного вероятностного анализа.

#### 4.1 Базовая методика оценки надёжности

Характеристики надёжности (включая значение ВБР) оборудования регламентируются требованиями технического задания на этапе проектирования аппаратуры [14]. Для организации расчетов требуется иметь: статистику (поток) отказов  $\xi_1, \xi_2, \dots, \xi_n$ , полученный эмпирическим путём в процессе эксплуатации аналогов или отдельных составных блоков; нормативные показатели надёжности каждого  $i$ -го элемента системы ( $P_i$ ), вычисляемого относительно экспоненциального априорного закона динамики отказов; топологические характеристики проектируемой аппаратуры, включая способы подключения резервных комплектов.

Далее производится оценка вероятности отказа оборудования для всего технического устройства  $P$ , например, в случае последовательного соединения блоков оценка может быть сведена к формуле (4.1)

$$P(t) = 1 - \prod_{i=1}^m P_i(t) \quad (4.1)$$

Подход к расчёту ВБР по известным формулам типа (4.1) считается классическим и его справедливость во многом подтверждается реальными апостериорными данными по эксплуатации оборудования [18]. Но если поставить задачу повышения точности прогноза значения ВБР, то сразу становятся очевидны следующие допущения: достоверность оценки  $P_i$  на прямую зависит от объёма исходных эмпирических данных о потоке отказов  $\{\xi\}$  для  $i$ -го элемента системы. При рассмотрении оборудования специального назначения только в исключительных случаях количество данных имеется в достаточном для анализа объёме.

Одним из способов повышения достоверности оценок показателей надежности является использование интервального анализа. В этом случае оценки показателей представляются в виде интервалов. Для оценки показателей надежности в этом случае может быть использована интервальная арифметика [17,19]. Например, по этому пути пошли в работе [14], предложив использовать для оценки показателей надёжности  $P(t)$  интервальные оценки. Заметим, что интервальные оценки ничего не говорят о специфике поведения исследуемой величины внутри интервала.

Рассмотрим другой подход, основанный на применении численного вероятностного анализа (ЧВА). Для этих целей необходимо преобразовать исходные данные в виде функций плотностей вероятности для функций  $P_i(t)$ . Численный вероятностный анализ имеет развитую арифметику для работы с плотностями вероятности и позволяет вычислять плотности вероятностей для широкого класса функций от случайных аргументов [2,16,22]. В этом случае исходным типом представления  $P_i(t)$  будет плотность вероятности, представленной различными способами, например в виде гистограммы, аналитической функции или в виде гистограмм второго порядка. Рассмотрим подробнее применение ЧВА и численной арифметики для оценки значения ВБР.

В тех случаях, когда это возможно, численные операции над плотностями вероятности случайных величин позволяют существенно поднять достоверность расчетов при сравнительно небольшом объеме вычислений.

Обратим внимание на одну особенность решения задач по определению характеристик надежности. Наблюдаемой случайной величиной в рассматриваемой задаче является время (наработка до отказа). Область определения данного параметра  $[0, \infty)$ . Следовательно, априори известно, что обрабатываемые случайные величины относятся к классу

неотрицательных величин. Ниже определим основные показатели надежности.

*Вероятность безотказной работы*  $P(t)$  – вероятность того, что в пределах заданной наработки отказ не возникает (*наработка* – это продолжительность или объем работы):

$$P(t) = P(T \geq t), \quad (4.2)$$

где  $T$  – случайное время работы объекта до отказа;  $t$  – заданная наработка. Этот показатель обладает следующими свойствами:  $P(t) = 1$ , т.е. до начала работы ( $t = 0$ ) система являлась, безусловно, работоспособной;  $P(t)$  – невозрастающая функция времени;

$\lim_{t \rightarrow \infty} P(t) = 0$ , т.е. объект не может сохранять свою работоспособность неограниченно долго.

*Вероятность отказа*  $Q(t)$  – вероятность того, что в пределах заданной наработки отказ объекта возникает:

$$Q(t) = 1 - P(t). \quad (4.3)$$

Она характеризует вероятность того, что случайное время  $t$  работы объекта до отказа меньше заданного времени,  $T < t$ . Под  $T$  понимается непрерывная случайная величина, для которой существует плотность распределения наработки до отказа:

*Интенсивность отказов* (ИО) – это отношение условной вероятности того, что случайная наработка до отказа будет принимать значения из полуинтервала  $[t; t + \Delta t)$  бесконечно малой длины  $\Delta t$  при условии, что отказа до момента времени  $t$  не было, к длине этого полуинтервала  $\Delta t$ :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq \xi < t + \Delta t | t \leq \xi)}{\Delta t} = \frac{f(t)}{P(t)}, \quad (4.4)$$

Заметим, что  $f(t) = P'(t)$

$$\lambda(t) = -\frac{P'(t)}{P(t)}. \quad (4.5)$$

Интенсивность отказов показывает, какая часть элементов выходит из строя в единицу времени по отношению к среднему числу исправно работающих элементов.

Решая соотношение (4.5) как линейное однородное дифференциальное уравнение первого порядка относительно функции безотказности, получаем

$$P(t) = e^{\int_0^t -\lambda(t)dt}. \quad (4.6)$$

В частном случае, когда  $\lambda = \text{const}$ , выражение (4.6) представляет собой экспоненциальный закон надежности. По этому закону вероятность безотказной работы элементов, обладающих интенсивностью отказов, убывает со временем по экспоненциальной кривой.

Рассмотрим методы вычисления показателей надежности. В случае достаточного объема выборок для построения  $f_{\xi}(t)$  используют гистограммы. Соответственно далее можно использовать в полном объеме численный вероятностный анализ. Однако в случае уникального оборудования объем выборок весьма не велик  $n \ll O(1)$  и часто  $n < 10$ . В этих случаях использование гистограмм весьма ограничено, поэтому в качестве альтернативы применяют ядерные оценки, которые впервые были введены в работах Парзена [6] и Розенблатта [7].

Приведем формулу для построения ядерной оценки плотности распределения непрерывной случайной величины

$$f_n(t) = \frac{1}{n} \sum_{i=1}^n V(t - \xi_i), \quad (4.7)$$

где  $V$  ядро. Приведем в качестве  $V$  гауссовское ядро

$$V(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-t^2/(2\sigma^2)), \quad (4.8)$$

где  $n$  – размер выборки,  $\sigma$  – параметр.

Заметим, что при малых  $n$  ядерные оценки часто дают большие погрешности.

С целью повышения достоверности оценок показателей надежности рассмотрим подход, основанный на сглаживании эмпирической функции распределения.

Пусть перед нами стоит задача оценивания функции распределения одномерной непрерывной случайной величины  $X$  по выборке  $x_1, x_2, \dots, x_n$ . Не ограничивая общности, будем считать, что последовательность  $x_i$  монотонно возрастающая. Поскольку функция распределения  $F(x)$  определяется как вероятность события  $(X \leq x)$ , то оценка этой вероятности будет оценкой для  $F(x)$  в точке  $x$ . Известно, что оценкой вероятности событий является его относительная частота. В данном случае число благоприятных исходов равно числу выборочных значений, не превышающих заданную величину  $x$ . Можно записать:

$$F_n = P(X \leq x) = \frac{m_x}{n} \quad (4.9)$$

Если теперь рассматривать  $F_n$  как функцию  $x$  то формула (4.9) является оценкой функции распределения  $F(x)$ . Выражение (4.9) называется *эмпирической функцией распределения*. Важно отметить, что эмпирическая функция распределения проходит через точки  $(x_i, i/n)$ .

Поскольку эмпирическая функция распределения не является непрерывно дифференцируемой, то функцию плотности вероятности получить непосредственно дифференцированием нельзя. Для этих целей произведем сглаживание эмпирической функции распределения. Для сглаживания будем использовать кубические сплайны Эрмита  $s$ . Функционал для сглаживания можно записать в виде

$$\Phi(z) = \sum_{i=1}^n (s(x_i) - i/n)^2 + \alpha \|s''\|^2, \quad (4.10)$$

где  $\alpha$  – параметр сглаживания,  $\|\cdot\|$  – евклидова норма.

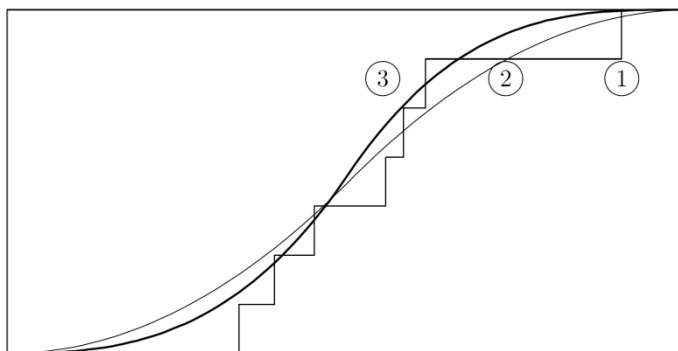


Рисунок 13 – Сглаживание эмпирической функции распределения

На рис. 13 представлен пример сглаживания эмпирической функции распределения по выборке  $x_1, x_2, \dots, x_n$ ,  $n=7$ , распределенной по треугольному закону на отрезке  $[0,2]$ , с вершиной в точке  $(1,1)$ . Цифрами представлены соответственно: 3 – эрмитов сплайн, представляющий сглаженную функцию распределения, 2 – точная функция распределения, 1 – эмпирическая функция распределения.

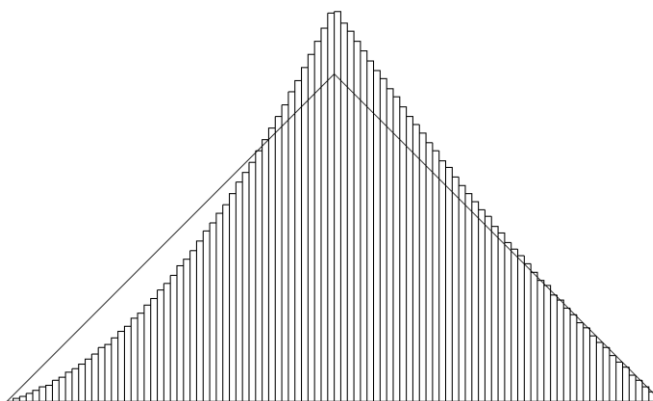


Рисунок 14 – Гистограмма, аппроксимирующая функцию плотности вероятности и точная функции плотности

На рис. 14 представлена гистограмма, построенной по сглаженной функции распределения, и точная функции плотности вероятности.

Используя данный подход и ЧВА, можно построить оценку для интенсивности отказов  $\lambda(t)$ . Для этого рассмотрим основные элементы ЧВА.



## 4.2 Оценки интенсивности отказов

Для оценки интенсивности отказов используются методы численного моделирования. В случае модели надежности (6) необходимо определять функцию  $\lambda(t)$ .

Пусть, вектор  $(\xi_1, \xi_2, \dots, \xi_n)$  – поток отказов, полученный эмпирическим путём в процессе эксплуатации аналога или отдельного блока оборудования ответственного назначения. Тогда для  $\lambda(t)$  имеет место соотношение

$$-\ln(z_i) = \int_0^{\xi_i} \lambda(\xi) d\xi, \quad (4.11)$$

где  $z_i = P(\xi_i)$ . Поскольку истинные значения  $z_i$  не известны, для построения оценок интенсивности отказов заменим  $z_i = P(\xi_i)$  их математическими ожиданиями  $M[z_i]$ . Заметим, в векторе  $(z_1, z_2, \dots, z_n)$ , все компоненты  $z_i$  – равномерно распределенные величины на отрезке  $[0,1]$  и  $z_1 < z_2 < \dots < z_n$ . В этом случае  $M[z_i] = i/(n+1)$  и точки  $(\xi_i, i/(n+1))$ ,  $i = 1, \dots, n$  лежат на эмпирической функции распределения.

Для нахождения оценок  $\lambda(t)$  будем использовать метод наименьших квадратов, суть которого состоит в следующем: пусть  $\varphi_1, \varphi_2, \dots, \varphi_m$  – линейно независимые функции. Тогда

$$\Phi(a_1, \dots, a_m) = \sum_{i=1}^n (z_i - \sum_{j=1}^m a_j \varphi_j(\xi_i))^2 \rightarrow \min. \quad (4.12)$$

$$A\vec{a} = b \quad (4.13)$$

где  $A = (a_{ij})$  – матрица Грама,  $b = (b_i)$   $a_{ij} = (\varphi_i, \varphi_j)$   $b_i = (z, \varphi_i)$  и

$$(x, y) = \sum_{i=1}^n x(\xi_i) y(\xi_i). \quad (4.14)$$

Используя вместо  $z_i$  их совместные функции плотности вероятности, можно построить вероятностные расширения для  $\lambda(t)$ . Пусть  $p_z$  – совместная плотность распределения вектора  $(z_1, z_2, \dots, z_n)$ , тогда заменяя  $z$  их совместной плотностью вероятности. Далее используя численный вероятностный анализ, строим оценки плотности вероятности для компонент вектора  $\vec{a}$ , в результате, получаем вероятностное расширение  $\lambda(t)$ .

### 4.3 Модельный пример

Пусть дана повторная выборка наработок на отказ  $\{0.0155, 0.0389, 0.2855, 0.5318, 0.7412, 1.0118, 1.1267, 1.2327, 1.8594\}$ . Предположим  $\lambda(t)$  имеет вид  $\lambda(t) = a_0 + a_2 x^2$  и

$$P(t) = \exp\left(-\int_0^t a_0 + a_2 x^2 dx\right), \quad (4.15)$$

где  $a_0 = 1$ ,  $a_2 = 0.3$ . Используя метод наименьших квадратов и полагая  $z_i = i / (n+1)$ ,  $\varphi_1 = 1$ ,  $\varphi_2 = x^2$ , были получены следующие оценки  $a_0 = 0.95$ ,  $a_2 = 3.23$ . Если заменить  $z_i$  на их совместную функцию плотности вероятности и использовать численный вероятностный анализ, получаем гистограммные оценки функции плотности вероятности для  $a_0, a_2$  рис. 15, 16.

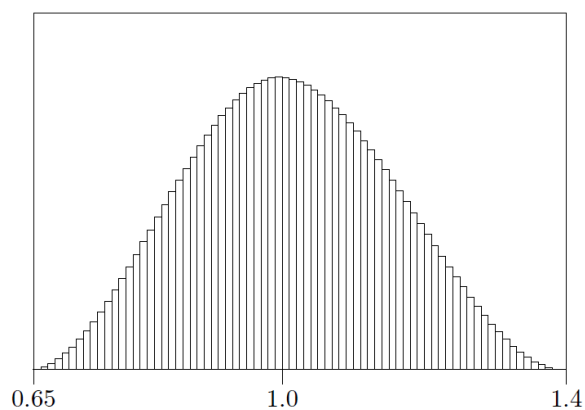


Рисунок 15 – Гистограммное представление оценки  $a_0$

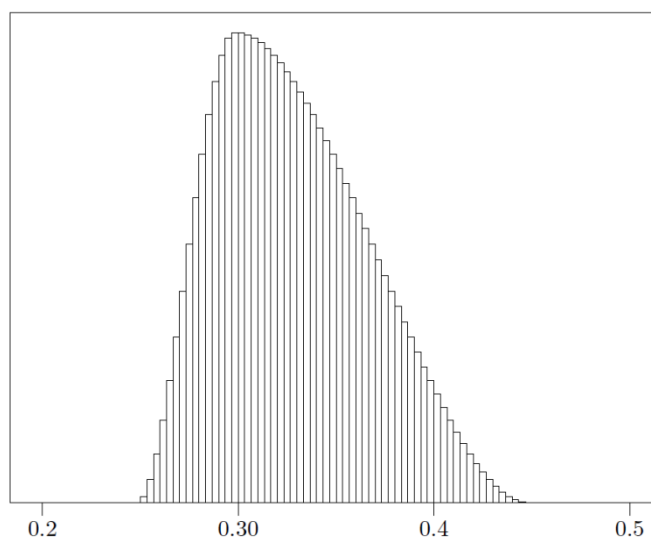


Рисунок 16 – Гистограммное представление оценки  $a_2$

Полученные оценки плотности вероятности для  $a_0, a_2$  и вероятностные расширения, можно вычислить оценку плотности вероятности для случайной функции  $P(t)$ .

Подставляя в выражение для  $\lambda(t)$  гистограммные оценки коэффициентов  $a_0, a_2$  можно получить все множество возможных значений интенсивности отказов  $\lambda(t)$ . Используя ЧВА можем вычислить гистограммную оценку значений  $P(t)$  в любой момент времени.

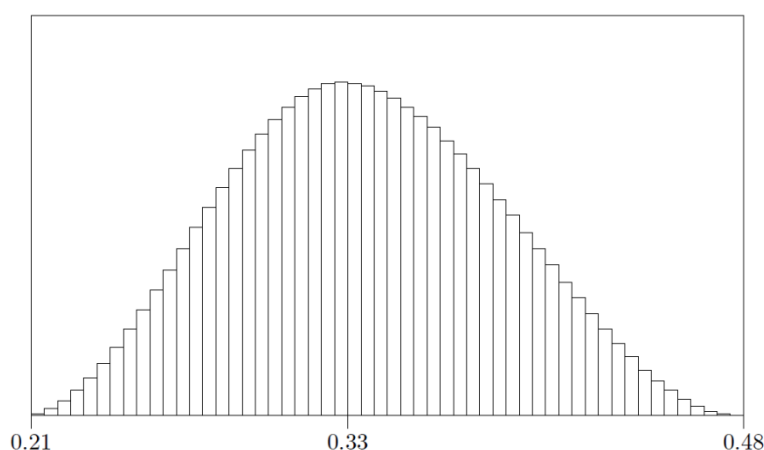


Рисунок 17 – Гистограммное представление оценки  $P(t)$

На рис. 17 приведена гистограммная оценка  $P(t)$  в момент времени  $t = 1$ .

#### **4.4 Информационно - аналитический модуль построения гарантированных оценок показателей надежности**

На основе численного вероятностного анализа был разработан информационно – аналитический модуль, который состоит из следующих блоков:

- блок обработки и представления данных;
- блок численного вероятностного анализа;
- блок оценки показателей надежности;
- блок визуализации.

Входные данные, представляют собой измерения исследуемых параметров и показателей функционирования ТСОИ. Входные данные представляют собой выборку, состоящую из значений величины или параметра (например, интенсивность отказов, вероятность безотказной работы, данные проектной документации по значениям параметров, и т.д.)

Блок численно вероятностного анализа включает в себя математические алгоритмы, включая способы представления данных в виде гистограмм, гистограмм второго порядка, кусочно – линейных функций и сплайнов, а также процедуры численной реализации арифметических операций, которые необходимы для дальнейшего вычисления показателей надежности; процедуры построения эмпирических функций распределения и процедуры построения случайных сплайнов.

Блок оценки показателей надежности включает в себя базовую методику оценки надежности показателей функционирования объекта исследования, процедуры построения гарантированных областей значений показателей надежности по эмпирическим данным

Блок визуализации включает в себя –представление гарантированной области значений показателей надежности в виде визуально – интерактивной модели с возможностью анализа информации

Предполагается интеграция данного модуля с другими внешними компьютерными системами, такими как ПАП DEDUCTOR, ППП SPSS и среда программирования R, для проведения статистических расчетов, вычисления и представления порядковых статистик и решения других задач моделирования и анализа данных.

## ЗАКЛЮЧЕНИЕ

В диссертации рассматривается актуальная проблема повышения надежности функционирования ТСОИ в условиях неопределенности и ограниченного объема информации на основе построения гарантированных оценок показателей надежности. Для этого применяется численный подход к построению гарантированных оценок показателей надежности на основе численного вероятностного анализа. Для этих целей необходимо преобразовать исходные данные в виде функций плотностей вероятности для функций  $P_i(t)$ . Численный вероятностный анализ имеет развитую арифметику для работы с плотностями вероятности и позволяет вычислять плотности вероятностей для широкого класса функций от случайных аргументов. В этом случае исходным типом представления  $P_i(t)$  будет плотность вероятности, представленной различными способами, например в виде гистограммы, аналитической функции или в виде гистограмм второго порядка. В тех случаях, когда это возможно, численные операции над плотностями вероятности случайных величин позволяют существенно поднять достоверность расчетов при сравнительно небольшом объеме вычислений [21-23]. Данная методика оценки интенсивности отказов оборудования ответственного назначения, позволяет строить достоверные оценки показателей надежности сложных технических систем в условиях малых выборок. Достоверность оценок обеспечивается построением вероятностных расширений для  $\lambda(t)$ . Предложенный подход может использоваться для оценки различных рисков в сложных технических системах в условиях ограниченного объема информации. В диссертации разработан алгоритм оценки показателей надежности, который включает такие этапы, как предварительный анализ и статистическая обработка экспериментальных данных. где изучается характер имеющейся неопределенности и далее определяется способ представления данных; выбор соответствующего метода вычисления показателей надежности.

основанный на *сглаживании эмпирической функции распределения*; анализ результатов и использование визуально-интерактивного моделирования (ВИМ). В рамках ВИМ реализуется графическое представление полученных результатов, возможность интерактивной настройки модели. Данный подход для оценки интенсивности отказов оборудования ответственного назначения, позволяет строить достоверные оценки показателей надежности сложных технических систем в условиях малых выборок. Достоверность получаемых оценок показателей надежности обеспечивается процедурами вычисления их вероятных расширений и алгоритмами построения эмпирических законов функций распределения. Предложенный подход может использоваться для оценки рисков в различных сложных технических системах в условиях ограниченного объема информации и в условиях ответственного функционирования.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1) Becker P., Jensen F. Design of systems and circuits for maximum Reliability or maximum production yield. New York: McGraw-Hill. 1977.
- 2) Dobronets B.S., Popova O.A. Numerical probabilistic analysis under aleatory and epistemic uncertainty. // *Reliable Computing*, 2014. Vol. 19, P. 274–289.
- 3) Dobronets B., Popova O. Numerical probabilistic approach for data nonparametric analysis// *Applied methods of statistical analysis. Nonparametric approach proceedings of the international workshop*. 2015. С. 376-384.
- 4) Uglev V.A., Popova O.A., Dobronets B.S. The accuracy calculation control of reliability indices for equipment responsible appointment// *2015 International Siberian Conference on Control and Communications, SIBCON 2015 - Proceedings 2015*. С. 7147248.
- 5) Husmeier D. *Neural Networks for Conditional Probability Estimation*. – London: Springer-Verlag, 1999.
- 6) MIL-HDBK-217F. *Military handbook. Reliability prediction of electronic equipment*, 1995. – 150 p.
- 7) NASA Technical Standard (NASA-STD-8729.1) *Planning, developing and managing an effective Reliability and Maintainability (R&M) program*, 1998. – 77 p.
- 8) Parzen E. On estimation of a probability density function and mode // *Annals of Mathematical Statistics*. – 1962. – №33. – P. 1065–1076.
- 9) Rozenblatt M. Remark on some nonparametric estimates of a density function // *Annals of Mathematical Statistics*. – 1956. – №27. – P. 832–837.
- 10) Usher J.S., Kamal A.H., and Syed W.H. Cost optimal preventive maintenance and replacement scheduling. // *ИЕ Trans.* – 1998. – №30. – P.1121-1128.
- 11) Williams J.H., Davies A., and Drake P.R. *Condition-Based Maintenance and Machine Diagnostics*. – London: Chapman and Hall, 1998.
- 12) Абрамов, О. В. К проблеме предотвращения аварий технических объектов ответственного назначения / О. В. Абрамов // *Надежность и качество сложных систем*. – 2013. – № 1. – С. 11–16.
- 13) Абрамов О.В. Мониторинг и прогнозирование технического состояния систем ответственного назначения // *Информатика и системы управления*. 2011. – № 2. С. 4–15.
- 14) Абрамов О.В., Розенбаум А.Н. *Управление эксплуатацией систем ответственного назначения*. – Владивосток: Дальнаука, 2000.



- 15) Альберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. М.: Мир, 1972.
- 16) Антонов А.В., Малковик К.Н., Чумаков И.А. Интервальная оценка характеристик надёжности уникального оборудования // Фундаментальные исследования. – №12. 2011. С. 71–76.
- 17) Берман А.Ф., Николайчук О.А. Пространство технических состояний уникальных механических систем // Проблемы машиностроения и надёжности машин. – 2007. – № 1. – С.14-22.
- 18) Герасимов В.А., Добронев Б.С., Шустров М.Ю. Численные операции гистограммной арифметики и их применения // АиТ. 1991. №2. С. 83—88.
- 19) Добронев Б. С. Интервальная математика. Красноярск: КГУ, 2004. 216 с.
- 20) Добронев Б.С. Надёжность информационных систем Электронный ресурс. : учеб.-метод. пособие Сиб. федерал. ун-т. – Красноярск : СФУ, 2012. – 159 с.
- 21) Добронев Б. С. Приближения множеств решений параметрическими множествами // Журнал Сибирского федерального университета. Серия: Математика и физика. 2009. Т. 2. № 3. С. 305-311.
- 22) Добронев Б.С., Попова О.А. Представление и обработка неопределённости на основе гистограммных функций распределения и r-boxes // Информатизация и связь. 2014. № 2. С. 23-26.
- 23) Добронев Б.С., Попова О.А. Численные операции над случайными величинами и их приложения // Журнал Сибирского федерального университета. Серия: Математика и физика. 2011. Т. 4. № 2. С. 229—239.
- 24) Добронев Б.С., Попова О.А. Численный вероятностный анализ неопределённых данных: монография. Красноярск: Сиб. федер. ун-т, 2014. 167 стр.
- 25) Добронев Б.С., Попова О.А. Элементы численного вероятностного анализа // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. 2012. № 2. С. 19–23.
- 26) Добронев Б.С., Попова О.А., Углев В.А. Методика построения гарантированных оценок показателей надёжности для технических систем ответственного назначения // Безопасность и живучесть технических систем Материалы и доклады: в 3-х томах. 2015. С. 154-159.
- 27) Дэйвид Г. Порядковые статистики. М.: Наука. 1979, 336 стр.
- 28) Карлин С., Стадден В. Чебышевские системы и их применение в анализе и статистике/ пер. с англ., под ред. С.М. Ермакова. – М.: Наука, 1976.

29) Максимов Ю.Д. Математика. Теория и практика по математической статистике. Конспект-справочник по теории вероятностей.: Учеб.пособие / Ю.Д. Максимов; под ред. В.И. Антонова. – СПб.: Изд-во Политехнического ун-та, 2009. – 395 с.

30) Положинцев Б.И. Введение в математическую статистику: Учеб. пособие. – СПб., Изд-во Политехнического ун-та 1994 .– 56с.

31) Пооль Д.Е., Обработка данных на основе использования сплайнов для построения гистограммных функций распределения – Новосибирск: Научные исследования и разработки молодых ученых, 2015, с. 124 – 128.

32) Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере/ под ред. Фигурнова В.Э. –М.: ИНФРА – М. 1998. – 528 с., ил.

33) Хайкин С. Нейронные сети: Полный курс/ пер. с англ. – М.: ООО «И.Д. Вильямс», 2006.

## ПРИЛОЖЕНИЕ А

### Плакаты презентации

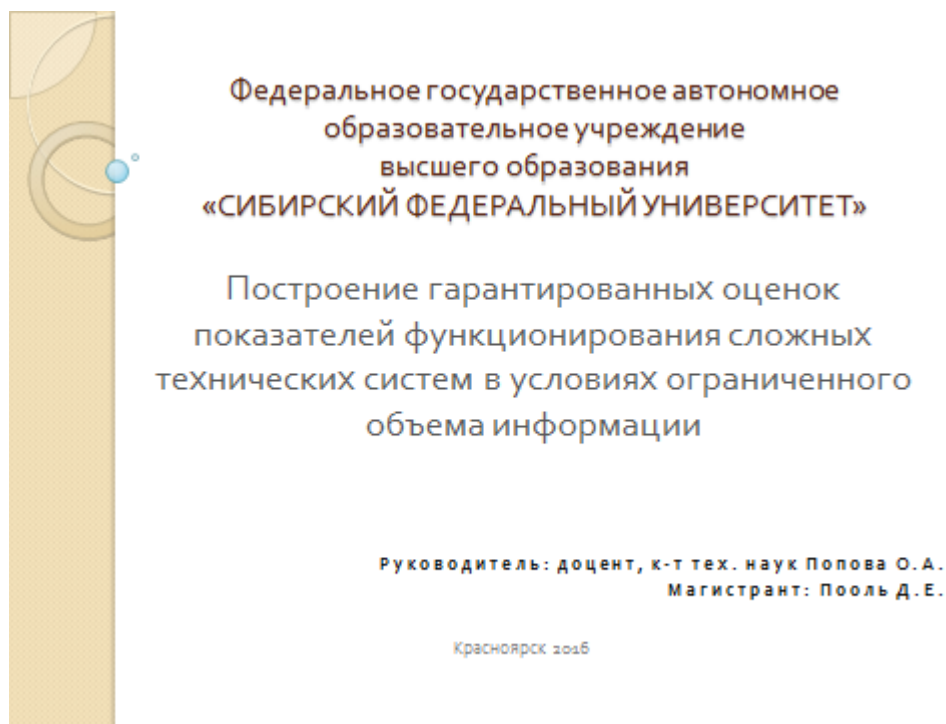


Рисунок А.1 – Плакат презентации №1

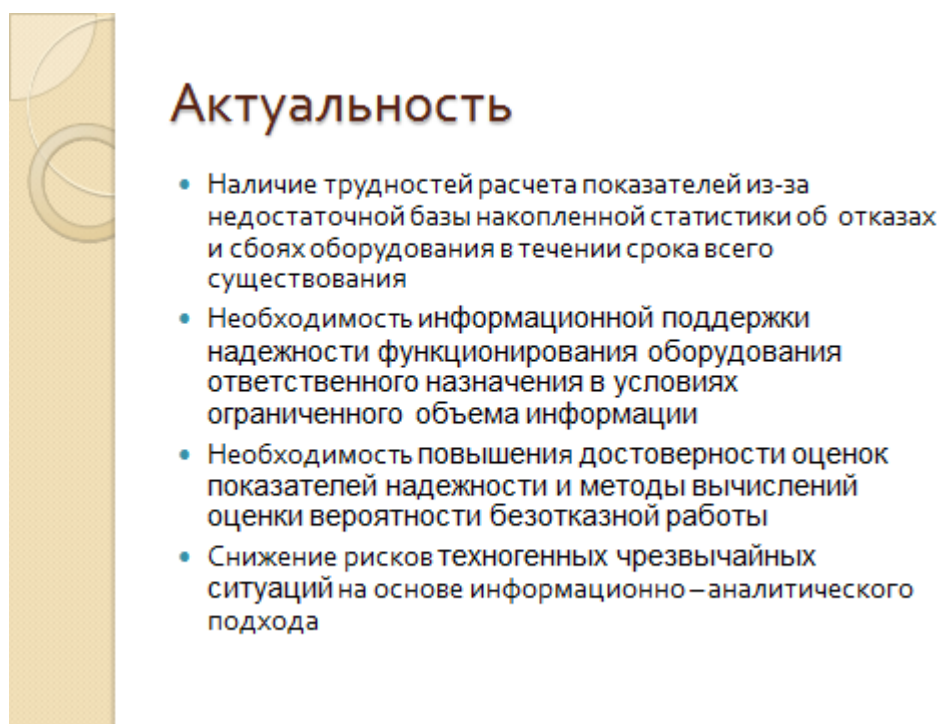
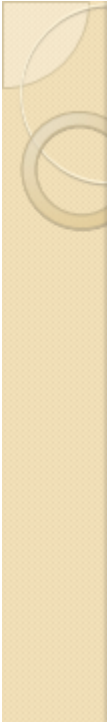


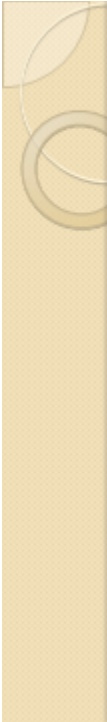
Рисунок А.2 – Плакат презентации №2



## Цель

- Повышение эффективности анализа надежности сложных технических систем ответственного назначения на основе построения гарантированных оценок показателей функционирования в условиях неопределенности и ограниченного объема информации

Рисунок А.3 – Плакат презентации №3



## Задачи

1. Провести проблемный анализ по теме исследования на основе изучения научных публикаций
2. Изучить теоретические основы численного вероятностного анализа и применить его методы для построения численных оценок показателей надежности
3. Разработать алгоритм построения эмпирической функции распределения с использованием сглаживающего сплайна
4. Провести анализ численных оценок показателей надежности в условиях ограниченного объема информации
5. Разработать модуль для построения гарантированных оценок и показателей надежности

Рисунок А.4 – Плакат презентации №4

## Информационно – аналитический модуль

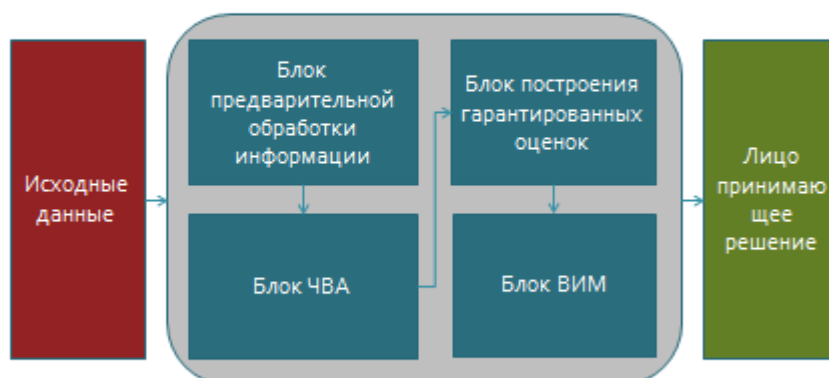


Рисунок А.5 – Плакат презентации №5

## Процедуры построения эмпирических ФРВ и ФПВ

Пусть перед нами стоит задача оценивания функции распределения одномерной непрерывной случайной величины  $X$  по выборке  $x_1, x_2, \dots, x_n$ .

Не ограничивая общности будем считать, что последовательность  $x_i$  монотонно возрастающая.

Поскольку функция распределения  $F(x)$  определяется как вероятность события  $(X \leq x)$ , то оценка этой вероятности будет оценкой для  $F(x)$  в точке  $x$ .

Известно, что оценкой вероятности событий является его относительная частота. В данном случае число «благоприятных исходов»  $m_x$  равно числу выборочных значений, не превышающих заданную величину. Можно записать:

$$F_n = P(X \leq x) = \frac{m_x}{n} \quad (2)$$

Рисунок А.6 – Плакат презентации №6

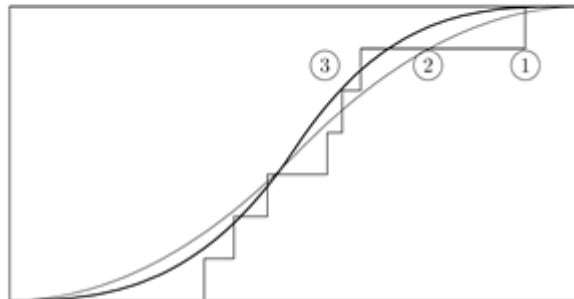
## Процедуры сглаживания эмпирической функций распределения сплайнами.

Пусть  $\omega = \{a = x_0 < x_1 < x_2 < \dots < x_n < b = x_{n+1}\}$  – сетка,  $h = O(1/n)$ .  
Проведем через точки  $(x_i, z_i)$  на сетке  $\omega$  кубический сплайн  $S$ .  
Тогда

$$\|F - S\| \leq Kh^4 \|F^{(4)}\| = O\left(\frac{1}{n^4}\right).$$

Рисунок А.7 – Плакат презентации №7

## Процедуры аппроксимации функций распределения кубическими сплайнами



На рисунке представлен пример аппроксимации функции распределения по некоторой выборке, где:

- 1 — эмпирическая функция распределения,
- 2 — точная функция распределения,
- 3 — сплайн, аппроксимирующий функцию распределения

Рисунок А.8 – Плакат презентации №8

## Порядковые статистики

$(x_1, \dots, x_n)$  – повторная выборка с функцией распределения  $F$ .  
Пусть

$$z_i = F(x_i), i = 1, \dots, n.$$

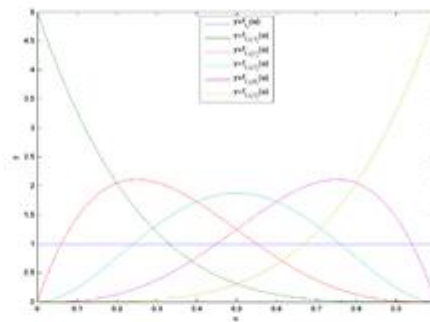
Заметим, что  $z_i, i = 1, \dots, n$  – равномерно распределенные случайные переменные  $z_i \in [0, 1]$ .

Если  $z_1 \leq z_2 \leq \dots \leq z_n$ , то  $\{z_i\}$  – образуют порядковую статистику и  $M[z_i] = i/(n+1)$ .

Рисунок А.9 – Плакат презентации №9

## Порядковые статистики

На рисунке представлены плотности стандартного непрерывного равномерного распределения и его порядковых статистик для случая  $n=5$



Плотность распределения  $k$ -й порядковой статистики

$$p_k(z) = \frac{n!}{(n-k)!(k-1)!} z^{k-1} (1-z)^{n-k}, z \in [0, 1].$$

Рисунок А.20 – Плакат презентации №10

## Процедуры построения сплайнов

- Задача построения сплайнов сводится к построению и решению алгебраических уравнений с тридиагональными матрицами
- Матрицы этих систем детерминированные, а правые части содержат случайные величины. Таким образом, в силу детерминированности матриц, решение может быть представлено как линейная комбинация элементов правых частей.

$$\lambda_j m_{j-1} + 2m_j + \mu_j m_{j+1} = d_j,$$

$$2m_0 + m_1 = 3(z_1 - z_0)/h_1 - h_1 z_0^2/2,$$

$$2m_N + m_{N-1} = 3(z_N - z_{N-1})/h_N + h_N z_N^2/2,$$

$$d_j = 3\lambda_j(z_j - z_{j-1})/h_j + 3\mu_j(z_{j+1} - z_j)/h_{j+1}, \quad j = 1, \dots, N-1.$$

где  $m_j = s'(x_j)$ .

На отрезке  $[x_{j-1}, x_j], j = 1, \dots, N$  сплайн может быть представлен в виде:

$$s(x) = m_{j-1}(x_j - x)^2(x - x_{j-1})/h_j^2 - m_j(x - x_{j-1})^2(x_j - x)/h_j^2 + z_{j-1}(x_j - x)^2(2(x - x_{j-1}) + h_j)/h_j^3 + z_j(x - x_{j-1})^2(2(x_j - x) + h_j)/h_j^3,$$

$$\vec{m} = B\vec{z}.$$

Рисунок А.31 – Плакат презентации №11

## Процедуры построения сплайнов

Поскольку коэффициенты  $m_j$  можно линейно выразить через  $z_j$

$$\vec{m} = B\vec{z},$$

то на отрезке  $[x_{j-1}, x_j], j = 1, \dots, N$  сплайн может быть представлен в виде

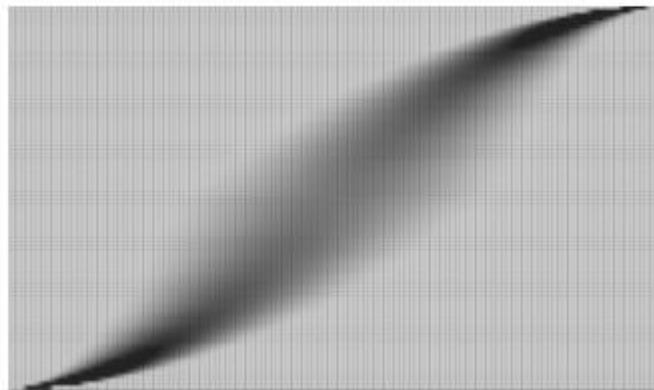
$$s(x) = \sum_{i=1}^N z_i \psi_i^j(x), \quad j = 1, \dots, N,$$

где  $\psi_i^j(x)$  – некоторые вещественные функции. Для  $\{z_i\}$  известна совместная функция плотности вероятности  $\rho(z_1, \dots, z_n)$  (свойство порядковых статистик). Используя численный вероятностный анализ, можно построить вероятностное расширение для сплайна  $s$ , что позволит построить гистограммную функцию распределения.

Рисунок А.42 – Плакат презентации №12



## Процедуры построения гистограммных функций плотности вероятностей (Р-боксов)



На рисунке показана гистограммная функция распределения, где значения гистограмм представлены оттенками серого.

Рисунок А.53 – Плакат презентации №13

## Процедуры построения гарантированных оценок интенсивности отказов

$$P(t) = e^{-\int_0^t \lambda(t) dt}$$

Пусть,  $\xi_1, \xi_2, \dots, \xi_n$  – поток отказов, полученный эмпирическим путём в процессе эксплуатации аналогов или отдельных составных блоков. Имеет место равенство

$$-\ln(z_i) = \int_0^{\xi_i} \lambda(t) dt,$$

где  $z_i = P(\xi_i)$ .

Поскольку истинные значения  $z_i$  не известны, их заменяют математическими ожиданиями  $Mz_i$  –  $i$ -компоненты в векторе  $(z_1, z_2, \dots, z_n)$ , все  $z_i$  – равномерно распределённые величины на  $[0, 1]$  и  $z_1 < z_2 < \dots < z_n$ . В этом случае  $Mz_i = i/(n+1)$ . Точки  $(\xi_i, i/(n+1))$ ,  $i = 1, \dots, n$  лежат на эмпирической функции распределения.

Используя сплайны можно построить аппроксимацию для  $\lambda(t)$ .

$$\lambda(t) = \sum_{i=1}^n a_i \varphi_i(t). \quad \Phi(a_1, \dots, a_n) = \sum_{i=1}^n (z_i - \sum_{j=1}^n a_j \varphi_j(\xi_i))^2 \rightarrow \min. \quad A \bar{a} = b,$$

Рисунок А.64 – Плакат презентации №14

## Пример

Пусть дана повторная выборка наработок на отказ  $\{0.0155, 0.0389, 0.2855, 0.5318, 0.7412, 1.0118, 1.1267, 1.2327, 1.8594\}$ , и последовательность имеет функцию распределения

$$P(t) = \exp\left(-\int_0^t a_0 + a_2 x^2 dx\right),$$

где  $a_0 = 1$ ,  $a_2 = 0.3$ .

Используя метод наименьших квадратов и полагая  $Z_i = i/(n+1)$ ,  $\varphi_1 = 1$ ,  $\varphi_2 = x^2$ , были получены следующие оценки  $a_0 = 0.95$ ,  $a_2 = 3.23$ .

Рисунок А.75 – Плакат презентации №15

## Надежная оценка функции плотности вероятности

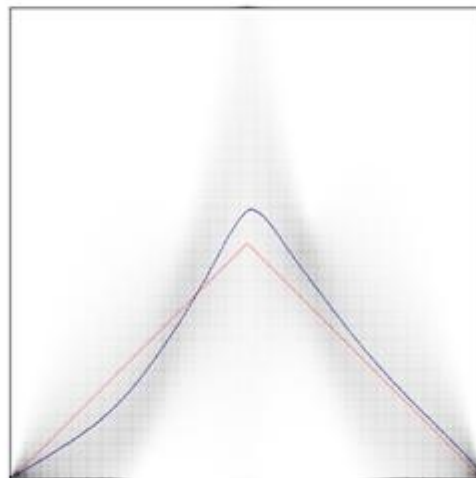


Рисунок А.86 – Плакат презентации №16

## ПРИЛОЖЕНИЕ Б

### Исходный программный код

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using ZedGraph;

namespace GraphicCore
{
    public partial class Form1 : Form
    {
        int[] Gis = new int[0];

        public Form1()
        {
            InitializeComponent();
        }

        private void button1_Click(object sender, EventArgs e)
        {
            Gis = buildArray2();
            num_Zi.Maximum = numForm.Value;
        }

        private void drawZedGraph(int[] masGis)
        {
            GraphPane pane = zedGraph.GraphPane;

            pane.CurveList.Clear();

            double[] values = new double[masGis.Length];
            for (int i = 0; i < masGis.Length; i++)
            {
                values[i] = masGis[i];
            }

           BarItem bar = pane.AddBar("Гистограмма", null, values, Color.Blue);
```

```

pane.BarSettings.MinClusterGap = 0.0f;

zedGraph.AxisChange();
zedGraph.Invalidate();

}

private void drawGistogramm(int[] masGis)
{
    pictureBox.Refresh();

    pictureBox.Image = new Bitmap(pictureBox.Width, pictureBox.Height);
    Graphics g = Graphics.FromImage(pictureBox.Image);
    Pen blackPen_0 = new Pen(Brushes.Black, 1);
    Pen redPen_0 = new Pen(Brushes.Red, 1);
    Rectangle[] columnsGis = new Rectangle[masGis.Length];
    Point[] massPointGis = new Point[masGis.Length];

    g.TranslateTransform(1, pictureBox.Width - 1);
    g.ScaleTransform(1, -1);
    int maxValue = masGis.Max();
    int weigtColumn = pictureBox.Width / masGis.Length;

    for (int i = 0; i < masGis.Length; i++)
    {
        columnsGis[i] = new Rectangle(new Point(weigtColumn * i, 105), new
        Size(weigtColumn, ((pictureBox.Height) / maxValue) * masGis[i]));
        massPointGis[i] = new Point(weigtColumn * i + weigtColumn / 2,
        ((pictureBox.Height) / maxValue) * masGis[i] + 105);
    }

    g.DrawRectangles(blackPen_0, columnsGis);
    g.DrawCurve(redPen_0, massPointGis, 1.0f);
}

public int[] buildArray2()
{
    int num = Convert.ToInt32(numForm.Value); // размерность массива zi
    int accuracy = Convert.ToInt32(accuracyForm.Value); //точность
    ВЫЧИСЛЕНИЙ
    int numMass = Convert.ToInt32(numMassForm.Value); // число опытов
    double numStolb = Convert.ToDouble(numStolbForm.Value); // кол-во
    столбцов гистограммы
}

```

```

double[] zi = new double[num];
double[] xi = new double[num];
double[,] Mass_xi = new double[num, numMass]; // здесь храним все
ОПЫТЫ
double[,] Mass_zi = new double[num, numMass];
int[] gistogramm = new int[Convert.ToInt32(numStolb) + 1];
int[] gistogramm_zi = new int[Convert.ToInt32(numStolb) + 1];

Random rand = new Random();
double tempRand;
int tempAccuracy = Convert.ToInt32(Math.Pow(10,
Convert.ToDouble(accuracy)));
int tempValue;

double TVMin = 0; // максимальное значение xi для отрезка zi[0, 0.5]
double TV = 50; // минимальное значение xi для отрезка zi[0.5, 1]

for (int nM = 0; nM < numMass; nM++)
{
    for (int i = 0; i < zi.Length; i++) // заполняем массив zi дробными
значениями от 0 до 1
    {
        if (checkBox_Random.Checked == true)
        {
            tempRand = Convert.ToDouble(rand.Next(tempAccuracy)) /
tempAccuracy;
            zi[i] = tempRand;
        }
        else
            zi[i] = ((1.0 / num) * i) + (Convert.ToDouble(nM) /
Convert.ToDouble(numMass * 10));
    }

    Array.Sort(zi);

    tempValue =
Convert.ToInt32(Math.Round(zi[Convert.ToInt32(num_Zi.Value) - 1] / (1 /
numStolb)));
    if (tempValue >= numStolb) { tempValue -= 1; }
    gistogramm_zi[tempValue] += 1;
}

```

```

        for (int i = 0; i < zi.Length; i++) // заполняем массив xi через решение
уравнений
    {
        if (zi[i] < 0.5)
        {
            Mass_xi[i, nM] = Math.Round(Math.Sqrt(zi[i] * 2), accuracy);
            tempValue = Convert.ToInt32(Mass_xi[i, nM] / (2 / numStolb));
            gistogramm[tempValue] += 1;
        }
        else
        {
            Mass_xi[i, nM] = Math.Round(quadraticEquation(zi[i]), accuracy);
            tempValue = Convert.ToInt32(Mass_xi[i, nM] / (2 / numStolb));
            gistogramm[tempValue] += 1;
        }
    }

}
textBox1.Text = Convert.ToString(TVMin * (2 / numStolb));
textBox2.Text = Convert.ToString(TV * (2 / numStolb));
textBox3.Text = Convert.ToString(gistogramm.Max());
if (checkBox_Zi.Checked == false)
{
    return gistogramm;
}
else
    return gistogramm_zi;
}

public double quadraticEquation(double zi)
{
    double x1, x2;
    double radical;
    int temp2x = 0;

    radical = Math.Sqrt(16 - 8 * (1 + zi));

    if (radical < 0)
    {
        return 0;
    }
    else if (radical == 0)
    {
        return 2;
    }
}

```

```

    }
else
{
    x1 = (4 + radical) / 2;
    x2 = (4 - radical) / 2;

    if ((x1 >= 1 & x1 <= 2) & (x2 >= 1 & x2 <= 2))
    {
        temp2x = temp2x++;
    }

    if (x1 >= 1 & x1 <= 2)
    {
        return x1;
    }
    else if (x2 >= 1 & x2 <= 2)
    {
        return x2;
    }

    return 0;
}
}

```

```

private void RBGis_CheckedChanged(object sender, EventArgs e)
{
    zedGraph.Visible = true;
    pictureBox.Visible = false;
}

```

```

private void RBPicture_CheckedChanged(object sender, EventArgs e)
{
    zedGraph.Visible = false;
    pictureBox.Visible = true;
}

```

```

private void button_Xi_Click(object sender, EventArgs e)
{
    drawZedGraph(Gis);
    drawGistogramm(Gis);
}

```

```

private void button_Zi_Click(object sender, EventArgs e)

```

{  
}  
}  
}