

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  
подпись      инициалы, фамилия  
« \_\_\_\_ » \_\_\_\_\_ 2019 г.

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

Разработка программного модуля для моделирования задач со случайными  
входными данными

09.04.02 Информационные системы и технологии

09.04.02.01 Информационно-управляющие системы

Руководитель	_____	доц, канд.техн.наук	О.А. Попова
	подпись, дата		
Выпускник	_____	КИ17-02-1М	А.М. Мерко
	подпись, дата		
Рецензент	_____	профессор, д-р. техн. наук	М.Н. Фаворская
	подпись, дата		
Нормоконтролер	_____	доц, канд.техн.наук	О.А. Попова
	подпись, дата		

Красноярск 2019

## РЕФЕРАТ

Выпускная квалификационная работа в форме магистерской диссертации по теме «Разработка программного модуля для моделирования задач со случайными входными данными» содержит 102 страницы текстового документа, 1 приложение и 40 использованных источников.

**ЧИСЛЕННЫЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ, АРИФМЕТИЧЕСКИЕ ОПЕРАЦИИ, НЕОПРЕДЕЛЕННЫЕ ДАННЫЕ, СЛУЧАЙНЫЕ ПЕРЕМЕННЫЕ, ФУНКЦИЯ ПЛОТНОСТИ ВЕРОЯТНОСТИ, РЕКУРСИВНЫЕ ВЫЧИСЛЕНИЯ, ВЕРОЯТНОСТНЫЕ РАСШИРЕНИЯ.**

Магистерская диссертация посвящена актуальной проблеме, связанной с решением задач с неопределенными входными данными. Присутствие неопределенности во входных данных при решении многих практических задач создает необходимость применения методов, которые бы увеличивали скорость выполнения задач и повышали точность вычислений.

Целью диссертационной работы является разработка программного модуля, реализующего новые методы и подходы к вычислению функций со случайными аргументами, которые обеспечивают надежность и снижение вычислительной сложности численных процедур.

Основные задачи:

- исследование новых методов и подходов в области решения задач со случайными данными на основе численного моделирования;
- применение методов вычислительного вероятностного анализа в задачах вычисления функций от случайных аргументов;
- разработка рекурсивно – параллельной схемы вычислительного процесса для вычисления вероятностных расширений;
- разработка программного модуля и проверка на тестовых примерах.

Основные результаты:

- Разработан программный модуль для решения задач с случайными

входными данными;

- Выполнено тестирование данного модуля.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	6
1 Анализ предметной области .....	9
1.1 Анализ типов случайной неопределенности.....	9
1.2 Анализ методов.....	10
1.3 Численные вероятностный анализ.....	32
1.3.1 Назначение ЧВА .....	32
1.3.2 ЧВА – арифметика.....	34
1.3.3 Кусочно-полиномиальные модели .....	38
1.4 Вероятностные расширения.....	42
1.5 Основная теорема о вероятностных расширениях .....	48
2 Техника рекурсивных вычислений .....	58
2.1 Рекурсивные задачи и вычисления.....	58
2.2 Рекурсивные функции.....	63
3 Разработка программного модуля .....	66
3.1 Постановка задачи и необходимые теоретические сведения.....	66
3.2 Структура модуля и описание его основных блоков .....	74
3.3 Организация работы программного модуля .....	75
3.4 Распараллеливание программы на языке PascalABC .....	81
3.5 Тестирование работы программного модуля.....	84
ЗАКЛЮЧЕНИЕ .....	92
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	93
ПРИЛОЖЕНИЕ А .....	97

Слайды презентации .....97

## ВВЕДЕНИЕ

Диссертация посвящена актуальной теме, связанной с вопросами исследования различных задач со случайными входными данными на основе методов численного моделирования и вычислительного вероятностного анализа.

Целью диссертационной работы является разработка программного модуля, реализующего новые методы и подходы к вычислению функций со случайными аргументами, которые обеспечивают надежность и снижение вычислительной сложности численных процедур.

Для многих практических задач требуется вычисление функций от некоторого числа случайных аргументов, которые являются важной частью многих сложных, многомерных, многопараметрических математических моделей, реализующих информационный подход, связанный с представлением и обработкой данных больших объемов. К таким задачам можно отнести ряд задач цифровой экономики, задачи обработки и анализа данных дистанционного зондирования земли, гидрологических данных и так далее. В подобных задачах часто необходимо численными методами исследовать модели со случайными входными данными, заданными своими плотностями вероятности. В тех случаях, когда их функции плотности вероятности известны, исследование таких моделей можно свести к вычислению функций со случайными аргументами. Универсальным подходом к решению таких задач является метод Монте-Карло, который давно и успешно используется для решения разнообразных задач в условиях неопределенности. При всех его положительных качествах этот метод обладает рядом недостатков. Один из самых существенных есть низкая скорость сходимости и большой объем вычислений, что представляет дополнительные сложности при работе с данными большого объема. Многомерность и случайный характер функциональных параметров, необходимость обработки данных массивов

большой размерности увеличивают время выполнения и вычислительную сложность численных процедур, что существенным образом сказывается, как на возможности получения конечного численного результата при заданных ограничениях по времени, так и надежности вычислительных процедур в отношении точности и достоверности полученных результатов. В настоящее время решение данной проблемы возможно на основе применения активно развивающегося подхода, который сформировался на стыке таких понятий, как параллельные вычислительные системы и параллельные численные методы. В работе предлагается подход к вычислению функций от случайных аргументов, представляющий собой технику быстрых вычислений, основанную на идее распараллеливания вычислительного процесса и применении методов численного вероятностного анализа в условиях случайной неопределенности [6].

Основными задачами диссертационного исследования являются:

- исследование новых методов и подходов в области решения задач со случайными данными на основе численного моделирования;
- применение методов вычислительного вероятностного анализа в задачах вычисления функций от случайных аргументов;
- разработка рекурсивно – параллельной схемы вычислительного процесса для вычисления вероятностных расширений;
- разработка программного модуля и проверка на тестовых примерах.

Диссертация состоит из трех глав. В первой главе приводится анализ предметной области, рассматриваются основные методы и даются необходимые теоретические сведения. Для решения поставленной задачи обосновывается применение численного вероятностного анализа и приводятся основные теоремы и утверждения для применения вероятностных расширений к вычислению функций от случайных аргументов. Во второй главе рассматриваются алгоритмы и методы рекурсивных вычислений, на основе которых реализуется техника быстрых вычислений. Описывается организация

рекурсивного процесса распараллеливания на основе вычисления вероятностных расширений. Третья глава посвящена вопросам разработки программного модуля и тестирования его работы.



## **1 Анализ предметной области**

### **1.1 Анализ типов случайной неопределенности**

Для многих практических задач требуется вычисление функций от случайных аргументов, которые являются важной составляющей многих сложных, многомерных, многопараметрических математических моделей, реализующих информационный подход, связанный с представлением и обработкой данных больших объемов. К таким задачам можно отнести ряд задач цифровой экономики, задачи обработки и анализа данных дистанционного зондирования земли, гидрологических данных и так далее. В подобных задачах часто необходимо численными методами исследовать модели со случайными входными данными, заданными своими плотностями вероятности. В тех случаях, когда их функции плотности вероятности известны, исследование таких моделей можно свести к вычислению функций со случайными аргументами.

Анализ публикаций показал, что неопределенности в данных можно классифицировать как элиаторную и эпистемистическую.

Выделяют элиаторную неопределенность, которая определяется изменчивостью процессов и явлений, и эпистемическую неопределенность, которая определяется малым количеством или нехваткой знаний о системе и характеризуется неопределенностью вероятностных оценок.

Входные данные, в которых содержится случайная неопределенность, задаются, как правило, определенными вероятностными распределениями или функцией плотности вероятности. Заметим, что важную роль играет объем имеющейся информации.

В этом случае выделяют два вида неопределенности: элиаторную неопределенность, которая определяется изменчивостью процессов и явлений и достаточным объемом информации, и эпистемическую неопределенность,

которая определяется малым количеством данных или нехваткой знаний о системе, а также неопределенностью их вероятностных оценок.

Примером эпистемической неопределенности является интервальная неопределенность, которая характеризуется состоянием неполного или частичного знания о интересующем нас значении, когда известна только ее принадлежность какому-либо интервалу, т. е. мы можем задать пределы их изменения или границы вероятных значений этой величины. Например, интервальная неопределенность присутствует в измерении объемов продаж, объемов производства, прибыли, оценке себестоимости и других экономических показателях. Это позволяет сделать вывод об интервальном характере ряда экономических данных, которые служат основой принятия экономических решений.

## **1.2 Анализ методов**

Важно отметить, что каждому типу неопределенности в данных соответствует своя арифметика. Так, для интервальных данных разработана интервальная арифметика, которая является частью интервального анализа. Основная идея интервального анализа – это замена вещественных функций и арифметических операций и над вещественными числами интервальными функциями и операциями, преобразующими интервалы, которые содержат эти числа. Для нечетких данных в 1965 году Заде предложил теорию нечетких множеств, которая в настоящее время превратилась в детально разработанную область с широким спектром приложений к задачам практического характера. Нечеткая арифметика строится на таких понятиях, как нечеткая величина, степень принадлежности, функция принадлежности. Случайная неопределенность и арифметические операции над случайными числами рассматриваются в теории вероятностей.

## *Методы распространения неопределенностей*

Анализ и обработка информации в условиях информационной неопределенности приводит к необходимости применения специальных методов исследования. В настоящее время во многих работах современных авторов, которые посвящены исследованию систем в условиях информационной неопределенности, часто применяется подход, который в зарубежных источниках называют «*propagation of uncertainty*», либо «распространение неопределенности» в переводе на русский язык. Анализируя содержательный смысл этого понятия выделяют один важный аспект процедуры распространения неопределенности в данных. Данный способ основан на получении дополнительных оснований или знаний для исследования входных данных. Иначе говоря, если присутствует условие недостаточности данных или полного отсутствия эмпирической информации о входных данных, а также оснований для выдвижения идей и предположений о каком-либо распределении входных параметров данных, то существует необходимость распространить (в пер. *propagate*) существующую неопределенность данных для получения достаточных выводов в соответствии с принципом недостаточного основания. Для получения необходимых оснований оценки или восстановления какого-либо входного распределения на основе неточной или неполной информации, применяют различные процедуры. Примером может стать рассмотрение распределение вероятностей, которые имеют максимальную энтропию, которая допускается имеющейся априорной информацией. При этом необходимо следовать принципу: выбирать тип представления, который соответствует количеству имеющейся информации и оставаться верным имеющейся в данный момент информации, включая присутствующие информационные пробелы. Применение данного принципа недостаточности либо достаточности оснований позволяет нам существенно расширить знания о формах представления неопределенностей.

Используя метод вероятностных границ можно достичь «распространение неопределенности». Основная идея данного метода основывается на следующем: «Есть нечто, что мы можем сказать о неизвестном распределении. В частности, это его кумулятивная функция распределения вероятностей, которая должна лежать в области – ящике и быть ограничена нулем и единицей от минимума/максимума горизонтально и по вертикали. Истинная же функция распределения, какой бы она ни являлась, должна находиться в этой области». Эта идея построения вероятностных границ оказалась очень продуктивной. Еще один имеющийся подход к распространению неопределенности в данных – это Облака Неймайера, которые позволяют нам неполную стохастическую информацию представить понятным, четким и вычислительно привлекательным способом, который позволяет визуализировать нашу неопределенность и обладает четкой семантикой. Также он выступает посредником между самим понятием нечеткого множества и его вероятностным распределением.

### ***Методы статистических испытаний. Метод Монте-Карло***

При исследовании систем в условиях элитарной неопределенности используются самые разнообразные методы, опирающиеся на теорию вероятностей, в том числе такие подходы, которые основаны на описании законов распределения случайных величин. В таком случае важно отметить, что вероятностные оценки законов распределения носят детерминированный характер. Одним из самых распространенных и широко применяемых методов для решения задач в условиях неопределенностей является метод Монте-Карло.

Имитационное моделирование Монте-Карло в общем случае является процедурой, при помощи которой математическая модель нахождения какого-либо показателя подвергается определенному количеству имитационных прогонов при помощи компьютера. Во время этого процесса имитации

происходит построение последовательных сценариев при использовании исходных данных, которые в проекте являются неопределенными данными, в связи с этим во время анализа приходится полагаться на случайные величины. Сам процесс имитации происходит со следующим условием: случайный выбор значений из определенных вероятностных распределений не должен нарушать существования предполагаемых или известных отношений корреляции между переменных. Несмотря на широкие возможности, метод Монте-Карло имеет ряд недостатков: Монте-Карло требует большого объема данных; не распространяет частичное незнание ни при какой частотной интерпретации; не может оценить превышение риском определенного уровня; не может быть применен для эффекта не сворачиваемости при решении задач обратных расчетов.

Рассмотрим еще один метод, автором которого является R.E. Moore, показывает, как можно определить совокупную функцию распределения арифметической операции над случайными величинами, когда случайные величины являются объединение непересекающихся интервалов, но с определенной долей вероятности встречаемости в пределах каждого интервала Мур использует только небольшое количество подгрупп или он все-таки получает результаты, которые соответствуют результатам, полученным методом Монте-Карло. Это может быть отчасти связано с тем, что кумулятивные распределения, как правило, выглядят лучше, чем соответствующие плотности, из-за сглаживающего эффекта интеграции.

Численный вероятностный анализ (ЧВА) является еще одним способом распространения информационной неопределенности, в том числе для практических задач, когда имеющиеся вероятностные оценки входных данных и параметров носят неопределенный характер. При достижении цели, которая требует снижения уровня информационной неопределенности в данных и получения дополнительной информации о распределении параметров входных

данных в условиях информационной неопределенности и недостаточности предложено использование гистограммного подхода.

Идея гистограммного подхода заключается в следующем: наряду с общими представлениями случайных величин своими плотностями в виде непрерывных функций, можно рассматривать случайные величины, плотность распределения которых представляет гистограмму.

Для нахождения решения этих задач предлагается использование интервальных гистограмм и гистограмм второго порядка. В отдельных случаях, когда отсутствует возможность получения точной функции распределения случайной величины, задаются оценки плотности распределения снизу и сверху. Данные оценки очень удобно аппроксимировать с помощью интервальных гистограмм. Называть гистограмму интервальной принято том случае, если значения этой гистограммы принимают только интервальные значения.

Вместе с тем для интервальных гистограмм в рамках ЧВА предложено использование гистограммы второго порядка, т.е. это гистограммы, каждый столбец которой – гистограмма.

### *Приближение функций*

Интерполяционный многочлен Лагранжа

Пусть известны значения некоторой функции  $f$ , заданной на отрезке  $[a,b]$  в  $n + 1$  различных точках  $a = x_0 < x_1 < x_2 \dots < x_n = b$ , которые обозначим следующим образом:

$$x_i = f(x_i), i = 0, 1, \dots, n$$

Множество точек  $\{x_i\}$  далее будем называть сеткой. Возникает задача приближенного восстановления функции  $f$  в произвольной точке  $x$ . Построим для этого алгебраический многочлен  $L_n(x)$  степени  $n$ :

$$L_n = f(x_i), i = 0, 1, \dots, n \quad (1)$$

Такой многочлен будем называть *интерполяционным*, точки  $x_i$  узлами *интерполяции*. Рассмотрим общий вид интерполяционного многочлена:

$$L_n(x) = \sum_{i=0}^n p_{ni}(x) f_i, \quad (2)$$

где

$$p_{ni} = \frac{(x-x_0)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_0)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)}, i = 0, 1, \dots, n. \quad (3)$$

Приведем примеры интерполяционного многочлена Лагранжа. Пусть  $n=1$ , тогда

$$L_1(x) = \frac{x-x_1}{x_0-x_1} f_0 + \frac{x-x_0}{x_1-x_0} f_1. \quad (4)$$

При  $n=2$

$$L_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f_0 + \\ + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f_2.$$

### *Погрешность интерполяции.*

Оценим функцию  $R_n$ :

$$R_n(x) = f(x) - L_n(x). \quad (5)$$

Предположим, что  $f \in C^{n+1}$ , т.е.  $n + 1$  раз непрерывно дифференцируема. Представим  $R_n(x)$  в следующем виде:

$$R_n(x) = \omega_n(x)r_n(x), \quad (6)$$

где

$$\omega_n(x) = (x - x_0)(x - x_1) \dots (x - x_n). \quad (7)$$

Зафиксируем произвольное значение  $x, x \neq x_i$  и рассмотрим следующую функцию от  $t$ :

$$\varphi(t) = L_n(t) + \omega_n(t)r_n(x) - f(t). \quad (8)$$

Эта функция обращается в нуль при  $t = x_i, t = x$  т.е. в  $n + 2$  точках.

По теореме Ролля производная по  $t$  от  $\varphi$  обращается в нуль по крайней мере в  $n + 1$  точке, вторая производная равна нулю в  $n$  точках и т.д. Таким образом, существует точка  $\xi$  такая, что  $\varphi^{(n+1)}(\xi) = 0$ . Поскольку  $L_n^{(n+1)}(\xi) = 0, \omega_n^{(n+1)}(\xi) = (n + 1)!$ , получаем

$$(n + 1)!r_n(x) - f_n^{(n+1)}(\xi) = 0.$$



Следовательно,

$$r_n(x) = \frac{f_n^{(n+1)}(\xi)}{(n+1)!}.$$

Остаточный член выглядит следующим образом:

$$R_n(x) = \omega_n(x) \frac{f_n^{(n+1)}(\xi)}{(n+1)!}.$$

Если известна оценка  $\varphi_n^{(n+1)}(\xi) \leq M_{n+1}$ , то

$$|f(x) - L_n(x)| \leq |\omega_n(x)| \frac{M_{(n+1)}}{(n+1)!}. \quad (9)$$

В качестве примера рассмотрим оценку погрешности при линейной интерполяции  $n = 1$ . Заметим, что  $\omega_2(x) = (x - x_0)(x - x_1)$  и, следовательно,

$$\max_{[x_0, x_1]} |\omega_2(x)| = \max_{[x_0, x_1]} |(x - x_0)(x - x_1)| = h^2/4, \quad (10)$$

где  $h = x_1 - x_0$ . Таким образом,

$$\max_{[x_0, x_1]} |f(x) - L_1(x)| \leq h^2 \frac{M_2}{4}. \quad (11)$$

В дальнейшем нам потребуется следующая лемма.

**Лемма 1.** Пусть  $f \in C[a, b]$ ,  $\xi_i \in [a, b]$  – произвольные точки,  $i = 1, 2, \dots, n$ . Тогда существует такая точка  $\xi \in [a, b]$ , что

$$\frac{f(\xi_1) + f(\xi_2) + \dots + f(\xi_n)}{n} = f(\xi).$$

### ***Простейшие формулы численного дифференцирования***

Рассмотрим простейший случай вычисления первой производной. По определению,

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

Тогда, задавшись конечным  $\Delta x$ , положим

$$f'(x) \approx \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Пусть  $x_i = x_0 + ih$ ,  $h > 0$ . Обозначим  $f_i = f(x_i)$ ,  $f'_i = f'(x_i)$  и т.п. Несложно убедиться, что справедливы следующие соотношения:

$$f'_0 = \frac{f_1 - f_0}{h} - \frac{h}{2} f''(\xi), x_0 < \xi < x_1.$$

$$f'_0 = \frac{f_1 - f_{-1}}{2h} - \frac{h^2}{6} f'''(\xi), x_{-1} < \xi < x_1.$$

$$f''_0 = \frac{f_{-1} - 2f_0 + f_1}{h^2} - \frac{h^2}{12} f^{(4)}(\xi), x_{-1} < \xi < x_1.$$

Точка  $\xi$  в каждой из приведенных формул неизвестна. Следовательно, предположив малость  $h$ , мы можем отбросить остаточные члены и получить формулы численного дифференцирования.

Разностная производная

$$f'_0 \approx \frac{f_1 - f_0}{h}, \quad (12)$$

центральная разность

$$f'_0 \approx \frac{f_1 - f_{-1}}{2h}, \quad (13)$$

вторая разностная производная

$$f''_0 \approx \frac{f_{-1} - 2f_0 + f_1}{h^2}. \quad (14)$$

Оценим погрешность приведенных формул:

$$\left| f'_0 - \frac{f_1 - f_0}{h} \right| \leq \frac{h}{2} \max_{[x_0, x_1]} |f''(x)|,$$

$$\left| f'_0 - \frac{f_1 - f_{-1}}{2h} \right| \leq \frac{h^2}{6} \max_{[x_0, x_1]} |f'''(x)|.$$

Погрешность формулы (12) имеет *первый порядок относительно  $h$* , а погрешность формул (13), (14) имеет *второй порядок относительно  $h$* .

### ***Применение интерполяционного многочлена Лагранжа.***

Для нахождения приближенных значений производных и построения формул численного дифференцирования можно использовать интерполяционный многочлен Лагранжа. Для этого по известным значениям функции  $f$  в некоторых узлах  $x_i$  строят многочлен  $L_n(x)$  и полагают

$$f^{(m)}(x) \approx L_n^{(m)}(x), 0 \leq m \leq n.$$

**Теорема 1.** Пусть  $f \in C^{k+1}[x_0, x_n]$ . Тогда существует константа  $K$ , не зависящая от  $h$  и функции  $f$ , такая, что

$$\max_{[x_0, x_n]} |f^{(m)}(x) - L_n^{(m)}(x)| \leq Kh^{k+1-m} \max_{[x_0, x_n]} |f^{(k+1)}(x)|,$$

где  $L_n(x)$  – интерполяционный многочлен Лагранжа для функции  $f$ .

### ***Сплайны***

Пусть на отрезке  $[a, b]$  задана сетка  $a = x_0 < x_1 < x_2 \dots < x_n = b$ . Для целого  $k \geq 0$  через  $C^k[a, b]$  обозначим множество  $k$  раз непрерывно дифференцируемых на  $[a, b]$  функций.

**Определение.** Функция  $s$  называется *сплайном* степени  $n$ , дефекта  $v$ , если  $s \in C^{n-v}[a, b]$ , на каждом отрезке  $[x_i, x_{i+1}]$  функция  $s$  является многочленом степени  $n$ .

Простейшим примером сплайна может служить кусочно-линейная интерполяция. Предположим, что в узлах сетки  $x_i$  заданы значения функции  $f_i$ . Таким образом, на каждом отрезке  $[x_i, x_{i+1}]$  мы можем построить линейную функцию  $s(x)$  согласно формуле (4). Следовательно, мы построили непрерывную кусочно-линейную функцию. Такую функцию можно считать сплайном первой степени. Формула (10) дает нам ответ о величине погрешности интерполяции на каждом отрезке. Далее, как и прежде,  $h = \max |x_{i+1} - x_i|$ .

**Теорема 2.** Пусть  $f \in C^2[x_0, x_n]$ . Тогда существует константа  $K$ , не зависящая от  $h$  и функции  $f$  такая, что

$$\max_{[a,b]} |f^{(m)}(x) - s^{(m)}(x)| \leq Kh^{2-m} \max_{[a,b]} |f^{(2)}(x)|, m = 0, 1. \quad (15)$$

Рассмотрим наиболее распространенные на практике эрмитовы кубические сплайны. Предположим, что в узлах сетки  $x_i$  заданы значения функции и ее производных  $f_i, f'_i$ . Тогда на каждом отрезке  $[x_i, x_{i+1}]$  мы можем построить кубический полином  $s(x)$  третьего порядка. Действительно,

$$S(x) = a_0 + a_1x + a_2x^2 + a_3x^3$$

и для полного определения сплайна нам необходимо вычислить  $a_2, a_3$   $a_0, a_1, a_2, a_3$  – всего 4 неизвестных. Для их определения мы имеем 4 уравнения:

$$\begin{aligned} s(x_i) &= f_i, \\ s'(x_i) &= f'_i, \\ s(x_{i+1}) &= f_{i+1}, \\ s'(x_{i+1}) &= f'_{i+1}. \end{aligned} \quad (16)$$

Приведем для примера вид второго уравнения  $s'x_i = f'_i$ :

$$a_1 + a_2x_i + a_3x_i^2 = f'_i.$$

Решив эту систему, мы определим сплайн на отрезке  $[x_i, x_{i+1}]$ . Таким же образом мы можем построить сплайн на произвольном отрезке. У этого подхода есть один существенный недостаток: необходимо постоянно решать систему уравнений для определения коэффициентов сплайна. Оказывается, мы можем легко преодолеть это препятствие, достаточно представить наш сплайн в следующем виде:

$$\begin{aligned} s(x) = & f_i\phi_0((x-x_i)/h) + f'_ih\bar{\phi}_0((x-x_i)/h) + \\ & + f_{i+1}\phi_1((x-x_i)/h) + f'_{i+1}h\bar{\phi}_1((x-x_i)/h). \end{aligned} \quad (17)$$

Введенные функции  $\phi_0, \bar{\phi}_0, \phi_1, \bar{\phi}_1$  определяют следующим образом:

$$\begin{aligned} \phi_0(0) = 1, \phi_0(1) = 0, \bar{\phi}_0(0) = 0, \bar{\phi}_0(1) = 0, \\ \phi_1(0) = 0, \phi_1(1) = 1, \bar{\phi}_1(0) = 0, \bar{\phi}_1(1) = 0, \\ \bar{\phi}_0(0) = 0, \bar{\phi}_0(1) = 1, \phi_0(0) = 0, \phi_0(1) = 0, \\ \phi_1(0) = 0, \phi_1(1) = 0, \bar{\phi}_1(0) = 0, \bar{\phi}_1(1) = 1. \end{aligned}$$

Легко видеть, что мы можем найти эти функции как кубические полиномы, решив соответствующие системы линейных алгебраических уравнений. Приведем вид этих функций:

$$\phi_0(x) = (1-x)^2(1+2t), \bar{\phi}_0(x) = x(1-x)^2,$$

$$\phi_1(x) = x^2(3-2x), \bar{\phi}_1(x) = x^3 - x^2.$$

Другими словами, функции  $\phi_0, \bar{\phi}_0, \phi_1, \bar{\phi}_1$  представляют собой базис в пространстве решений системы (17).

Пусть  $f \in C^4[x_0, x_n]$ . Тогда существует константа  $K$ , не зависящая от  $h$  и функции  $f$ , такая, что

$$\max_{[a,b]} |f^{(m)}(x) - S^{(m)}(x)| \leq Kh^{4-m} \max_{[a,b]} |f^{(4)}(x)|, \quad m = 0, 1, 2, 3, \quad (18)$$

где  $S(x)$  – интерполяционный эрмитов сплайн для функции  $f$ .

### **Экстраполяция Ричардсона и правило Рунге**

Одно из первых правил для практической оценки погрешности было предложено К. Рунге в начале XX века. Это правило широко использовалось сначала в области квадратурных вычислений, затем в разностных методах и методе конечных элементов. Это правило основано на декомпозиции приближенного решения  $u^h$  как суммы [1]

$$u^h = u + h^k v + O(h^{k+m}), \quad (19)$$

где  $u$  есть искомое точное решение,

$v$  – неизвестная функция;

$h$  – малый параметр дискретизации, который чаще всего рассматривается, как шаг разностной сетки.

Целое значение  $k$  характеризует порядок точности приближенного решения, и  $m > 0$  характеризует порядок точности приближенного решения членом

погрешности  $h^k v$ . Поскольку  $u$  и  $v$  не зависят от  $h$ , для параметра  $h/2$  справедливо разложение:

$$u^{h/2} = u + \left(\frac{h}{2}\right)^k v + O(h^{k+m}). \quad (20)$$

Вычтем его из (20) избавляясь от  $u$ :

$$u^h - u^{h/2} = v \left(\frac{h}{2}\right)^k (2^k - 1) + O(h^{k+m}).$$

Отсюда можно определить главный член погрешности:

$$u^{h/2} - u \approx \frac{u^h - u^{h/2}}{2^k - 1}. \quad (21)$$

Поскольку в формуле (21) отброшен остаточный член порядка  $O(h^{k+m})$ , то полученное выражение не приводит к гарантированной оценке, но при достаточно малых  $h$  дает представление о величине погрешности численного решения.

Экстраполяция Ричардсона является общим методом для получения результатов высокой точности по формулам низкого порядка [20].

Объединим две аппроксимации таким способом, чтобы исключить ошибку порядка  $h^k$ . Умножая (20) на  $2^k$  и вычитая из (19), получаем

$$u = \frac{2^k}{2^k - 1} u^{h/2} - \frac{1}{2^k - 1} u^h + O(h^{k+m}).$$



## Непараметрические методы

В тех случаях, когда желательно по данным эксперимента построить оценку плотности вероятностей, экспериментаторы чаще всего прибегают к построению гистограммы. Процедура ее построения проста и состоит из следующих шагов. В области возможных значений измеряемой величины  $X$  строится сетка  $\omega = \{x_i | i = 1, 2, \dots, n\}$ . Определяется, сколько выборочных значений  $m_i$  от общего числа  $N$  оказалось в каждом интервале  $(x_{i-1}, x_i]$ .

Над каждым из интервалов строится вертикальный прямоугольник с площадью  $\frac{m_i}{N}$ . Высота прямоугольника  $P_i = \frac{m_i}{N(x_i - x_{i-1})}$ . Полученная совокупность прямоугольников и называется гистограммой. Другими словами гистограмма – кусочно-постоянная функция и определяется своей сеткой  $\omega$ , значениями  $\{P_i\}$ , принимающая на каждом интервале  $(x_{i-1}, x_i]$  постоянное значение  $P_i$ .

Основанием для использования гистограммы  $p_h(x)$  в качестве оценки неизвестной плотности вероятностей  $p(x)$  является кусочно-интегральная сходимость  $p_h(x)$  к  $p(x)$ , которая следует из того, что относительная частота  $\frac{m_i}{N}$  события  $X \in (x_{i-1}, x_i]$  сходится к его вероятности  $p_i$

$$\frac{m_i}{N} \rightarrow \int_{x_{i-1}}^{x_i} p(x) dx.$$

Такой сходимости в ряде практических случаев оказывается достаточно; однако всегда желательно по возможности улучшить оценку при заданных ограничениях. Несколько факторов влияет на качество гистограммы: объем выборки  $N$  величина интервалов группировки. Все осложняется еще и тем, что степень влияния этих факторов зависит от неизвестного экспериментатору до опыта истинного распределения вероятностей  $p(x)$ . Поэтому на практике гистограммы строят с некоторым учетом свойств полученной выборки,

например, величина интервала группировки выбирается так, чтобы не сгладить существенные особенности распределения. Объем выборки связывают с тем, чтобы в ячейке с наименьшим числом измерений их насчитывалось не менее пяти. Размещение интервалов связывают с положением наименьшего и наибольшего выборочных значений, и т. п. Отметим также, что на качество гистограммы влияет и точность измерений [27].

Теоретическая задача оптимизации гистограммы может быть сформулирована в нескольких вариантах, однако ее решение, связано с трудностями, так что конкретных результатов можно добиться лишь при некоторых частных предположениях.

Основной трудностью при построении гистограммы является выбор разбиения  $\omega$ . Рекомендуется интервалы  $(x_{i-1}, x_i]$  выбирать так, чтобы в каждый из них попадало одинаковое количество членов выборки (5-10, если  $N$  порядка 100). При увеличении  $N$  количество членов выборки в каждом интервале необходимо увеличивать. Если обозначить  $h_{max}$  длину наибольшего интервала, то  $h_{max}$  играет роль параметра регуляризации при оценивании плотности по гистограмме. При  $N \rightarrow \infty$ ,  $h_{max}$  должен уменьшаться согласованно с ростом  $N$ .

Правило Стёрджеса – эмпирическое правило определения оптимального количества интервалов, на которые разбивается наблюдаемый диапазон изменения случайной величины при построении гистограммы плотности её распределения. Названо по имени американского статистика Герберта Стёрджеса (Herbert Arthur Sturges, 1882–1958). Количество интервалов  $n$  определяется как:

$$n = 1 + [\log_2 N]$$

Обоснование правила следует из примера. Построим гистограмму со столбцами, каждый шириной 1 и по центру точки  $i = 0, 1, \dots, k - 1$ .

Предположим, что  $i$ -й столбец имеет высоту равную биномиальному коэффициенту  $C_i^{k-1} = \binom{k-1}{i}$ . С ростом  $k$ , гистограмма принимает форму нормальной плотности с средним  $(k-1)/2$  и дисперсией  $(k-1)/4$ . Общий размер выборки вытекает из биномиального разложения

$$n = \sum_{i=0}^{k-1} C_i^{k-1} = (1+1)^{k-1} = 2^{k-1}$$

из которого правило Стёрджеса вытекает немедленно.

### *Ядерные оценки функции плотности вероятности*

Для оценки функция плотности вероятности часто используются непараметрические методы, например, методы построения ядерных оценок. Отметим, что вплоть до середины 50-х годов в качестве единственного подхода для построения непараметрической оценки функция плотности вероятности использовалась гистограмма. Первые важные результаты в области применения ядерных оценок для функции плотности вероятности были получены в работах М.Розенблатта, Э.Парзена и Н. Ченцова.

В общем виде ядерная оценка может быть записана в виде

$$\hat{f}^h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - \xi_i}{h}\right) = \frac{1}{Nh} \sum_{i=1}^N K_h(x - \xi_i),$$

где

$$K_h(t) = K(t/h)/h.$$

Обозначим

$$K_h(x, \xi_i) = K\left(\frac{x - \xi_i}{h}\right).$$

где  $\xi$  случайная величина с функцией плотности вероятности  $f(x)$ .

Тогда математическое ожидание

$$E[\hat{f}^h(x)] = E[K_h(x, \xi)]$$

и

$$\sigma_N = D[\hat{f}^h(x)] = \frac{1}{N} D[K_h(x, \xi)].$$

Значение математического ожидания можно записать, как

$$E[K_h(x, \xi)] = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x-t}{h}\right) f(t) dt = \int_{-\infty}^{\infty} K(\eta) f(x - h\eta) d\eta.$$

Заметим, что

$$f(x - h\eta) = f(x) - hf'(x)\eta + \frac{h^2}{2} f''(x)\eta^2 + \frac{h^3}{6} f^{(3)}(x)\eta^3 + O(h^4).$$

$$\begin{aligned} E[K_h(x, \xi)] &= f(x) \int_{-\infty}^{\infty} K(\eta) d\eta - hf'(x) \int_{-\infty}^{\infty} \eta K(\eta) d\eta + \\ &+ \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} K(\eta) \eta^2 d\eta + \frac{h^3}{6} f^{(3)}(x) \int_{-\infty}^{\infty} K(\eta) \eta^3 d\eta + O(h^4). \end{aligned}$$

Пусть ядро  $K$  удовлетворяет требованиям

$$\int_{-\infty}^{\infty} K(\eta) d\eta = 1, \int_{-\infty}^{\infty} \eta K(\eta) d\eta = 0$$

и

$$\int_{-\infty}^{\infty} \eta^3 K(\eta) d\eta = 0.$$

Обозначим

$$\int_{-\infty}^{\infty} \eta^2 K(\eta) d\eta = \sigma_K^2.$$

Тогда

$$E[\hat{f}^h(x)] = E[K_h(x, \xi)] = f(x) + \sigma^2 h^2 f''(x)/2 + O(h^4)$$

и

$$E[\hat{f}^h(x) - f(x)] = \sigma^2 h^2 f''(x)/2 + O(h^4).$$

Определим  $f^h(x) f^h(x)$

$$f^h(x) = E[\hat{f}^h(x)] = f(x) + \sigma^2 h^2 f''(x)/2 + O(h^4). \quad (22)$$

и  $f^{2h}(x)$

$$f^{2h}(x) = E[\hat{f}^{2h}(x)] = f(x) + \sigma^4 h^2 f''(x)/2 + O(h^4). \quad (23)$$

Далее оценим

$$D[K_h(x, \xi)] = E\left[\left(\frac{1}{h} K\left(\frac{x-\xi}{h}\right)\right)^2\right] - \left(E\left[\frac{1}{h} K\left(\frac{x-\xi}{h}\right)\right]\right)^2.$$

$$\begin{aligned} E\left[\left(\frac{1}{h} K\left(\frac{x-\xi}{h}\right)\right)^2\right] &= \frac{1}{h} \int_{-\infty}^{\infty} K^2(\eta) f(x-h\eta) d\eta = \\ &= \frac{f(x)}{h} \int_{-\infty}^{\infty} K^2(\eta) d\eta + \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} K^2(\eta) \eta^2 d\eta + O(h^4). \end{aligned}$$

В итоге получим

$$\sigma^2(x) = D[\hat{f}(x)] = \frac{f(x)}{Nh} \|K\|_2^2 + \frac{f(x)^2}{N} + O(h/N).$$

Без ограничения общности, можно записать, что [7]

$$M[(\hat{f}^h(x) - f(x))^2] = \sigma^2(x) + f''(x)^2 h^4/4 + O(h^6)$$

и

$$\sigma_I^2 = \int_{-\infty}^{\infty} \sigma^2(x) dx = \frac{\|K\|_2^2}{Nh} - \frac{\|f\|_2^2}{N} + O(h/N).$$

$$M \|\hat{f}^h - f\|_2^2 = \sigma_I^2 + \frac{\|f''\|_2^2 h^4}{4} + O(h^6) \quad (24)$$

Применим экстраполяцию Ричардсона к  $f^h(x)$  и  $f^{2h}(x)$ . Далее умножим (23) на 1/4 и вычтем результат из (24). Исключая  $\sigma^2 h^2 f''(x)/2$  из (23) и (24), мы получим

$$f(x) = \frac{4}{3} f^h(x) - \frac{1}{3} f^{2h}(x) + O(h^4).$$

Заметим, что мы построили приближение к функции  $f(x)$

$$f_{cor}^h(x) = \frac{4}{3} \hat{f}^h(x) - \frac{1}{3} \hat{f}^{2h}(x). \quad (25)$$

с точностью  $O(h^4)$ .

С другой стороны, применяя правило Рунге, мы можем получить оценку

$$f''(x) = 2(f^h(x) - f^{2h}(x))/(3\sigma h^2) + O(h^2)$$

или

$$\|\hat{f}''\| = \frac{2}{3\sigma h^2} \|\hat{f}^h - \hat{f}^{2h}\| \quad (26)$$

Таким образом, используя экстраполяцию Ричардсона, получены приближение к функции плотности вероятности с точностью  $O(h^4)$ . Далее это будет использовано для построения сплайновых аппроксимаций эмпирических функции плотности вероятности.

## 1.3 Численные вероятностный анализ

### 1.3.1 Назначение ЧВА

ЧВА существенно отличается от существующих подходов. Он оперирует с плотностями случайных величин, которые представлены гистограммами, а также кусочно-полиномиальными и дискретными функциями.

Заметим, что функция плотности вероятности является одним из имеющихся способов описания и представления стохастической неопределенности в данных. Аналитические формулы, которые приводятся в рамках классической теории вероятности, наглядно показывают выполнение простейших арифметических операций над случайными величинами. На практике же использование данных процедур либо практически невозможно, либо существенно затруднено. Наличие информации о плотности вероятности случайных величин приводит к возможности при расчетах учитывать и получать результаты в виде случайных величин с построенной плотностью вероятности.

Сравнение подходов интервальной математики и численных операций над плотностями вероятности случайных величин, показывает, что практически значимые плотности вероятности занимают лишь небольшую часть полученных интервалов. Поэтому, в таких случаях, границы решений мало информативны. Часто при работе со случайными величинами ограничиваются вычислением лишь небольшого числа характеристик: математическое ожидание и дисперсия [17].

Численные операции над плотностями вероятности случайных величин позволяют существенно поднять точность расчетов при сравнительно небольшом объеме.

В отличие от большинства существующих арифметик над неопределенными величинами (интервальная арифметика, арифметика



нечетких значений), арифметики, построенные на основе понятия вероятностного расширения функций, могут быть использованы как для независимых, так и для зависимых случайных величин, что является крайне актуальным для ряда практических задач.

Сравнение таких подходов, как интервальная математика и численные операции над плотностями вероятностей случайных величин наглядно доказывает то, что значимым на практике плотностям вероятностей относятся к небольшой части полученных интервалов. Из этого следует, что в таких случаях границы решений очень мало информативны. Очень часто во время работы со случайными величинами ограничиваются вычислением лишь небольшого числа характеристик: математического ожидания и дисперсии. Метод Монте-Карло является одним из подходов учета случайного характера входных данных. При всех его положительных качествах этот метод обладает рядом недостатков. Один из самых существенных его недостатков – низкая скорость сходимости. В определенных случаях, когда это представляется возможным, численные операции над плотностями вероятностей случайных величин позволяют существенно увеличить точность расчетов при сравнительно небольшом объеме вычислений.

Необходимость разработки операций над численными значениями, которые определяют предмет арифметики, определяется следующими основаниями. Функция плотности вероятности – это один из способов описания и представления стохастической неопределенности в данных. Основываясь на классической теории вероятности имеются уже готовые аналитические формулы выполнения самых простейших арифметических операций над случайными величинами. На практике – применение данных процедур либо существенно затруднено, либо практически полностью невозможно.

### 1.3.2 ЧВА – арифметика

ЧВА – арифметика начала разрабатываться с использования гистограммных моделей представления данных и получила название гистограммной арифметики.

Для создания арифметики было сформулировано понятие гистограммной переменной (гистограммное число) и разработаны численные процедуры гистограммной арифметики. Сама идея гистограммного подхода заключается в том, что помимо представления плотности случайной величины в виде непрерывных функций, можно также рассматривать случайные величины, плотность распределения которых представляет гистограмму [16].

Из всех существующих видов арифметических операций нас интересуют исключительно операции над двумя и более случайными величинами. Например необходимо вычислить некоторую арифметическую операцию  $*$   $\in \{+, \uparrow, -, /, \cdot, \}$  над двумя дискретными независимыми величинами  $\zeta$  и  $\eta$ , которые принимают соответствующие значения  $x_i$  с вероятностью  $p_i$  и  $y_i$  с вероятностью  $q_i$ . Результат – это случайная величина  $\psi$ , которая принимает все возможные значения  $x_i * y_i$  с вероятностью  $p_i q_i$ . Так как существует вероятность «комбинаторного взрыва», то возникает необходимость преобразования случайных величин в другие типы, например, гистограммы.

Рассмотрим систему двух непрерывных случайных величин  $(x_1, x_2)$  с плотностью распределения  $p(x_1, x_2)$ . Для определения плотности вероятности, полученной в результате арифметических действий над случайными величинами, существуют аналитические формулы. Например, для нахождения плотности вероятности  $p_{x_1+x_2}$  для суммы двух случайных величин  $x_1 + x_2$  используется соотношение

$$p_{x_1+x_2}(x) = \int_{-\infty}^{\infty} p(x-v, v)dv = \int_{-\infty}^{\infty} p(v, x-v)dv, \quad (27)$$

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

$$p_{x_1/x_2}(x) = \int_{-\infty}^{\infty} vp(xv, v)dv - \int_{-\infty}^{\infty} vp(v, xv)dv, \quad (28)$$

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

$$p_{x_1x_2}(x) = \int_{-\infty}^{\infty} (1/v)p(x/v, v)dv - \int_{-\infty}^{\infty} (1/v)p(v, x/v)dv, \quad (29)$$

Но для численных расчетов эти формулы не всегда удобны. На примере операции сложения продемонстрируем основные принципы разработки гистограммных операций. Пусть  $z = x_1 + x_2$ , носители  $x_1 - [a_1, a_2]$ ,  $x_2 - [b_1, b_2]$ ,  $p(x_1, x_2)$  – это плотность распределения вероятностей случайного вектора  $(x_1, x_2)$ . Отметим, что прямоугольник  $[a_1, a_2] \times [b_1, b_2]$  – это наш носитель плотности распределения вероятностей  $p(x_1, x_2)$ , который представлен на рисунке 7.

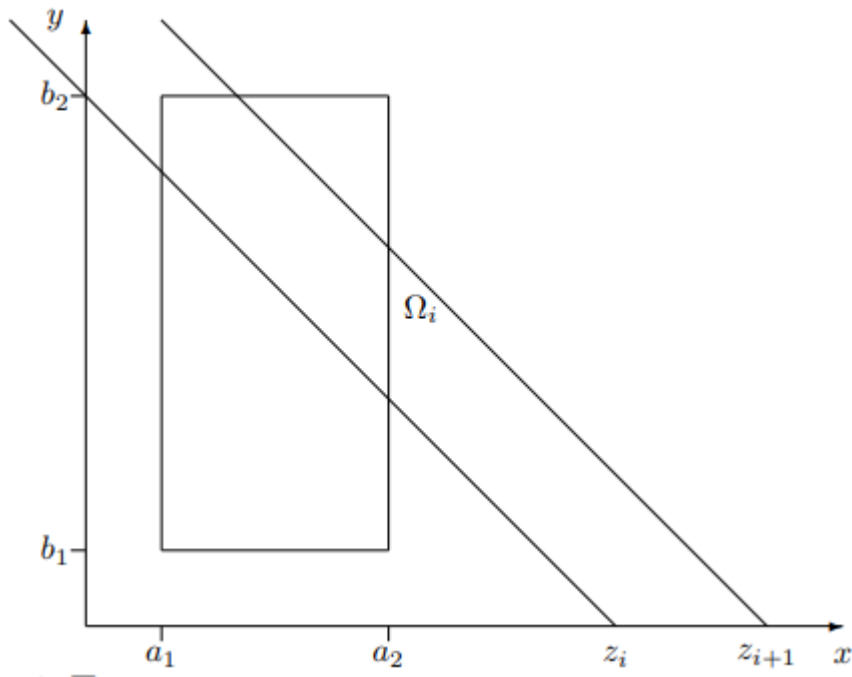


Рисунок 1 – Построение гистограммы суммы двух случайных величин

Плотность вероятности  $z$  на интервале  $[a_1 + b_1, a_2 + b_2]$  отлична от нуля. Отметим, что  $z_i, i = 0, 1, \dots, n$  – это точки деления нашего интервала на  $n$  отрезков. Из этого следует, что вероятность попадания величины  $z$  в наш интервал  $[z_i, z_{i+1}]$  определяется по формуле

$$p(z_i < z < z_{i+1}) = \iint_{\Omega_i} p(x_1, x_2) dx_1 dx_2 \quad (30)$$

где  $\Omega_i = \{(x_1, x_2) | z_i \leq x_1 + x_2 \leq z_{i+1}\}$ . И окончательно  $p_{z_i}$  имеет вид

$$p_{z_i} = \iint_{\Omega_i} p(x_1, x_2) dx_1 dx_2 / (z_{i+1} - z_i). \quad (31)$$

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

$$z = a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (32)$$

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

$$P(z_i < z < z_{i+1}) = \int \dots \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \quad (33)$$

где

$$\Omega_i = \{(x_1, x_2, \dots, x_n) | z_i < a_1x_1 + a_2x_2 + \dots + a_nx_n < z_{i+1}\}, p_{zi}$$

имеет вид

$$P(z_i) = \int \dots \int_{\Omega_i} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n / (z_{i+1} - z_i), \quad (34)$$

Рассмотрим в качестве примера операцию  $\max(x, y)$ . Вероятность того, что  $\max(x, y) < z$  определяется через функцию распределения  $F$ :

$$F(z) = \int_{-\infty}^z P_x(\xi) d\xi - \int_{-\infty}^z P_y(\xi) d\xi, \quad (35)$$

Используя функцию распределения  $F$ , можно построить гистограмму для  $\max(x, y)$  на сетке  $\{z_i\}$ . Тогда  $p_i = F(z_{i+1}) - F(z_i)$  [13].

### 1.3.3 Кусочно-полиномиальные модели

Рассмотрим кусочно-полиномиальные модели для аппроксимации эмпирических функций плотности вероятности. прежде всего это:

- кусочно-постоянные функции (гистограммы);
- кусочно-линейные функции (частотные полигоны);
- кусочно-кубические функции (сплайны).

Все кусочно-полиномиальные функции определяются сеткой  $\omega = \{x_i, i = 0, 1, \dots, n\}$ , где  $n \geq 2$  – целое и дополнительными массивами  $F = \{f_i, i = 1, \dots, n\}$ . Таким образом, гистограмма определяются сеткой  $\omega = \{x_i, i = 0, 1, \dots, n\}$  и массивов  $P = \{p_i, i = 1, \dots, n\}$ . На каждом отрезке  $[x_{i-1}, x_i)$  гистограмма принимает постоянное значение  $p_i$ .

Частотный полигон определяются сеткой  $\omega$   $f = \{f_i, i = 0, 1, \dots, n\}$ . На каждом отрезке  $[x_{i-1}, x_i)$  частотный полигон задается линейной функцией

$$l(x) = f_{i-1} \frac{x_i - x}{x_i - x_{i-1}} + f_i \frac{x - x_{i-1}}{x_i - x_{i-1}}.$$

Кубический сплайн определяются сеткой  $\omega$  и двумя массивами  $f = \{f_i, i = 0, 1, \dots, n\}$  и  $m = \{m_i, i = 0, 1, \dots, n\}$ . На каждом отрезке  $[x_{i-1}, x_i)$  сплайн задается кубической функцией

$$s(x) = f_{i-1}\phi_0(t) + f_i\phi_1(t) + hm_{i-1}\psi_0(t) + hm_i\psi_1(t),$$

$$\phi_0(x) = (1-x)^2(1+2t), \psi_0(x) = x(1-x)^2,$$

$$\phi_1(x) = x^2(3-2x), \psi_1(x) = x^3 - x^2.$$

Рассмотрим вопрос построения сплайна  $s$ , аппроксимирующего  $\hat{f}_{cor}(x)$ , так чтобы выполнялась оценка

$$\|f - s\| \leq Ch^4.$$

Заметим, что точность построенных ядерных оценок определяется из соотношения

$$M[f(x) - \hat{f}_{cor}(x)] = O(h^4),$$

что значительно точнее стандартных ядерных оценок. Важно отметить, что прямые вычисления  $\hat{f}_{cor}(x)$  в произвольной точке  $x$  требуют значительных вычислительных затрат [15].

Для этих целей построим в области  $[a, b]$  носителя функции плотности вероятности  $f(x)$  сетки  $\omega_z = \{z_i = a + ih_z, i = 0, \dots, N_z\}$ ,  $\omega_x = \{x_i = a + ih_x, i = 0, \dots, N_x\}$ .

На сетке  $\omega_z$  вычислим значения  $f_i = \hat{f}_{cor}(z_i)$ . Сплайн  $s$  будем строить на сетке  $\omega_x$ . Краевые условия выберем следующим образом  $s(a) = 0, s'(a) = 0, s(b) = 0, s'(b) = 0$

$$\sum_{i=1}^{N_z} (s(z_i) - f_i)^2 \rightarrow \min. \quad (36)$$

В случае кубических сплайнов, как классических так, и эрмитовых, задача (10) сводится к решению пятидиагональной системы линейных алгебраических уравнений.

Для кубических сплайнов справедлива следующая оценка

$$\|f^\nu - s^\nu\| \leq Kh_x^{4-\nu} \|f^{(4)}\|, \quad (37)$$

где  $K$  – константа, не зависящая от  $h_x$ .

Задачу можно упростить, если вычислить в узлах сетки  $\omega_x$  значения  $\hat{f}_i$ . Для этих целей будем использовать значения  $\hat{f}_{cor}(z_i)$  и процедуры сглаживания. Например, для классических кубических сплайнов можно использовать метод скользящего среднего, метод взвешенной локальной регрессии, фильтр Савицкого–Голея. Следует стремиться, чтобы выполнялись оценки

$$|\hat{f}_i - f(x_i)| = O(h_x^4).$$

В этом случае построение сплайна сводится к решению трех диагональной системы линейных алгебраических уравнений и будет выполнена оценка (10).

В итоге кубический сплайн на отрезках  $[x_{j-1}, x_j], j = 1, \dots, n$  имеет два представления [8]:

$$\begin{aligned} s(x) = & M_{j-1}(x_j - x)^3/(6h_j) + M_j(x - x_{j-1})^3/(6h_j) + \\ & + (f_{j-1} - M_{j-1}h_j^2/6)(x_j - x)/h_j + (f_j - M_jh_j^2/6)(x - x_{j-1})/h_j, \end{aligned} \quad (38)$$

или

$$\begin{aligned} s(x) = & m_{j-1}(x_j - x)^2(x - x_{j-1})/h_j^2 - m_j(x - x_{j-1})^2(x_j - x)/h_j^2 + \\ & + f_{j-1}(x_j - x)^2(2(x - x_{j-1}) + h_j)/h_j^3 + f_j(x - x_{j-1})^2(2(x_j - x) + h_j)/h_j^3, \end{aligned} \quad (39)$$



где

$$M_j = s''(x_j), m_j = s'(x_j), f_j = f(x_j)$$

Для эрмитовых кубических сплайнов в узлах сетки  $\omega_x$  необходимо вычислить  $\hat{f}_i$  и значения производных  $\hat{f}'_i$ . Будем использовать фильтр Савицкого–Голея с кубическими полиномами [6]. В этом случае для построения эрмитовых кубических сплайнов достаточно локальных вычислений. На каждом интервале  $[x_{j-1}, x_j], j = 1, \dots, n$  эти сплайны представимы в виде

$$s(x) = f_{j-1}v((x - x_{j-1})/h_j) + f'_{j-1}w((x - x_{j-1})/h_j) + f_jv((x - x_j)/h_j) + f'_jw((x - x_j)/h_j),$$

где

$$v(x) = (|x| - 1)^2(2|x| + 1); w(x) = x(|x| - 1)^2.$$

### ***Арифметические операции над кусочно-полиномиальными моделями***

Реализация арифметических операций над двумя случайными величинами  $x, y$ , представленных кусочно-полиномиальными моделями, основана на работе с их совместной функцией плотности вероятности  $p(x, y)$ .

В случае, когда случайные величины  $x, y$  являются независимыми и имеют плотности вероятности, представленные кубическими сплайнами,  $S_x S_y$ . Совместную плотность вероятности можно представить в виде произведения  $p(x, y) = s_x s_y$ . Поскольку кубический сплайн на каждом отрезке сетки

представляет кубический полином, то  $p(x,y)$  в случае вычисления интегралов (37)–(39) будет кусочно-полиномиальной функцией шестой степени. Наиболее удобными в этом случае будут квадратуры Гаусса с четырьмя внутренними узлами, которые точны на полиномах седьмой степени [21].

В качестве примера, рассмотрим построение сплайна, аппроксимирующего  $p_{x_1+x_2}$ . Для этих целей в области носителя  $p_{x_1+x_2}$  построим сетку  $\omega = \{x_0, x_1, \dots, x_n\}$  и вычислим значения  $f_i = p_{x_1+x_2}(x_i)$ . Используя значения  $f_i$  и краевые условия  $s'(x_0) = 0$ ,  $s'(x_n) = 0$  на сетке  $\omega$  построим кубический сплайн  $s$ . В этом случае справедлива оценка

$$\|p_{x_1+x_2}^{(\nu)} - s^{(\nu)}\| \leq Kh^{4-\nu} \|p_{x_1+x_2}^{(4)}\|, \nu = 1, 2, 3.$$

Далее вычислим

$$\text{norm} = \int s(x) dx,$$

если  $\text{norm} \neq 1$ , то  $s(x) := s(x)/\text{norm}$ .

В случае, когда случайные величины  $x$ ,  $y$  являются зависимыми, совместную функцию плотности вероятности необходимо вычислять отдельной процедурой.

## 1.4 Вероятностные расширения

Важным понятием численного вероятностного анализа является понятие вероятностного расширения. В рамках гистограммного подхода определим понятие гистограммного расширения. Для этого рассмотрим задачу определения закона распределения функции нескольких случайных аргументов.

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

$$z = f(x_1, \dots, x_n).$$

Тогда плотность вероятности случайной величины  $z$ , будем называть вероятностным расширением функции  $f$ .

На основе понятия вероятностного расширения определим гистограммное вероятностное расширение. Пусть, гистограмма  $F$  определяется сеткой  $\{z_i | i = 0, \dots, n\}$ . Определим область  $\Omega_i = \{(x_1, \dots, x_n) | z_i < f(x_1, \dots, x_n) < z_{i+1}\}$ . Тогда значение гистограммы  $F_i$  на отрезке  $[z_i, z_{i+1}]$  имеет вид [4]

$$F_i = \int_{\Omega_i} p(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n / (z_{i+1} - z_i). \quad (40)$$

Гистограмму  $F$ , построенную по (40), будем называть вероятностным гистограммным расширением  $f$  [9].

Далее построим гистограммные вероятностные расширения для арифметических операций над случайными величинами. Пусть  $P$  – гистограмма плотности вероятности  $z = x * y$ , где  $* \in \{+, -, \cdot, /, \uparrow\}$ . Тогда величина  $P_i$  на интервале  $[z_i, z_{i+1}]$  определяется по формуле

$$P_i = \int_{\Omega_i} p(x_1, x_2) dx_1 dx_2 / (z_{i+1} - z_i),$$

где

$$\Omega_i = \{(x_1, x_2) | z_i \leq x_1 * x_2 \leq z_{i+1}\}.$$

Операция  $\max(x,y)$ . определяется через функцию распределения  $F$

$$F(z) = \int_{-\infty}^z P_x(\xi) d\xi \int_{-\infty}^z P_y(\xi) d\xi.$$

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.

**Теорема 3.** 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  $O(h^\alpha)$ .

Доказательство проведем по индукции. Для  $n = 2$  утверждение справедливо (10). Пусть справедливо для  $n = k$  и гистограмма  $F_k$  аппроксимирует плотность вероятности функции  $f(x_1, \dots, x_k)$  с некоторой точностью  $O(h^\alpha)$ . Покажем, что это справедливо и для  $n = k + 1$ . Действительно,  $F_{k+1} = F_k * x_{k+1}$ , но

$$F_{k+1} - f(x_1, \dots, x_k, x_{k+1}) = (F_k - f(x_1, \dots, x_k)) * x_{k+1} \leq Ch^\alpha * \text{supp} \{x_{k+1}\}.$$

Теорема доказана.

Рассмотрим пример, который иллюстрирует Теорему 1. Пусть функция представлена в следующем виде

$$f(x, y) = xy + x + y + 1 = (x+1)(y+1).$$

Заметим, что только второе представление функции в виде произведения двух сомножителей попадает под условие Теоремы 1 и, следовательно,

естественное гистограммное расширение будет аппроксимировать вероятностное с некоторой точностью  $O(h^\alpha)$ .

Теорема 1 легко обобщается на следующий случай.

**Теорема 4.** Пусть для функции  $f(x_1, \dots, x_n)$  возможна замена переменных, такая что  $f(z_1, \dots, z_k)$  – рациональная функция от переменных  $z_1, \dots, z_k$ , удовлетворяющая условиям теоремы 1 и  $z_i$  – функции от множества переменных  $x_i, i \in Ind_i$ , причем множества  $Ind_i$  попарно не пересекаются. Пусть для каждой  $z_i$  можно построить вероятностное расширение. Тогда естественное расширение  $f(z_1, \dots, z_k)$  будет аппроксимировать вероятностное расширение  $f(x_1, \dots, x_n)$ .

**Пример 1.** Пусть  $f(x_1, x_2) = (-x_1^2 + x_1) \sin x_2$ . Тогда  $z_1 = (-x_1^2 + x_1)$  и  $z_2 = \sin x_2$ . Заметим, что можно построить вероятностные расширения функций  $z_1, z_2$ , где  $f = z_1, z_2$  – рациональная функция, попадающая под условия теоремы 1. Следовательно, ее естественное расширение будет аппроксимировать вероятностное расширение  $f(x_1, x_2)$ .

Рассмотрим случай, когда для  $f(x_1, \dots, x_n)$  необходимо найти вероятностное расширение  $f$ , но не удастся построить замену переменных согласно теореме 2. Пусть для определенности только  $x_1$  встречается несколько раз. Заметим, что если подставить вместо случайной величины  $x_1$  детерминированную  $t$ , то для функции  $f(t, x_2, \dots, x_n)$  можно построить естественное вероятностное расширение. Пусть  $t$  – дискретная случайная величина, аппроксимирующая  $x_1$  следующим образом:  $t$  принимает значения  $t_i$  с вероятностью  $P_i$  и пусть для каждой  $f(t_i, x_2, \dots, x_n)$  можно построить естественное вероятностное расширение  $\varphi_i$ . Тогда вероятностное расширение  $f$  функции  $f(x_1, \dots, x_n)$  можно аппроксимировать плотностью вероятности  $\varphi$  следующим образом:

$$\varphi(\xi) = \sum_{i=1}^n P_i \varphi_i(\xi).$$

**Пример 2.** Пусть  $f(x_1, x_2) = x^2y + x$ , где  $x, y$  – равномерные случайные величины, заданные на  $[0, 1]$ . Заменим  $x$  дискретной случайной величиной  $t$ ,  $\{t_i | t_i = (i - 0.5)/n, i = 1, 2, \dots, n\}$ ,  $P_i = 1/n$ . Далее вычислим естественные вероятностные расширения  $\varphi_i$ . Сравнение  $\varphi$  и  $f$ , вероятностного расширения  $f(x, y)$ , показало хорошее приближение. В таблице 1 приведена норма  $l_2$  разности  $\varphi$  и  $f$  в зависимости от  $n$ . Анализ полученных результатов показал, что в данном примере  $\varphi$  аппроксимирует  $f$  с порядком  $\alpha = 1.4998$ .

Таблица 1 – Погрешность аппроксимации вероятностных расширений

n	$\ f - \varphi\ _2$
10	1.2887825282E-03
20	4.5592973952E-04
40	1.6120775967E-04
80	5.6996092139E-05
16	2.0151185588E-05
0	

Важно отметить, что число арифметических операций для вычисления  $x * y$  имеет порядок  $O(n^2)$ , где  $n$  – размерность сетки гистограмм. Гистограмма, как кусочно-постоянная функция, приближает плотность вероятности с точностью  $O(1/n)$ . Однако, как нетрудно видеть, уже средние точки гистограмм приближают плотность вероятности с точностью  $O(1/n^2)$ . Как показано в (40), нижняя оценка скорости сходимости гистограммных расширений к вероятностным имеет значение  $\alpha = 1$ . Реально это значение несколько выше. Численные эксперименты показали, что в условиях теоремы 1. при правильном согласовании сеток гистограмм и использовании более тонких способов интегрирования, гистограммные расширения имеют скорость сходимости  $\alpha \approx 2$ . Далее сравним полученные результаты с методом Монте-

Карло. Известно, что метод Монте-Карло имеет сходимость  $O(1/\sqrt{N})$ , где  $N$  – число повторов [3]. Гистограммные расширения имеют скорость сходимости  $O(1/n^\alpha)$ .

Пусть необходимо достигнуть точности  $\varepsilon$ , число операций метода Монте-Карло при этом составит  $O(\varepsilon^{-2})$  в сравнении с гистограммными расширениями  $O(\varepsilon^{-2/\alpha})$ , следовательно, гистограммная арифметика при  $\alpha > 1$  эффективней метода Монте-Карло примерно в  $O(\varepsilon^{-2(1-1/\alpha)})$  раз.

Проиллюстрируем эффективность гистограммной арифметики на численном примере.

**Пример 3.** Пусть требуется найти сумму трех равномерно распределенных на  $[0,1]$  случайных величин. Вычисления при  $n = 30$  показали, что ошибка при использовании численных операций гистограммной арифметики в норме  $l_2$  составляет 1.1680E-03. Ошибка метода Монте-Карло при числе бросаний  $N = 10^5$  равна 5.4601E-03, а при числе бросаний  $N = 10^6$  – 2.0284E-03, что сравнимо с погрешностью гистограммной арифметики. При этом число арифметических операций гистограммной арифметики определяется порядком  $30^2$  или  $10^3$ , а метода Монте-Карло величиной порядка  $10^6$ . Таким образом, в данном примере гистограммная арифметика эффективней Монте-Карло примерно в тысячу раз.

Для решения ряда практических задач необходимо решать системы линейных и нелинейных уравнений. Методы их решений рассмотрены в работе [38].

В случае нахождения корня одномерного уравнения  $f(x, k) = 0$ , где  $k$  – случайный вектор параметров. Предположим, что корень локализован на отрезке  $[a, b]$ ,  $\phi_z$  – вероятностное расширение  $f(x, k)$ ,  $z \in [a, b]$ . Тогда,  $P(z)$  есть вероятность, что корень лежит левее (правее) точки  $z$ :

$$P(z) = \int_{-\infty}^0 \phi_z(\xi) d\xi. \quad (41)$$

## 1.5 Основная теорема о вероятностных расширениях

Одной из наиболее важных проблем, с которыми сталкивается ЧВА, является построение вероятности функции плотности случайных величин.

Начнем с общего случая, когда  $(x_1, \dots, x_n)$  - система непрерывных случайных величин с совместными функциями плотности вероятности  $p(x_1, \dots, x_n)$  и случайная величина  $z$  является функцией  $f(x_1, \dots, x_n)$

$$z = f(x_1, \dots, x_n) \quad (42)$$

Под вероятностным расширением функции  $f$  мы имеем в виду функцию плотности вероятности случайной величины  $z$ .

**Определение.** Носитель  $f$ ,  $\text{supp}(f)$ , является множеством точек  $x \in R$ , где  $f$  отлична от нуля

$$\text{supp}(f) = \{x \in R \mid f(x) > 0\}.$$

Рассмотрим процедуру вычисления вероятностных расширений для одномерного случая. Пусть задана функциональная зависимость

$$z = f(x)$$

где  $x$  – случайная величина. Пусть  $\varphi(x)$  – функция плотности вероятности случайной величины  $x$  с носителем  $[\underline{x}, \bar{x}]$ . Далее  $\{x_i(z) \in [\underline{x}, \bar{x}] \mid i = 1, \dots, n\}$  являются корнями уравнения  $z = f(x)$ .

Необходимо найти функцию плотности вероятности  $f$  случайной величины  $z$  как вероятностное расширение  $f(\cdot, x)$  функции  $f(x)$ . Для вычисления вероятностного расширения мы используем



$$f(\xi, x) = \sum_{i=1}^n \frac{x(x_i(\xi))}{|f'(x_i(\xi))|}.$$

**Пример 4.** рассмотрим функцию  $z = x + t$  где  $t$  вещественное число. Тогда вероятностные расширения могут быть представлены как

$$f(\xi, x, t) = x(\xi - t).$$

Если  $z = tx$  и  $t > 0$ , тогда вероятностные расширения могут быть представлены как

$$f(\xi, x, t) = \frac{1}{t} x\left(\frac{\xi}{t}\right).$$

**Теорема 5.** Пусть  $f(\cdot, x_1, x_2, \dots, x_n)$  вероятностное расширение  $f(x_1, x_2, \dots, x_n)$  и для всех вещественных  $t$  функция  $f(\cdot, t, x_2, \dots, x_n)$  есть вероятностное расширение функции  $f(t, x_2, \dots, x_n)$ . Тогда

$$f(z, x_1, x_2, \dots, x_n) = \int x_1(t) f(z, t, x_2, \dots, x_n) dt \quad (43)$$

**Замечание.** Теорема 6 дает возможность рекурсивных вычислений для общей формы вероятностных расширений и редукции к вычислению одномерного случая.

Рассмотрим вычисление интеграла (43). Для простоты мы представляем (43) как квадратуру

$$\int x_1(t) f(z, t, x_2, \dots, x_n) dt \approx \sum_{l=1}^m \gamma_l x_1(t_l) f(z, t_l, x_2, \dots, x_n)$$

Далее, для вычисления  $f(z, t_1, x_2, \dots, x_n)$  мы можем также использовать числовые квадратуры и т. д. В общем, это NP-hard проблема с необходимостью распараллеливания. На нижнем уровне необходимо вычислить вероятностные расширения только для одной переменной. Обратите внимание, что все вычисления на каждом уровне независимы и могут быть вычислены одновременно [33].

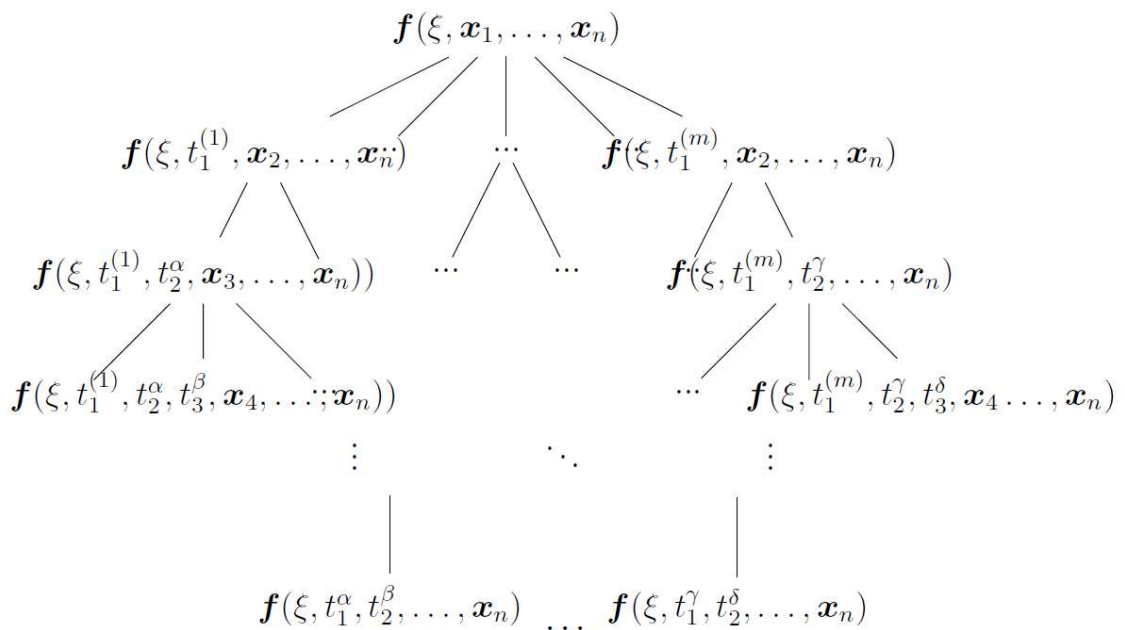


Рисунок 4 – Дерево параллельной рекурсии

На рис. 4 представлено Дерево параллельной рекурсивной, организация вычислительного процесса.

Пусть  $x, y$  система двух непрерывных случайных величин с функциями плотности вероятности  $x, y$ .

Плотности, возникающие в результате арифметических операций над случайными величинами, имеют вид

$$(x + y)(\xi) = \int_{\underline{x}}^{\bar{x}} x(t)y(\xi - t)dt;$$

если  $\text{supp}(x)[\underline{x}, \bar{x}]$  и  $\underline{x} > 0$  тогда

$$(xy)(\xi) = \int_{\underline{x}}^{\bar{x}} \frac{1}{t} x(t) y\left(\frac{\xi}{t}\right) dt,$$

$$(y/x)(\xi) = \int_{\underline{x}}^{\bar{x}} tx(t) y(t\xi) dt.$$

Эти арифметические операции полностью совпадают с классическими операциями над функциями плотности вероятности.

Пусть  $f(x_1, \dots, x_n)$  – рациональная функция. Чтобы построить вероятностные расширения  $f$ , мы заменяем арифметические операции арифметическими операциями над случайные величины, а переменные  $x_1, x_2, \dots, x_n$  заменяются функциями плотности вероятности. Имеет смысл назвать полученные вероятностные расширения  $f$  как естественное вероятностное расширение [33].

**Случай 1.** [7] Пусть  $x_1, \dots, x_n$  независимые случайные переменные. Если  $f(x_1, \dots, x_n)$  является рациональным выражением, где каждая переменная  $x_i$  встречается не более одного раза, то естественное вероятностное расширение аппроксимирует вероятностное расширение.

**Случай 2.** [7] Пусть функция  $f(x_1, \dots, x_n)$  может быть заменой переменных, приведена к виду что  $f(z_1, \dots, z_k)$  является рациональной функцией от переменных  $z_1, \dots, z_k$ , удовлетворяющих условиям случая 1. Переменная  $z_i$  является функцией  $x_i$ ,  $i \in \text{Ind}_i$ . и  $\text{Ind}_i$  взаимно не пересекаются. Предположим, что для каждого  $z_i$  можно построить вероятностное расширение. Тогда естественное вероятностное расширение  $f(z_1, \dots, z_k)$  будет аппроксимировать вероятностное расширение.

**Пример 5.** Пусть  $f(x_1, x_2) = (-x_1^2 + x_1) \sin x_2$  и  $z_1 = (-x_1^2 + x_1)$ ,  $z_2 = \sin x_2$ . Обратите внимание, что можно построить вероятностную расширени для функций  $z_1, z_2$ , а затем вычислить  $f = z_1 * z_2$ , которая является рациональной функцией, удовлетворяющей условиям случая 1. Итак, естественное вероятностное расширение будет аппроксимировать вероятностное расширение для функции  $f(x_1, x_2)$

### *Одномерный случай*

Рассмотрим процедуру вычисления вероятностных расширений для одномерного случая. Дана функциональная зависимость

$$z = f(x)$$

где  $x$  случайная величина.

Пусть  $x$  функция плотности вероятности случайной величины  $x$  на  $[\underline{x}, \bar{x}]$ . В дальнейшем  $\{r_i(z) \in [\underline{x}, \bar{x}] | i = 1, \dots, n\}$  являются корнями уравнения  $z = f(x)$ .

Следуя основному методу (1) легко построить обобщение выражения (2). Мы можем представить вероятностные расширения  $f(\cdot, x)$  функции  $f(x)$  в виде

$$f(\xi, x) = \sum_{i=1}^n \frac{x(r_i(\xi))}{f'(r_i(\xi))}$$

**Пример 6.** В качестве примера рассмотрим построение вероятностной функции расширений

$$f = ax^2 + bx, a \geq 0, b \geq 0,$$

$x$  случайная величина, распределенная по  $[0,2]$  по треугольному закону

$$x(\xi) = \begin{cases} \xi & \text{if } \xi \in [0,1), \\ 2-\xi & \text{if } \xi \in [1,2]. \end{cases}$$

далее

$$r_1(z) = -\frac{\sqrt{4az+b^2}+b}{2a}, r_2(z) = \frac{\sqrt{4az+b^2}-b}{2a}.$$

корни  $z = f(r_i)$ . Выберите положительный корень  $r(\xi) = r_2(\xi) = \frac{\sqrt{4a\xi+b^2}-b}{2a}$ , и

$$[ax^2+bx]' = 2ax+b = \sqrt{4az+b^2}.$$

В итоге

$$f(\xi, x) = x(r(\xi))/\sqrt{4a\xi+b^2}.$$

пусть  $a = 0$  и  $b = 0$

$$f(\xi, x) = \begin{cases} 1/2 & \text{if } \xi \in [0,1), \\ 1/\sqrt{\xi} - 1/2 & \text{if } \xi \in [1,4]. \end{cases}$$

## Численный подход

Рассмотрим численный подход для построения вероятностных расширений функции  $f(x)$ . Для этих целей мы строим  $[\underline{x}, \bar{x}]$  функции плотности вероятности  $x$  сетки  $\{\xi_i | \xi_i \in [\underline{x}, \bar{x}], i = 0, 1, 2, \dots, n\}$  и вычисляем  $\{z_i = f(\xi_i), i = 0, 1, 2, \dots, n\}$ . Далее мы устанавливаем

$$f_z(z_i) = \frac{\mathbf{x}(\xi_i)}{|f'(\xi_i)|}$$

И используя  $(z_i, f_z(z_i))$ , построим кусочно-полиномиальную интерполяцию [39].

Итак, для примера 1.  $[0, 2]$ ,  $\xi_i = i/10, i = 0, 1, \dots, 20$ . С  $a = 1, b = 0$  мы получаем

$$(z_i, f_z(z_i)) = \left( (i/10)^2, \frac{\mathbf{x}(i/10)}{i/5} \right), i = 0, 1, \dots, 20.$$

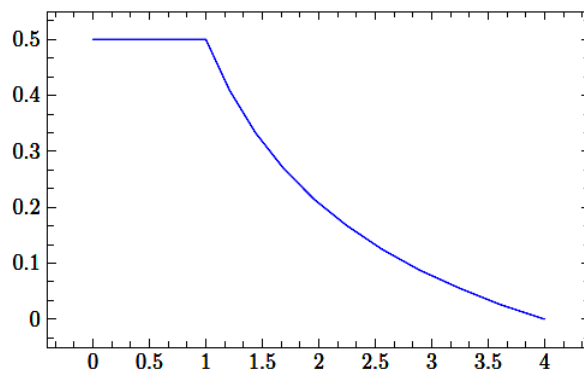


Рисунок 2 – Вероятностные расширения  $f(\cdot, x)$

Рисунок 2 показывает функцию плотности вероятности  $f(\cdot, x)$  из примера 1. В отличие от общего подхода, нет необходимости искать корни уравнения.

## *Двумерный случай*

Пусть  $(x, y)$  будет системой непрерывных случайных величин с совместной функцией плотности вероятности  $p(x, y)$  и случайная величина  $z$  является функцией  $f(x, y)$

$$z = f(x, y)$$

Необходимо найти функцию плотности вероятности случайной величины  $z$ . Определим

$$\Omega_z = \{(x, y) \mid z > f(x, y)\}$$

и  $F_z$  случайная величина  $z$

$$F_z = \int_{\Omega_z} p(x, y) dx dy$$

функция плотности вероятности  $z$

$$f_z = \frac{d}{dz} F_z.$$

Как следствие

$$\frac{d}{dz} F_z = \frac{d}{dz} \int_{\Omega_z} p(x, y) dx dy =$$

$$\lim_{dz \rightarrow 0} \frac{\int_{\Omega_{z+dz}} p(x, y) dx dy - \int_{\Omega_z} p(x, y) dx dy}{dz}$$

$$\lim_{dz \rightarrow 0} \frac{\int_{\Omega_{z+dz} - \Omega_z} p(x, y) dx dy}{dz}.$$

Предположим что  $dx = 0$  тогда

$$dz = f'_y dy$$

и

$$dS = dx dy = dx dz / |f'_y|.$$

Наконец получаем

$$f_z(z) = \int_{\Gamma_z} \frac{p(x, y)}{|f'_y(x, y)|} dx$$

где

$$\Gamma_z = \{(x, y) \mid z = f(x, y)\} = \{(x, y(x))\}.$$

Пусть

$$f_y(\xi, t) = \frac{p(t, y(\xi))}{|f'_y(\xi, y(t))|}$$



тогда  $f(t, y)$  вероятностное расширение.

Таким образом, расчет вероятностного расширения  $f_z$  сводится к вычислению интеграла вероятностных расширений  $f(t, y)$ .

**Пример 7.** Рассмотрим построение вероятностного расширения для функции

$$z = x^2 y + xy^2.$$

Пусть  $t$  реальное, предположим  $y = t$  и вероятностное расширение для функции

$$z_x = x^2 t + xt^2.$$

Это

$$z_x(\xi, t) = \mathbf{x}(\eta(\xi)) / \sqrt{4t\xi + t^4},$$

где

$$\eta(\xi) = (\sqrt{4t\xi + t^4} - b) / (2t).$$

Таким образом, вероятностное расширение для  $z$  можно представить в виде

$$z(\xi) = \int \mathbf{y}(t) z_x(\xi, t) dt = \int \mathbf{y}(t) \mathbf{x}(\eta(\xi)) / \sqrt{4t\xi + t^4} dt. \quad (44)$$

## 2 Техника рекурсивных вычислений

### 2.1 Рекурсивные задачи и вычисления

Понятие рекурсии связано с ситуацией, когда на наличие в определении объекта присутствуют ссылки на сам объект или проявление свойств самоповторения (при этом сколь угодно малая часть объекта подобна всему объекту в целом). Общий случай проявления рекурсивности может быть сформулирован как наличие циклических взаимных обращений в определении объекта, которые в итоге замыкаются на сам объект.

Рекурсивность в постановке задачи проявляется, если решение для общего случая сводится к аналогичным задачам для меньшего количества входных данных. В таком контексте под рекурсией понимают прием последовательного сведения решения некоторой задачи к решению совокупности «более простых» задач такого же класса и получению на этой основе решения исходной задачи.

Рекурсия в широком смысле – это определение объекта посредством ссылки на себя. Рекурсия в программировании – это пошаговое разбиение задачи на подзадачи, подобные исходной.

Рекурсивный алгоритм – это алгоритм, в определении которого содержится прямой или косвенный вызов этого же алгоритма.

В языках программирования процедурной парадигмы предусмотрено использование рекурсивных функций в решении задач [26].

Функция называется рекурсивной, если в своем теле она содержит обращение к самой себе с измененным набором параметров. При этом количество обращений конечно, так как в итоге решение сводится к базовому случаю, когда ответ очевиден.

В программировании выделяют прямую и косвенную рекурсию. Прямая рекурсия предусматривает непосредственное обращение рекурсивной функции к себе, но с иным набором входных данных.

Косвенная (взаимная) рекурсия представляет собой последовательность взаимных вызовов нескольких функций, организованная в виде циклического замыкания на тело первоначальной функции, но с иным набором параметров.

Для решения задач рекурсивными методами разрабатывают следующие этапы, образующие рекурсивную триаду:

- параметризация – выделяют параметры, которые используются для описания условия задачи, а затем в решении;
- база рекурсии – определяют тривиальный случай, при котором решение очевидно, то есть не требуется обращение функции к себе;
- декомпозиция – выражают общий случай через более простые подзадачи с измененными параметрами.

Целесообразность применения рекурсии в программировании обусловлена спецификой задач, в постановке которых явно или опосредовано указывается на возможность сведения задачи к подзадачам, аналогичным самой задаче. При этом эффективность рекурсивного или итерационного способов решения одной и той же задачи определяется в ходе анализа работоспособности программы на различных наборах данных. Таким образом, рекурсия не является универсальным способом в программировании. Ее следует рассматривать как альтернативный вариант при разработке алгоритмов решения задач [20].

Повысить эффективность рекурсивных алгоритмов часто представляется возможным за счет пересмотра этапов триады. Например, введение дополнительных параметров, не оговоренных в условии задачи, в реализации декомпозиции могут быть применены другие соотношения, а также можно организовать расширение базовых случаев с сохранением промежуточных результатов. Область памяти, предназначенная для хранения всех

промежуточных значений локальных переменных при каждом следующем рекурсивном обращении, образует рекурсивный стек. Для каждого текущего обращения формируется локальный слой данных стека (при этом совпадающие идентификаторы разных слоев стека независимы друг от друга и не отождествляются). Завершение вычислений происходит посредством восстановления значений данных каждого слоя в порядке, обратном рекурсивным обращениям. В силу подобной организации количество рекурсивных обращений ограничено размером области памяти, выделяемой под программный код. При заполнении всей предоставленной области памяти попытка вызова следующего рекурсивного обращения приводит к ошибке переполнения стека.

Примером рекурсивной задачи является Задача о Ханойских башнях.

Ханойская башня является одной из популярных головоломок XIX века. Даны три стержня, на один из которых нанизаны  $n$  колец, причем кольца отличаются размером и лежат меньшее на большем. Задача состоит в том, чтобы перенести пирамиду из  $n$  колец за наименьшее число ходов с одного стержня на другой. За один раз разрешается переносить только одно кольцо, причём нельзя класть большее кольцо на меньшее.

Существует древнеиндийская легенда, согласно которой в городе Бенаресе под куполом главного храма, в том месте, где находится центр Земли, на бронзовой площадке стоят три алмазных стержня. В день сотворения мира на один из этих стержней было надето 64 кольца. Бог поручил жрецам перенести кольца с одного стержня на другой, используя третий в качестве вспомогательного. Жрецы обязаны соблюдать условия: переносить за один раз только одно кольцо; кольцо можно класть только на кольцо большего размера или на пустой стержень.

Согласно легенде, когда, соблюдая все условия, жрецы перенесут все 64 кольца, наступит конец света. Для 64 колец это 18 446 744 073 709 551 615

перекладываний, и, если учесть скорость одно перекладывание в секунду, получится около 584 542 046 091 лет, то есть апокалипсис наступит нескоро.

Задача состоит в том, чтобы определить последовательность минимальной длины переноса колец. Решением задачи будем считать последовательность допустимых переносов, каждый из которых имеет вид:  $A \rightarrow B$ ,  $A \rightarrow C$ ,  $B \rightarrow A$ ,  $B \rightarrow C$ ,  $C \rightarrow A$ ,  $C \rightarrow B$ . Если кольцо всего одно, то задача решается за один перенос  $A \rightarrow B$ . Для перемещения двух колец требуется выполнить три действия:  $A \rightarrow C$ ,  $A \rightarrow B$ ,  $C \rightarrow B$ . Решение задачи для трех колец содержит семь действий, для четырех – 15. Напишем рекурсивную функцию, которая находит решение для произвольного числа колец. Параметризация. Функция имеет четыре параметра, первый параметр – число переносимых колец, второй параметр – стержень, на который первоначально нанизаны кольца. Третий параметр функции – стержень, на который требуется перенести кольца, и, наконец, четвертый параметр – стержень, который разрешено использовать в качестве вспомогательного. База рекурсии. Перенос одного стержня. Декомпозиция. Последовательность переноса колец.

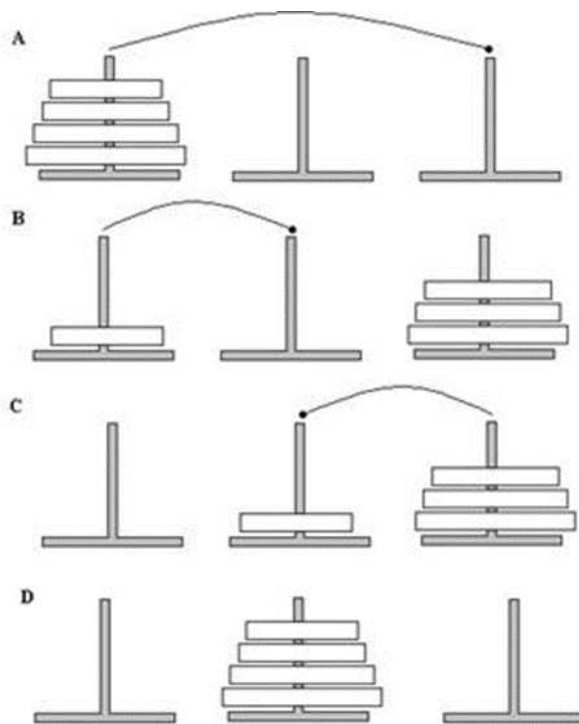


Рисунок 3 – Схема решения задачи о Ханойских башнях для четырех колец

Ключевыми терминами являются

База рекурсии – это тривиальный случай, при котором решение задачи очевидно, то есть не требуется обращение функции к себе.

Декомпозиция – это выражение общего случая через более простые подзадачи с измененными параметрами.

Косвенная (взаимная) рекурсия – это последовательность взаимных вызовов нескольких функций, организованная в виде циклического замыкания на тело первоначальной функции, но с иным набором параметров.

Параметризация – это выделение из постановки задачи параметров, которые используются для описания условия задачи и решения.

Прямая рекурсия – это непосредственное обращение рекурсивной функции к себе, но с иным набором входных данных.

Рекурсивная триада – это этапы решения задач рекурсивным методом.

Рекурсивная функция – это функция, которая в своем теле содержит обращение к самой себе с измененным набором параметров.

Рекурсивный алгоритм – это алгоритм, в определении которого содержится прямой или косвенный вызов этого же алгоритма.

Рекурсивный стек – это область памяти, предназначенная для хранения всех промежуточных значений локальных переменных при каждом следующем рекурсивном обращении.

Рекурсия в программировании – это пошаговое разбиение задачи на подзадачи, подобные исходной.

Рекурсия в широком смысле – это определение объекта посредством ссылки на себя.

Целесообразность применения рекурсии в программировании обусловлена спецификой задач, в постановке которых явно или опосредованно указывается на возможность сведения задачи к подзадачам, аналогичным самой задаче. При этом эффективность рекурсивного или итерационного способов решения одной и той же задачи определяется в ходе анализа работоспособности

программы на различных наборах данных. Таким образом, рекурсия не является универсальным способом в программировании. Ее следует рассматривать как альтернативный вариант при разработке алгоритмов решения задач [35].

Повысить эффективность рекурсивных алгоритмов часто представляется возможным за счет пересмотра этапов триады. Например, введение дополнительных параметров, не оговоренных в условии задачи, в реализации декомпозиции могут быть применены другие соотношения, а также можно организовать расширение базовых случаев с сохранением промежуточных результатов.

Область памяти, предназначенная для хранения всех промежуточных значений локальных переменных при каждом следующем рекурсивном обращении, образует рекурсивный стек. Для каждого текущего обращения формируется локальный слой данных стека (при этом совпадающие идентификаторы разных слоев стека независимы друг от друга и не отождествляются). Завершение вычислений происходит посредством восстановления значений данных каждого слоя в порядке, обратном рекурсивным обращениям [31]. В силу подобной организации количество рекурсивных обращений ограничено размером области памяти, выделяемой под программный код. При заполнении всей предоставленной области памяти попытка вызова следующего рекурсивного обращения приводит к ошибке переполнения стека.

## **2.2 Рекурсивные функции**

По сути индукция, рекурсия и рекуррентные соотношения – это один и тот же метод, применительно к различным областям и носит различные названия

Под индукцией понимается метод доказательства утверждений, который строится на базе индукции при  $n=0,1$ , затем утверждение полагается правильным при  $n=n$  и проводится доказательство для  $n+1$ .

Под рекурсией понимается метод определения функции через её предыдущие и ранее определенные значения, а также способ организации вычислений, при котором функция вызывает сама себя с другим аргументом.

Термин рекуррентные соотношения связан с американским научным стилем и определяет математическое задание функции с помощью рекурсии.

Основной задачей исследования рекурсивно заданных функций является получение  $f(n)$  в явной или как еще говорят «замкнутой» форме, т.е. в виде аналитически заданной функции.

Рекурсивная реализация алгоритмов состоит в следующем.

Большинство современных языков высокого уровня поддерживают механизм рекурсивного вызова, когда функция, как элемент структуры языка программирования, возвращающая вычисленное значение по своему имени, может вызывать сама себя с другим аргументом. Эта возможность позволяет напрямую реализовывать вычисление рекурсивно определенных функций. Отметим, что в силу тезиса Черча–Тьюринга аппарат рекурсивных функций Черча равномошен машине Тьюринга, и, следовательно, любой рекурсивный алгоритм может быть реализован итерационно [23].

Анализ трудоемкости рекурсивных реализаций алгоритмов, очевидно, связан как с количеством операций, выполняемых при одном вызове функции, так и с количеством таких вызовов. Графическое представление порождаемой данным алгоритмом цепочки рекурсивных вызовов называется деревом рекурсивных вызовов. Более детальное рассмотрение приводит к необходимости учета затрат как на организацию вызова функции и передачи параметров, так и на возврат вычисленных значений и передачу управления в точку вызова.



Можно заметить, что некоторая ветвь дерева рекурсивных вызовов обрывается при достижении такого значения передаваемого параметра, при котором функция может быть вычислена непосредственно. Таким образом, рекурсия эквивалентна конструкции цикла, в котором каждый проход есть выполнение рекурсивной функции с заданным параметром.

Дерево рекурсивных вызовов может иметь и более сложную структуру, если на каждом вызове порождается несколько обращений.

Механизм вызова функции или процедуры в языке высокого уровня существенно зависит от архитектуры компьютера и операционной системы. В рамках IBM PC совместимых компьютеров этот механизм реализован через программный стек. Как передаваемые в процедуру или функцию фактические параметры, так и возвращаемые из них значения помещаются в программный стек специальными командами процессора. Дополнительно сохраняются значения необходимых регистров и адрес возврата в вызывающую процедуру.

Анализ трудоемкости рекурсивных алгоритмов в части трудоемкости самого рекурсивного вызова можно выполнять разными способами в зависимости от того, как формируется итоговая сумма элементарных операций.

### 3 Разработка программного модуля

#### 3.1 Постановка задачи и необходимые теоретические сведения

Рассмотрим задачу восстановления функции плотности вероятности случайной величины  $\xi$  суммы  $n=4,6$  равномерно распределенных на  $[0,1]$  случайных величин с использованием теории вероятностных расширений. Как известно суммы равномерно распределенных на  $[0,1]$  случайных величин имеет функцию плотности вероятности распределения Ирвина-Холла.

Для дальнейших исследований приведем необходимые теоретические сведения

#### *Распределение Ирвина-Холла*

Распределение Ирвина-Холла, названное в честь Джозефа Оскара Ирвини и Филипп Холла, является распределением вероятностей для случайной величины, определяемой как сумма ряда независимых случайных величин, каждая из которых имеет равномерное распределение. По этой причине оно также известно, как равномерное распределение суммы.

Генерация псевдослучайных чисел, имеющих приблизительно нормальное распределение, иногда выполняется путем вычисления суммы числа псевдослучайных чисел, имеющих равномерное распределение; обычно ради простоты программирования. Изменение масштаба распределения Ирвина-Холла обеспечивает точное распределение генерируемых случайных величин.

Распределение Ирвина – Холла - это непрерывное распределение вероятностей для суммы  $n$  независимых и одинаково распределенных  $U(0, 1)$  случайных величин:

$$X = \sum_{k=1}^n U_k$$

Функция плотности вероятности (кусочно-полиномиальная функция) определяется как

$$f_x(x; n) = \frac{1}{2(n-1)!} \sum_{k=0}^n (-1)^k \binom{n}{k} (x-k)^{n-1} \operatorname{sgn}(x-k)$$

где  $\operatorname{sgn}(x-k)$  обозначает функцию знака :

$$\operatorname{sgn}(x-k) = \begin{cases} -1 & x < k \\ 0 & x = k \\ 1 & x > k. \end{cases}$$

Таким образом, кусочно-полиномиальная функция является сплайном степени  $n-1$  над узлами  $0, 1, \dots, n$ . Фактически, для  $x$  между узлами, расположенными в  $k$  и  $k+1$ , кусочно-полиномиальная функция равен

$$f_x(x; n) = \frac{1}{(n-1)!} \sum_{j=0}^{n-1} a_j(k, n) x^j$$

где коэффициенты  $a_j(k, n)$  могут быть найдены из рекуррентного соотношения по  $k$

$$a_j(k, n) = \begin{cases} 1 & k=0, j=n-1 \\ 0 & k=0, j < n-1 \\ a_j(k-1, n) + (-1)^{n+k-j-1} \binom{n}{k} \binom{n-1}{j} k^{n-j-1} & k > 0 \end{cases}$$

Коэффициенты также A188816 в OEIS . Коэффициенты для накопительного распределения A188668 .

Среднее и дисперсия являются  $n / 2$  и  $n / 12$ , соответственно.

- Для  $n = 1$   $X$  следует равномерному распределению :

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

- Для  $n = 2$   $X$  следует треугольному распределению :

$$f_X(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 \leq x \leq 2 \end{cases}$$

- Для  $n = 3$

$$f_X(x) = \begin{cases} \frac{1}{2}x^2 & 0 \leq x \leq 1 \\ \frac{1}{2}(-2x^2 + 6x - 3) & 1 \leq x \leq 2 \\ \frac{1}{2}(x - 3)^2 & 2 \leq x \leq 3 \end{cases}$$

- Для  $n = 4$

$$f_X(x) = \begin{cases} \frac{1}{6}x^3 & 0 \leq x \leq 1 \\ \frac{1}{6}(-3x^3 + 12x^2 - 12x + 4) & 1 \leq x \leq 2 \\ \frac{1}{6}(3x^3 - 24x^2 + 60x - 44) & 2 \leq x \leq 3 \\ \frac{1}{6}(x - 4)^3 & 3 \leq x \leq 4 \end{cases} \quad (45)$$

- Для  $n = 5$

$$f_X(x) = \begin{cases} \frac{1}{24}x^4 & 0 \leq x \leq 1 \\ \frac{1}{24}(-4x^4 + 20x^3 - 30x^2 + 20x - 5) & 1 \leq x \leq 2 \\ \frac{1}{24}(6x^4 - 60x^3 + 210x^2 - 300x + 155) & 2 \leq x \leq 3 \\ \frac{1}{24}(-4x^4 + 60x^3 - 330x^2 + 780x - 655) & 3 \leq x \leq 4 \\ \frac{1}{24}(x - 5)^4 & 4 \leq x \leq 5 \end{cases}$$

Заметим, что функция плотности вероятности суммы  $n$  равномерно распределенных случайных величин может быть записана в виде

$$p_n(x) = \frac{1}{(n-1)!} (x^{n-1} - C_n^1(x-1)^{n-1} + C_n^2(x-2)^{n-1} - \dots) \quad (46)$$

где  $C_n^k$  коэффициенты биномиального разложения. При каждом фиксированном значении аргумента  $x$  суммы в скобках приведены только для тех членов, для которых величина  $(x - k)$ ,  $k = 1, 2, \dots, n$  неотрицательна.

Рассмотрим пример, когда  $n=6$ . Тогда функция плотности вероятности имеет аналитическое представление в виде

$$p(x) = \begin{cases} x^5/120, & x \in [0, 1]; \\ (-5x^5 + 30x^4 - 60x^3 + 60x^2 - 30x + 6)/120, & x \in [1, 2]; \\ (10x^5 - 120x^4 + 540x^3 - 1140x^2 + 1170x - 474)/120, & x \in [2, 3]; \\ (-10x^5 + 180x^4 - 1260x^3 + 4260x^2 - 6930x + 4386)/120, & x \in [3, 4]; \\ (5x^5 - 120x^4 + 1140x^3 - 5340x^2 + 12270x - 10974)/120, & x \in [4, 5]; \\ (-x^5 + 30x^4 - 360x^3 + 2160x^2 - 6480x + 7776)/120, & x \in [5, 6]. \end{cases}$$

## Численное интегрирование. Квадратурные формулы

Для нахождения численных значений интегралов вида

$$I = \int_a^b f(x)dx. \quad (47)$$

Введем понятие квадратурной формулы.

**Определение.** Приближенное равенство

$$\int_a^b f(x)dx \approx \sum_{j=1}^n q_j f(x_j), \quad (48)$$

где  $q_j$  – некоторые числа, называемые *весами*,  $x_j$  – некоторые точки отрезка  $[a, b]$ , называется квадратурной формулой.

Говорят, что квадратурная формула точна для многочленов степени  $m$ , если при замене  $f$  на произвольный алгебраический многочлен степени  $m$  приближенное равенство (48) становится точным.

Одним из способов построения квадратурных формул является замена подынтегральной функции  $f$  ее некоторым приближением или аппроксимацией, от которой достаточно легко вычислить интеграл. Чем точнее мы найдем аппроксимацию, тем точнее мы вычислим интеграл.

### Формула прямоугольников

В этой формуле мы аппроксимируем нашу функцию  $f$  постоянным значением, взятым в середине интервала:

$$\int_a^b f(x)dx \approx (b-a)f((b-a)/2). \quad (49)$$

Интегрируя разложение функции  $f$  в ряд Тейлора, несложно убедиться, что

$$\int_a^b f(x)dx = (b-a)f((b-a)/2) + \frac{(b-a)^3}{24} f''(\xi), \xi \in [a,b], \quad (50)$$

т.е. мы получили *формулу прямоугольников с остаточным членом*.

### Формула трапеций

Для построения этой формулы мы заменяем функцию  $f$  полиномом Лагранжа первой степени, или, другими словами, линейной интерполяцией

$$\int_a^b f(x)dx \approx (b-a) \frac{f(a) + f(b)}{2}. \quad (51)$$

Остаточный член для формулы трапеций записывается в следующем виде:  $(b-a)^3/12f''(\xi)$ .

### Формула Симпсона

Следуя рассмотренной выше схеме построения квадратурных формул, перейдем от линейной интерполяции к интерполяционному многочлену Лагранжа второй степени. Как было показано выше, для его построения необходимо знать значения функции в трех точках. Пусть это точки  $x_{-1}, x_0, x_1$  и значения функции соответственно  $f_{-1}, f_0, f_1$ , далее для простоты

предположим  $h = x_0 - x_{-1} = x_1 - x_0$ . Тогда несложно убедиться, что интерполяционный многочлен Лагранжа второй степени задается уравнением

$$y = f_0 + \frac{f_1 - f_0}{2h} x + \frac{f_{-1} - 2f_0 + f_1}{2h^2} x^2.$$

Отсюда легко находим

$$\int_{-h}^h y dx = h(f_{-1} + 4f_0 + f_1)/3.$$

Таким образом, формула Симпсона, называемая также формулой парабол, имеет вид

$$\int_{-h}^h f(x) dx \approx h(f_{-1} + 4f_0 + f_1)/3. \quad (52)$$

Остаточный член для формулы Симпсона записывается в следующем виде:  $h^5/90f^4(\xi)$ .

### ***Усложненные квадратурные формулы***

Как видно из приведенных остаточных членов, при больших интервалах  $[a, b]$  точность квадратурных формул может быть низкой. Для повышения точности применяется следующий прием: интервал интегрирования разбивается на подинтервалы и на каждом подинтервале применяется своя квадратурная формула. Остановимся подробнее на применении формулы



трапеций. Разобьем интервал интегрирования на подинтервалы  $[x_i, x_{i+1}]$  таким образом, что

$$a = x_0 < x_1 < x_2 \dots < x_n = b, h = x_{i+1} - x_i.$$

Имеем

$$\int_{x_i}^{x_{i+1}} f(x) dx = h(f_i + f_{i+1})/2 + h^3/12 f''(\xi). \quad (53)$$

Суммируя по всем подинтервалам, получаем *усложненную формулу трапеций*

$$\int_a^b f(x) dx \approx h(f_0/2 + f_1 + f_2 + \dots + f_{n-1} + f_n/2) = I_h^{mp} \quad (54)$$

с остаточным членом  $h^2 \frac{(b-a)}{12} f''(\xi)$ . Аналогично получается и усложненная формула Симпсона

$$\int_a^b f(x) dx \approx h(f_0 + 4f_1 + 2f_2 + 4f_3 \dots + 4f_{2n-1} + f_{2n})/3 = I_h^C \quad (55)$$

с остаточным членом  $h^4 \frac{(a-b)}{180} f^4(\xi)$ .

Из выражений остаточных членов для квадратурных формул видно, что формулы прямоугольников и трапеций точны для многочленов первой степени, формула Симпсона – для многочленов третьей степени.

Погрешность формулы трапеций и формулы Симпсона удовлетворяет неравенствам

$$|I - I_h^{np}| \leq h^2 \frac{(b-a)}{12} \max_{[a,b]} |f''(x)|,$$

$$|I - I_h^C| \leq h^4 \frac{(b-a)}{180} \max_{[a,b]} |f^{(4)}(x)|.$$

### 3.2 Структура модуля и описание его основных блоков

Модуль моделирования задач со случайными входными данными предназначен для решения практических задач, в которых необходимо вычислять различные функциональные зависимости от случайных аргументов. Основу модуля составляют пять основных блоков и три вспомогательных.

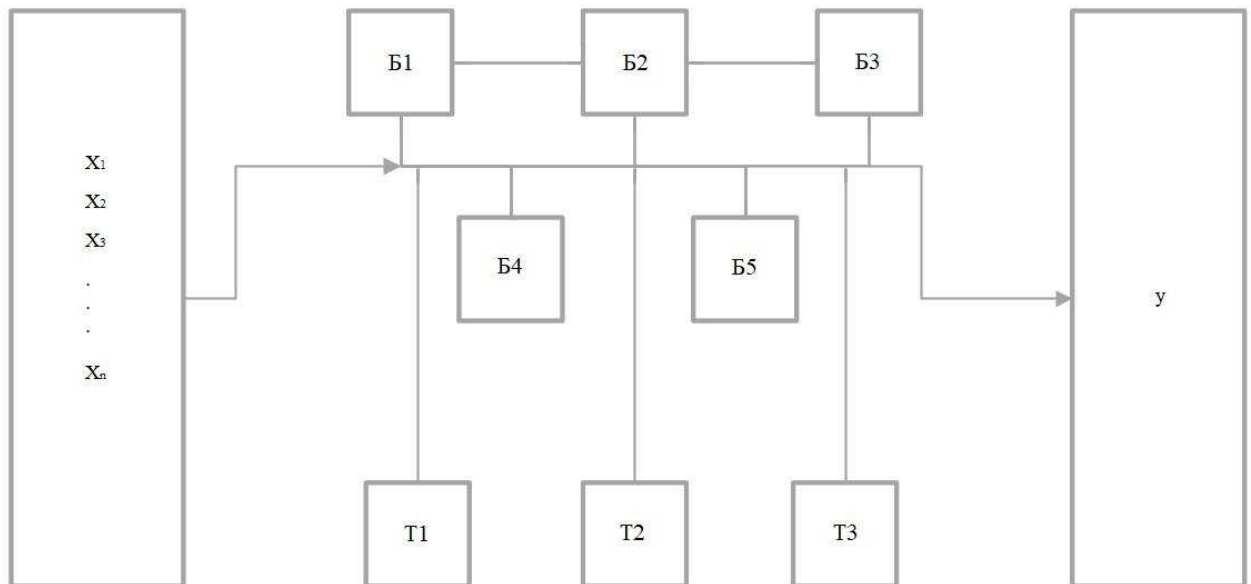


Рисунок 5 – Структура программного модуля

Основные блоки модуля:

Б1 – Построение кусочно-полиномиальной модели.

Б2 – Построение вероятностных расширений.

Б3 – Организация и вычисления рекурсий.

Б4 – Анализ результатов.

Б5 – Блок визуализации модуля.

Вспомогательные блоки:

Т1 – Распределение Ирвина-Холла.

Т2 – Численное интегрирование. Квадратурные формулы.

Т3 – Вероятностная арифметика.

### **3.3 Организация работы программного модуля**

Описание работы программного модуля представлена в виде пяти блок – схем.

На рисунке 6 представлено основное тело программы.

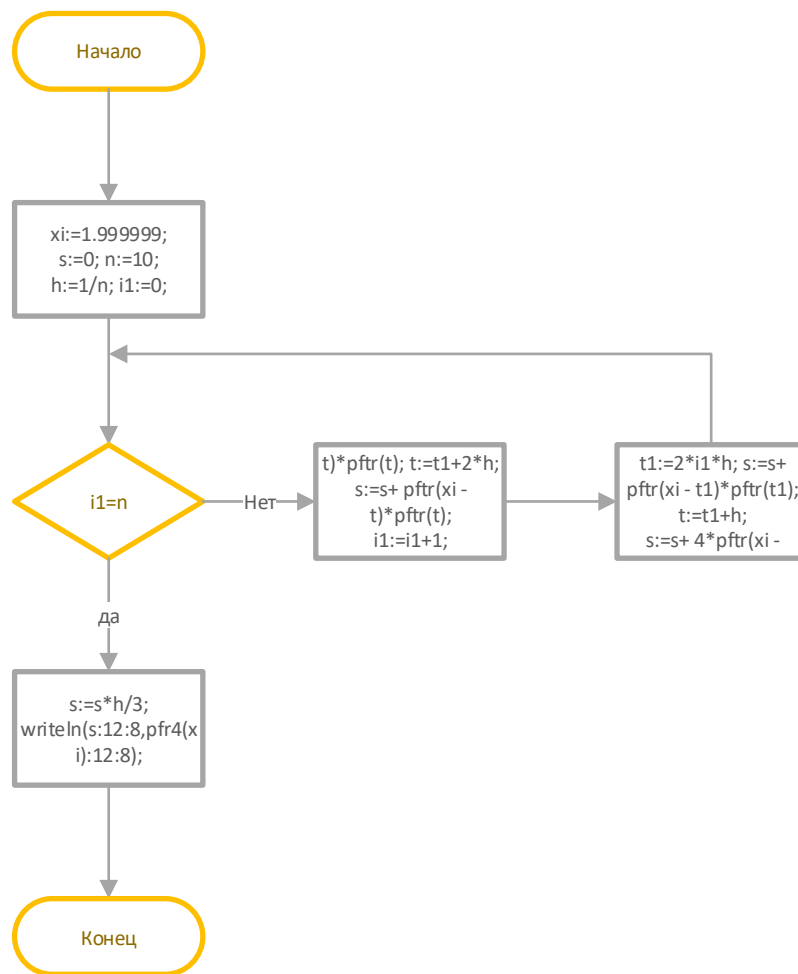


Рисунок 6 – Основное тело программы

На рисунке 7 представлена функция распределения случайной величины.

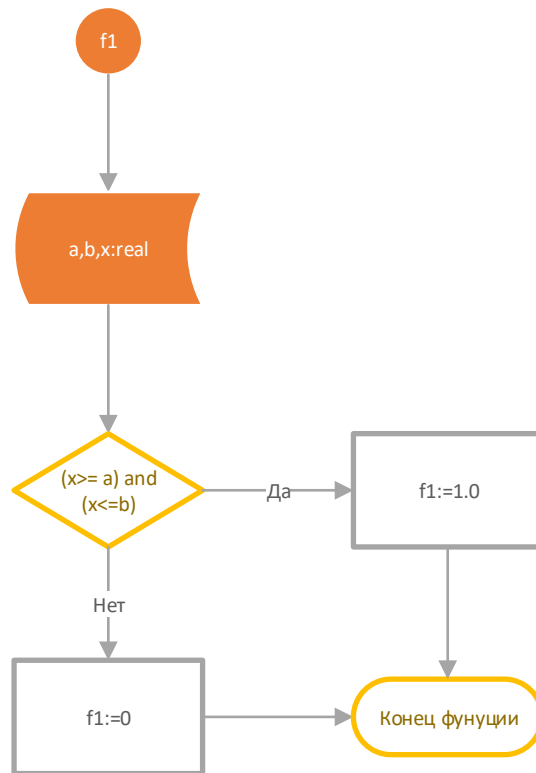


Рисунок 7 – Функция распределения случайной величины

На рисунке 8 приводится блок-схема вычисления интеграла с использованием квадратурных формул на примере формулы Симпсона.

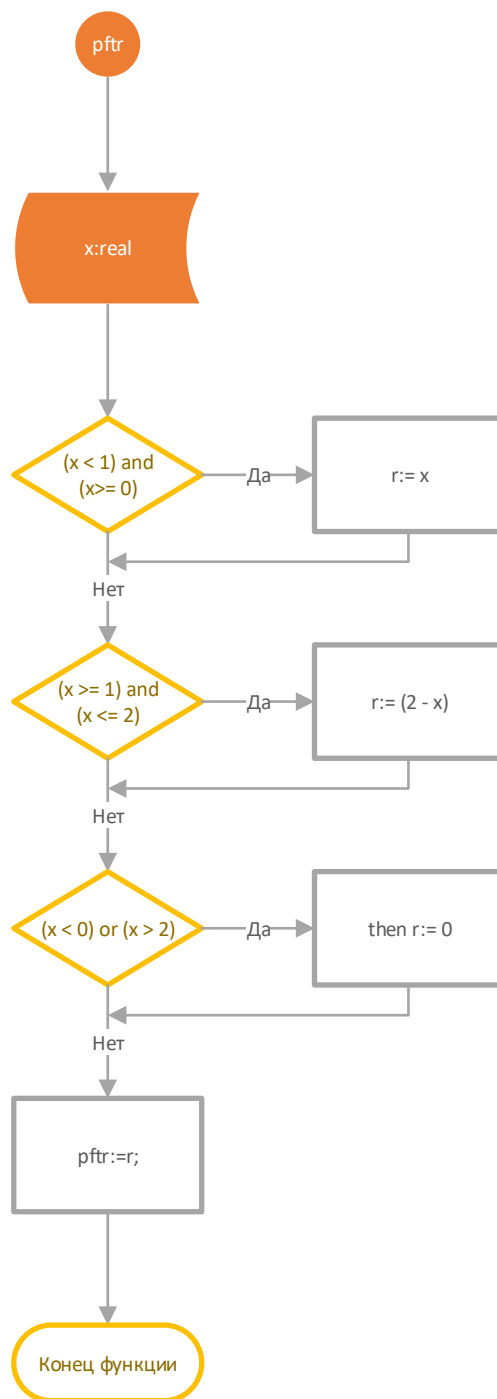


Рисунок 8 – Блок-схема вычисления интеграла

На рисунке 9 показана функция распределения Ирвин-Холла для  $n=4$ .

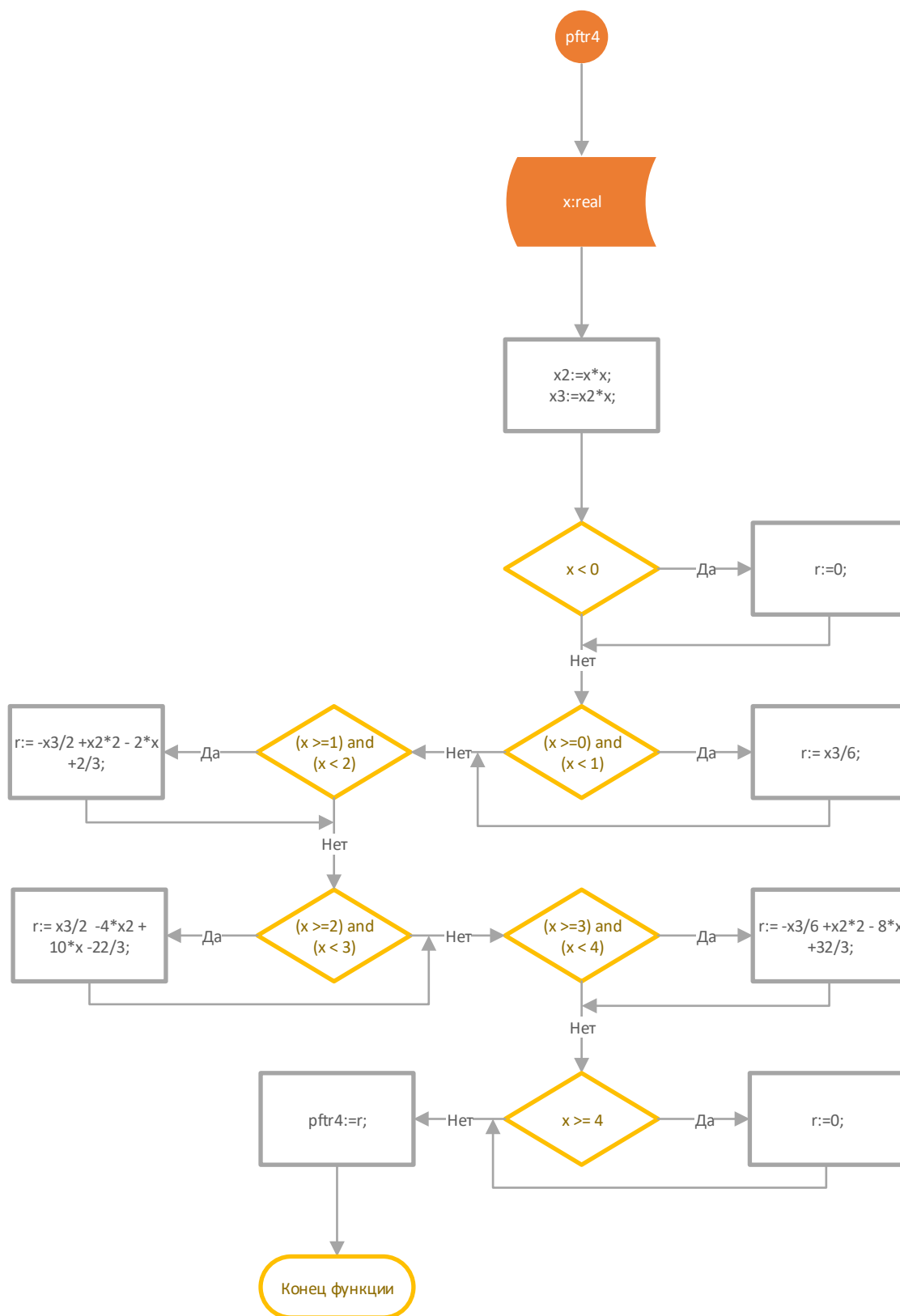


Рисунок 9 – Функция распределения Ирвина-Холла для n=4

На рисунке 10 приводится блок-схема вычисления интеграла методом прямоугольника.

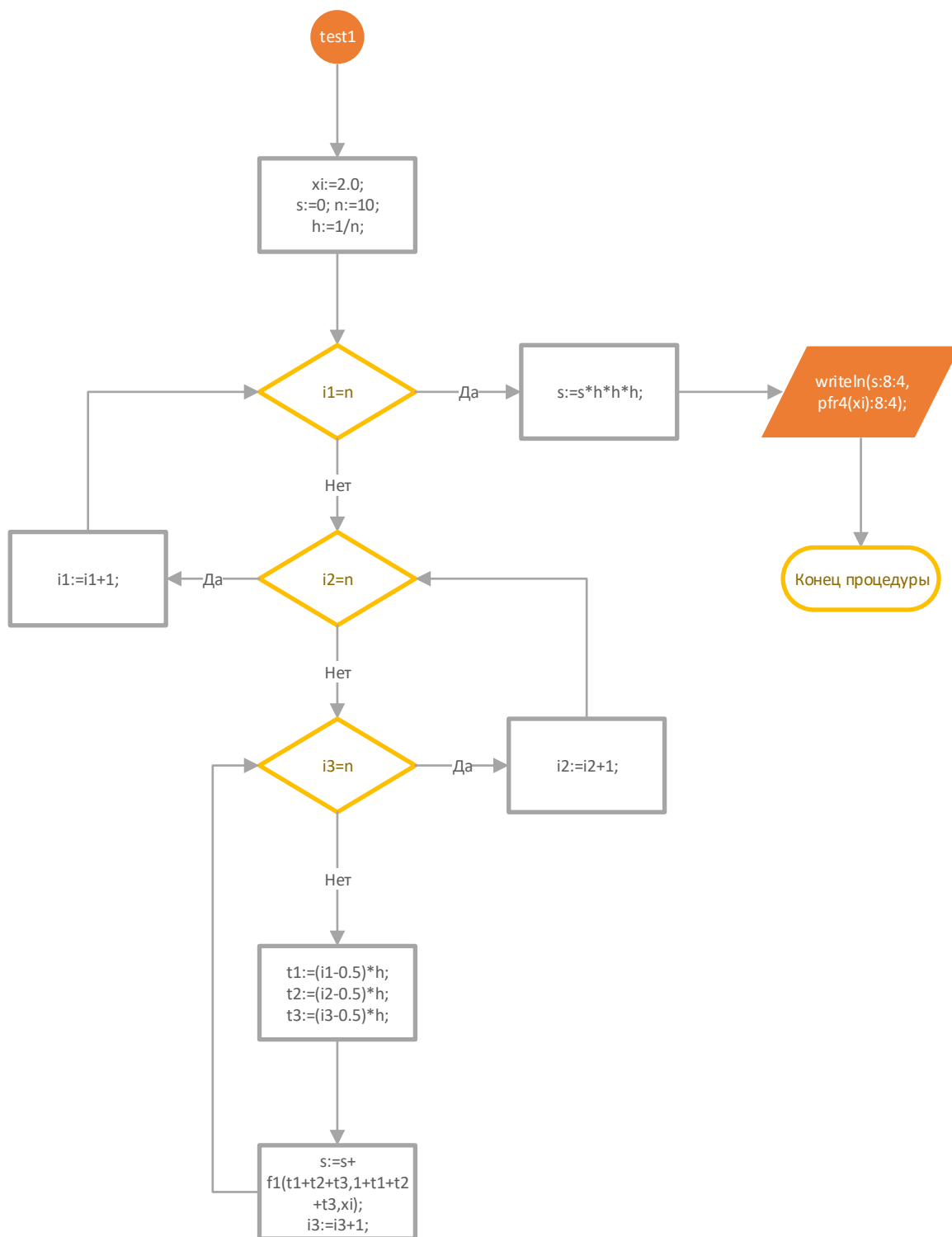


Рисунок 10 – Процедура интегрирования методом прямоугольника



### 3.4 Распараллеливание программы на языке PascalABC

Распараллеливание программ – процесс адаптации алгоритмов, записанных в виде программ, для их эффективного исполнения на вычислительной системе параллельной архитектуры (в последнее время, как правило, на многопроцессорной вычислительной системе). Заключается либо в переписывании программ на специальный язык, описывающий параллелизм и понятный трансляторам целевой вычислительной системы, либо к вставке специальной разметки (например, инструкций MPI или OpenMP).

OpenMP – открытый стандарт для распараллеливания программ на многопроцессорных системах с общей памятью (например, на многоядерных процессорах).

OpenMP реализует параллельные вычисления с помощью многопоточности: главный поток создает набор подчиненных потоков, и задача распределяется между ними.

OpenMP представляет собой набор директив компилятора, которые управляют процессом автоматического выделения потоков и данными, требуемыми для работы этих потоков.

В системе PascalABC.NET реализованы следующие элементы OpenMP:

- Конструкции для создания и распределения работы между потоками (директивы `parallel for` и `parallel sections`)

- Конструкции для синхронизации потоков (директива `critical`)

Директивы имеют следующий вид:

```
{ $omp directive-name [опция[[,] опция]...]}
```

Здесь `$omp` означает то, что это директива OpenMP, `directive-name` – имя директивы, например `parallel`, после чего могут быть опции. Директива относится к тому оператору, перед которым она находится.

***Программный код:***

```
program recurs00;
Var xi,s,t,t1,t2,t3,h : real;
    i,i1,i2,i3,n : integer;

function f1(a,b,x:real):real;
begin
if (x>= a) and (x<=b) then f1:=1.0 else f1:=0;
end;

function pfr(x:real):real;
Var r, t : real;
    i,j : integer;
begin
if (x < 1) and (x>= 0) then r:= x;
if (x >= 1) and (x <= 2) then r:= (2 - x);
if (x < 0) or (x > 2) then r:= 0;
pfr:=r;
end;

function pfr4(x:real):real;
Var r, t, x2,x3 : real;
    i,j : integer;
begin

x2:=x*x;
x3:=x2*x;
```

```

if x < 0 then r:=0;
if (x >=0) and (x < 1) then r:= x3/6;
if (x >=1) and (x < 2) then r:= -x3/2 +x2*2 - 2*x +2/3;
if (x >=2) and (x < 3) then r:= x3/2 -4*x2 + 10*x -22/3;
if (x >=3) and (x < 4) then r:= -x3/6 +x2*2 - 8*x +32/3;
if x >= 4 then r:=0;

```

```
pfr4:=r;
```

```
end;
```

```
procedure test1;
```

```
Var n : integer;
```

```
begin
```

```
xi:=2.0;
```

```
s:=0;
```

```
n:=10;
```

```
h:=1/n;
```

```
{ $omp parallel for }
```

```
for var i1:=1 to n do
```

```
for var i2:=1 to n do
```

```
for var i3:=1 to n do
```

```
begin
```

```
t1:=(i1-0.5)*h;
```

```
t2:=(i2-0.5)*h;
```

```
t3:=(i3-0.5)*h;
```

```
s:=s+ f1(t1+t2+t3,1+t1+t2+t3,xi);
```

```
end;
```

```
s:=s*h*h*h;
```

```
writeln(s:8:4,pfr4(xi):8:4);
```

```
end;
```

```

procedure test2;
begin
var i1: integer;
xi:=1.999999;
s:=0;
n:=10;
h:=1/n;
for i1:=0 to n do
begin
t1:=2*i1*h;
s:=s+ pftr(xi - t1)*pftr(t1);
t:=t1+h;
s:=s+ 4*pftr(xi - t)*pftr(t);
t:=t1+2*h;
s:=s+ pftr(xi - t)*pftr(t);
end;
s:=s*h/3;
writeln(s:12:8,pfr4(xi):12:8);
end;

begin
test1;
test2;
end.

```

### **3.5 Тестирование работы программного модуля**

В качестве тестового примера рассматривается задача вычисления функции плотности вероятности случайной величины суммы  $n=4,6$  равномерно

распределенных на  $[0,1]$  случайных величин с использованием теории вероятностных расширений. Как известно суммы равномерно распределенных на  $[0,1]$  случайных величин имеет функцию плотности вероятности распределения Ирвина-Холла.

Плотность вероятности равномерно распределенных на  $[0,1]$  случайных величин имеет вид

$$p = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Заметим, что операцию сложения случайной величины с некоторой вещественной переменной можно реализовать следующим способом. Пусть  $f(x)$  плотность вероятности случайной величины  $X$ , тогда  $f_r$  плотность вероятности случайной величины  $X + r$ :

$$f_r(x) = f(x - r).$$

Последнее равенство означает, что носитель случайной величины  $X + r$  сдвинут на  $r$ .

Если  $X$  равномерно распределенная на  $[0,1]$  случайная величина, то плотность вероятности случайной величины  $X + r$

$$p_r = \begin{cases} 1 & x \in [0+r, 1+r] \\ 0 & x \notin [0+r, 1+r] \end{cases}$$

Рассмотрим программную реализацию вычисления плотности вероятности  $S_m$  суммы  $m$  равномерно распределенных на  $[0,1]$  случайных величин. Согласно теореме о вероятностных расширениях вычисление

плотности вероятности  $S_m(\xi)$  можно свести к вычислению повторного интеграла. Для вычисления повторных интегралов будем использовать численные квадратуры. В данном случае – метод прямоугольников. Реализация численных квадратур выражается в виде вложенных  $m$  циклов. На нижнем слое вычисляются значения функции  $f_t(\xi)$ , где  $t = t_1 + t_2 + \dots + t_m$ .

**Пример 8.** Рассмотрим пример для вычисления суммы равномерно распределенных случайных величин на интервале  $[0,1]$

$$t = t_1 + t_2 + t_3 + t_4$$

Схема организации вычислительного процесса представлена на рисунке 11.

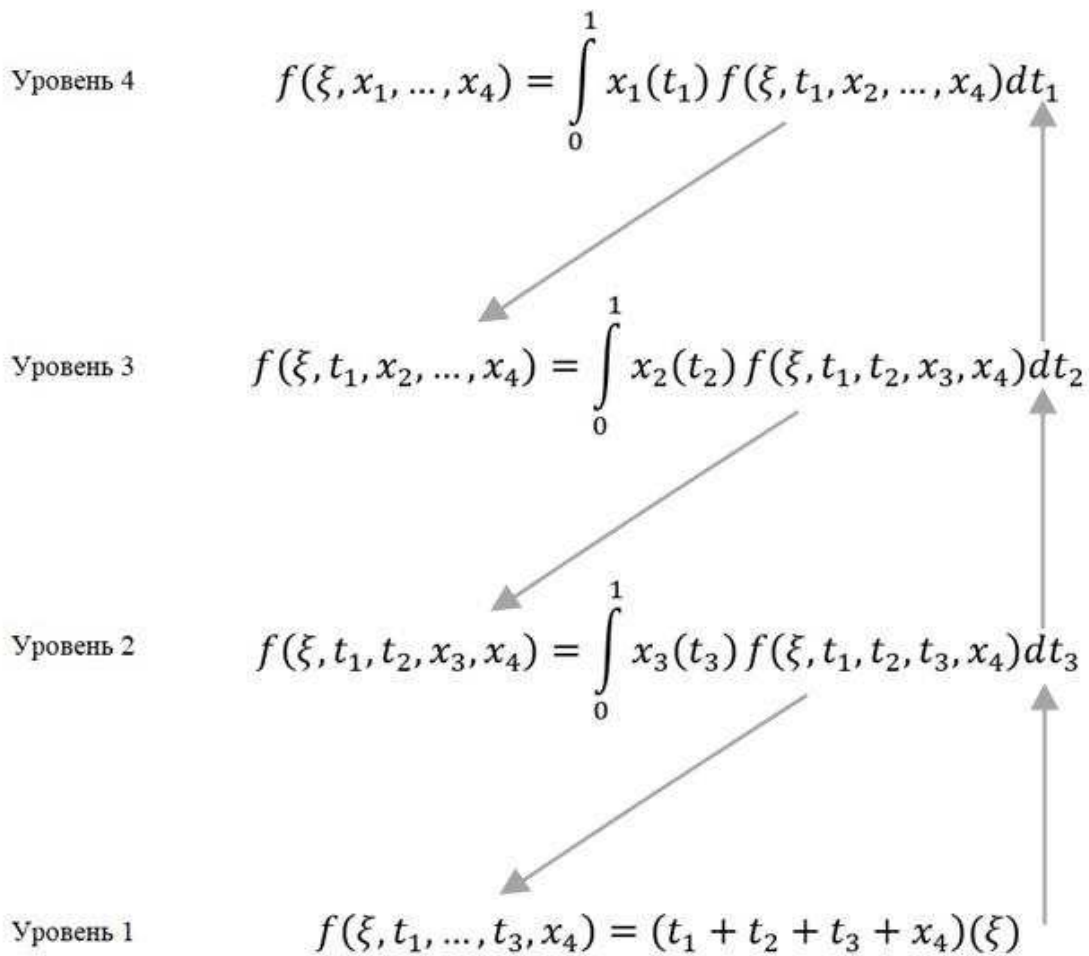


Рисунок 11 – Организация вычислительного процесса для тестового примера

Приведем оценки вычисления суммы четырех случайных величин в точке 2.0:  $S_4(2.0)$ . Точное значение функции  $S_4(2.0) = 2/3$ , при  $h = 0.1$  значение вероятностного расширения равно 0.6700, при  $h = 0.05$  – 0.66750. Таким образом, ошибка в первом случае составила  $\varepsilon_1 = 0.00333$ , во втором –

$\varepsilon_1 = 0.00083$ . Оценим порядок сходимости  $\alpha$ . Предположим, что выполняется зависимость  $\varepsilon(h) \approx Ch^\alpha$

Тогда

$$\ln(\varepsilon(h_i)) \approx \ln(C) + \alpha \ln(h_i), i = 1, 2.$$

Значение  $\alpha = 2.00434$ , это совпадает с теоретической оценкой порядка сходимости метода прямоугольников.

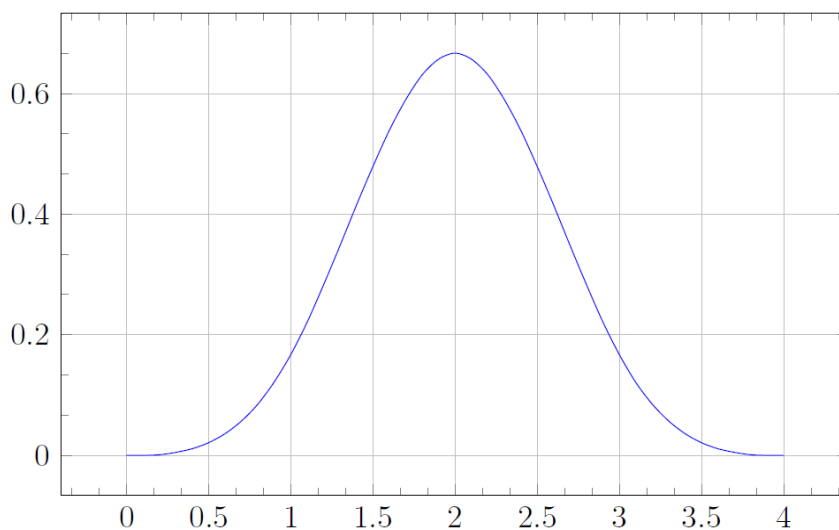


Рисунок 12 – Результат вычислений

### Пример 9. Расчет NPV и IRR.

Приведем пример формул, которые используются для оценки инвестиционных проектов предприятия, производящего товары.

Это прежде всего оценка денежных потоков  $CF_2$ , далее чистый дисконтированный доход  $NPV$  и внутренняя норма доходности  $IRR$ .

$$CF_Z = \sum_i G_i P_i (1 - AV_{Gi}) - \sum_j E_j (1 - AV_{Ej}) - \sum_k S_k W_k T_{Wk} - \\ - \sum_p H_p (A_p + T_{Hp}) - \sum_q T_q - \sum_n C_n R_n + Fin_z - Fout_z,$$

где  $G_i$  – количество продаж  $i$ -го товара;

$P_i$  – цена  $i$ -го товара;

$AV_{Gi}$  – налог на добавленную стоимость на  $i$ -ый товар;

$E_j$  – расходы  $j$ -го вида;

$AV_{Ej}$  – налог на добавленную стоимость на приобретаемые товары по  $j$ -му виду расходов;

$S_k$  – численность персонала  $k$ -ой категории;

$W_k$  – средняя заработная плата работников  $k$ -ой категории;

$T_{Wk}$  – коэффициент отчислений во внебюджетные фонды по  $k$ -ой категории работников;

$H_p$  – основные фонды  $p$ -го вида;

$A_p$  – норма амортизации по  $p$ -му виду основных фондов;

$T_{Hp}$  – ставка налога, базой для расчета которого выступает  $p$ -ый вид основных фондов (налог на имущество);

$T_q$  – сумма  $q$ -го налога;

$C_l$  – сумма  $l$ -го кредита;

$R_l$  – процент по  $l$ -ому кредиту;

$TPr_z$  – ставка налога на прибыль в  $z$ -ом году;

$Fin_z$  – прочие операционные, финансовые и инвестиционные поступления средств;



$Fout_z$  – прочие операционные, финансовые и инвестиционные выплаты средств.

$NPV$  – это сумма приведенных к текущему моменту времени чистых денежных потоков по инвестиционному проекту. Данный показатель определяется по следующей формуле:

$$NPV = \sum_{z=1}^T \frac{CF_z}{(1+d)^z}$$

где  $T$  – расчетный срок инвестиционного проекта в годах;

$d$  – ставка дисконтирования.

Инвестиционный проект признается эффективным в случае, если  $NPV > 0$ .

$IRR$  – расчетная ставка дисконтирования, при которой чистый дисконтированный доход ( $NPV$ ) равен нулю.  $IRR$  определяется из уравнения

$$\sum_{z=1}^T \frac{CF_z}{(1+IRR)^z} = 0.$$

Инвестиционный проект признается эффективным в случае если  $IRR > 0$ .

Как правило, для оценки инвестиционных проектов необходимо рассчитывать показатели на несколько лет вперед. В условиях высокой рыночной неопределенности такие показатели, например, как  $G_i$ ,  $P_i$ , имеют существенно стохастический характер. Однако при расчетах, как правило, используют только детерминированные оценки. Данные задачи заданы таблицей 2

Таблица 2 – Исходные данные задачи

	Год 0	Год 1	Год 2	Год 3
Цена упаковки		\$6,00	\$6,05	\$6,10
Количество проданных, шт.		802000	967000	1132000
Выручка		\$4 812 000	\$5 850 350	\$6 905 200
Себестоимость		\$2 646 600	\$3 217 693	\$3 797 860
Валовая прибыль		\$2 165 400	\$2 632 658	\$3 107 340
Операционные издержки		\$324 810	\$394 899	\$466 101
Чистый доход до налогов		\$1 840 590	\$2 237 759	\$2 641 239
Налоги		\$588 989	\$716 083	\$845 196
Стартовые инвестиции	-\$3 400 000			
Чистый доход	-\$3 400 000	\$1 251 601	\$1 521 676	\$1 796 043

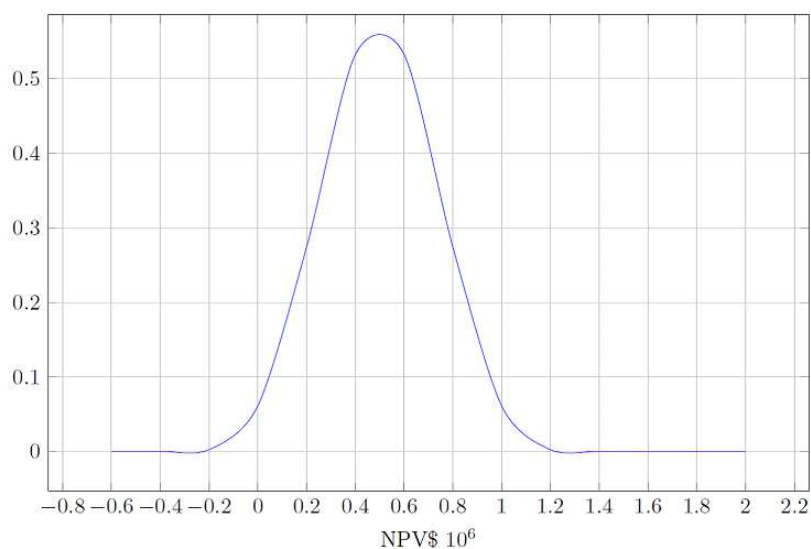


Рисунок 13 – функция плотности вероятности NPV

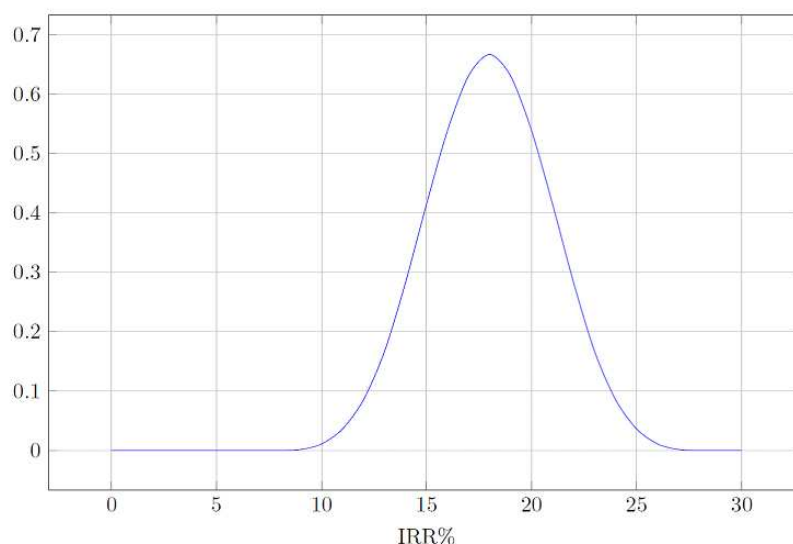


Рисунок 14 – Функция плотности вероятности IRR

На рис. 13 и рис. 14 приведены гистограммы *NPV* и *IRR* для рассмотренной задачи. Из анализа гистограмм *NPV* и *IRR* видно, что вероятны как крайне негативные исходы, так и возможна значительная прибыль в сравнении со стандартным анализом. Приведенный пример показывает, что применение рассмотренного подхода показывает, что рассмотренная выше организация вычислительного процесса с применением вероятностных расширений позволяет лицу, принимающему решения увидеть возможные варианты негативных исходов реализации проекта по сравнению со стандартным анализом, который дает только положительный ответ.

## ЗАКЛЮЧЕНИЕ

Разработанный программный модуль реализует новый подход к вычислению функций от случайных аргументов, представляющий собой технику быстрых вычислений, основанную на идее параллельных вычислений и применении численного вероятностного анализа. Для вычисления функции от случайных аргументов используется одно из основных понятий численного вероятностного анализа – вероятностное расширение. Для реализации техники быстрых вычислений применяется метод, основанный на параллельно – рекурсивных вычислениях.

Предлагается процесс вычисления вероятностных расширений представить в виде параллельного рекурсивного вычислительного процесса и представить его как технику быстрых вычислений для задач вычисления различных классов функций со случайными аргументами.

Построенная рекурсия обеспечивает численные рекурсивные операции, когда значение каждого очередного элемента структуры зависит от значений одного или нескольких предшествующих членов. Характерной особенностью применения построенного алгоритма распараллеливания является их независимость друг от друга, что означает, что эти задачи можно обрабатывать параллельно. В конечном итоге это создает тот ресурс параллелизма, который позволяет эффективно организовать вычислительный процесс для решения той или иной задачи. Данный модуль нацелен на решение широкого класса практических задач со случайными входными данными, например, где требуется вычисление функций от случайных аргументов. Например, задача оценки привлекательности инвестиционного проекта, для вычисления показателей  $NPV$  и  $IRR$ , а также задача оценки вероятности отказа оборудования ответственного назначения.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Багриновский К.А. Основы согласования плановых решений. М.: Наука. – 1977. – 244 с.
2. Велиходский А. С. Анализ подходов к обработке данных для оценки показателей инвестиционных проектов с учетом неопределенных факторов // Молодежь и наука: проспект Свободный: сб. материалов
3. Велиходский А. С., Минин А. С. Применение численного вероятностного анализа для оценки рисков инвестиционного проекта // Научные исследования и разработки молодых ученых: тр. VII МНПК. Новосибирск. – 2015. – С. 87-92.
4. Воробьев О. Ю. Современные теории неопределенности: эвентологический взгляд // Тр.VIII Международной конференции ФАМ. Красноярск: СФУ, 2009.
5. Добронев Б. С. Интервальная математика: Учеб. пособие // Б. С. Добронев. – Красноярск: Краснояр. гос. ун-т, 2004. – 216 с.
6. Добронев Б. С. Численный вероятностный анализ для исследования систем в условиях неопределенности / Б. С. Добронев, О. А. Попова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. - 2012. - № 4 - С. 39-46.
7. Добронев Б. С. Численный вероятностный анализ для оценки инвестиционных проектов // Б. С. Добронев, О. А. Попова, Е. В. Головчанская // XI Межд. конф. ФАМЭБ - Красноярск, 2012.
8. Добронев Б. С. Численный вероятностный анализ неопределенных данных: монография / Б. С. Добронев, О. А. Попова. - Красноярск: Сиб. фед. ун-т., – 2014. – 168 с.
9. Добронев Б. С., Попова О. А. Представление и обработка неопределенности на основе гистограммных функций распределения и P-Voxes // Информатизация и связь – 2014. – № 2. – С. 23-26.

10. Добронеев Б. С., Попова О. А. Гистограммные временные ряды // Тр. X Международной конференции ФАМЭТ-2011. Красноярск: КГТЭН, СФУ, 2011. С. 130–133.

11. Добронеев Б. С., Попова О. А. Элементы численного вероятностного анализа // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. 2012. – № 2 (42). – С. 19-23.

12. Добронеев 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.

их приложения // Б. С. Добронеев, О. А. Попова // Журнал Сиб. фед. ун-та, Математика и физика. – 2011 – 4(2) – С. 229-239.

13. Корчикина Д. И. Полиграммы для представления случайных данных // Научные итоги года: достижения, проекты, гипотезы. 2015. № 5. С. 148-152.

14. Крянев А. В., Лукин Г. В. Математические методы обработки неопределенных данных. – 2-изд., испр. – М.: ФИЗМАТЛИТ, 2006. – 216 с.

15. Крянев А.В. Лукин Г.В. Удумян Д.К. Метрический анализ и обработка данных. – Москва.: Издательство Физмалтит. – 2012. – 279 с.

16. Лукашев А. В. Метод Монте-Карло для финансовых аналитиков: краткий путеводитель. // Управление корпоративными финансами 2007. 18с.

17. Математическая энциклопедия: Электронный ресурс. URL: [http://dic.academic.ru/dic.nsf/enc\\_mathematics](http://dic.academic.ru/dic.nsf/enc_mathematics)

18. Определение среднего значения, вариации и формы распределения. Описательные статистики [Электронный ресурс]: // Режим доступа: <http://baguzin.ru/wp/?p=5381>.

19. A. G. Glen, L. M. Leemis, J. H. Drew, Computing the distribution of the product of two continuous random variables // Computational statistics and data analysis. – 2002. – 14 с.

20. Ahlberg J. H., Nilson E. N., Walsh J. L. (1967) The theory of splines and their applications. Academic Press, New York.

21. D. Dubois, H. Prade, Fuzzy real algebra: some results, *Fuzzy Sets and Systems* 2 (1979) 327–348.
22. Dempster A.P. Upper and lower probabilities induced by a multi-valued mapping // *Annals of Mathematical Statistics*. – 1967. – № 38. – C. 325-339.
23. Dobronets B. S., Krantsevich A. M., Krantsevich N. M. Software implementation of numerical operations on random variables // *Journal of Siberian Federal University. Mathematics & Physics*. – 2013. – № 6 (2). – C. 168–173.
24. Ferson S., W. Root, R. Kuhn, RAMAS Risk Calc: Risk assessment with uncertain numbers. *Applied Biomathematics*, Setauket, New York, 1999. Home page address: <http://www.ramas.com/riskcalc.htm>.
25. Jaroszewicz S., Korzen M. Arithmetic operations on independent random variables: a numerical approach // *Society for Industrial and Applied Mathematics*. – 2012. – 25c.
26. Kaufmann A., M. M. Gupta, *Fuzzy Mathematical Models in Engineering and Management Science*, Elsevier Science Publishers B.V., Amsterdam, 1988.
27. Liu B. *Theory and Practice of Uncertain Programming* (2nd Edition). Springer-Verlag, Berlin, 2009.
28. Moore R. E., *Interval arithmetic and automatic error analysis in digital computing*. Ph.D. Dissertation, Stanford University, 1962.
29. Neumaier A. Clouds, Fuzzy Sets and Probability Intervals // *Reliable Computing*. – 2004. – № 10. – C. 249-272.
30. R. C. Williamson. Probabilistic arithmetic // A thesis submitted for the degree of Doctor of Philosophy, Department of Electrical Engineering. – 1989. – 301c.
31. Swiler L. P., Giunta A. A. Aleatory and epistemic uncertainty quantification for engineering applications // *Sandia Technical Report*, SAND2007–2670.
32. W. Li, J. Mac Hyman. Computer arithmetic for probability distribution variables // *Reliability Engineering and System Safety*. – 2004. – 19c.

33. Ащепков Л. Т., Давыдов Д. В., Универсальные решения интервальных задач оптимизации и управления, Институт прикладной математики ДВО РАН, М., Наука, 2006.

34. Орлов А. И., Теория принятия решений: Учеб. Пособие. // М.: "Март", Экзамен – 2005. – 656 с.

35. Перепелица В. А., Тебуева Ф. Б., Дискретная оптимизация и моделирование в условиях неопределенности данных, М., Академия Естествознания – 2007.

36. Попова О. А. Гистограммы второго порядка для численного моделирования в задачах с информационной неопределенностью // Известия ЮФУ. Технические науки. 2014. № 6 (155). С. 6-14.

37. Попова О.А. Гистограммный информационно – аналитический подход к представлению и прогнозированию временных рядов // Информатизация и связь. – 2014. – № 2. – С. 43–47.

38. Турчак Л. И., Плотников П. В. Основы численных методов: учебное пособие. – 2-е изд. перераб. и доп. – М.: ФИЗМАТЛИТ, 2003. – 304 с. Тарасенко Ф. П. Непараметрическая статистика. Томск. ТГУ, – 1976. – 294с.

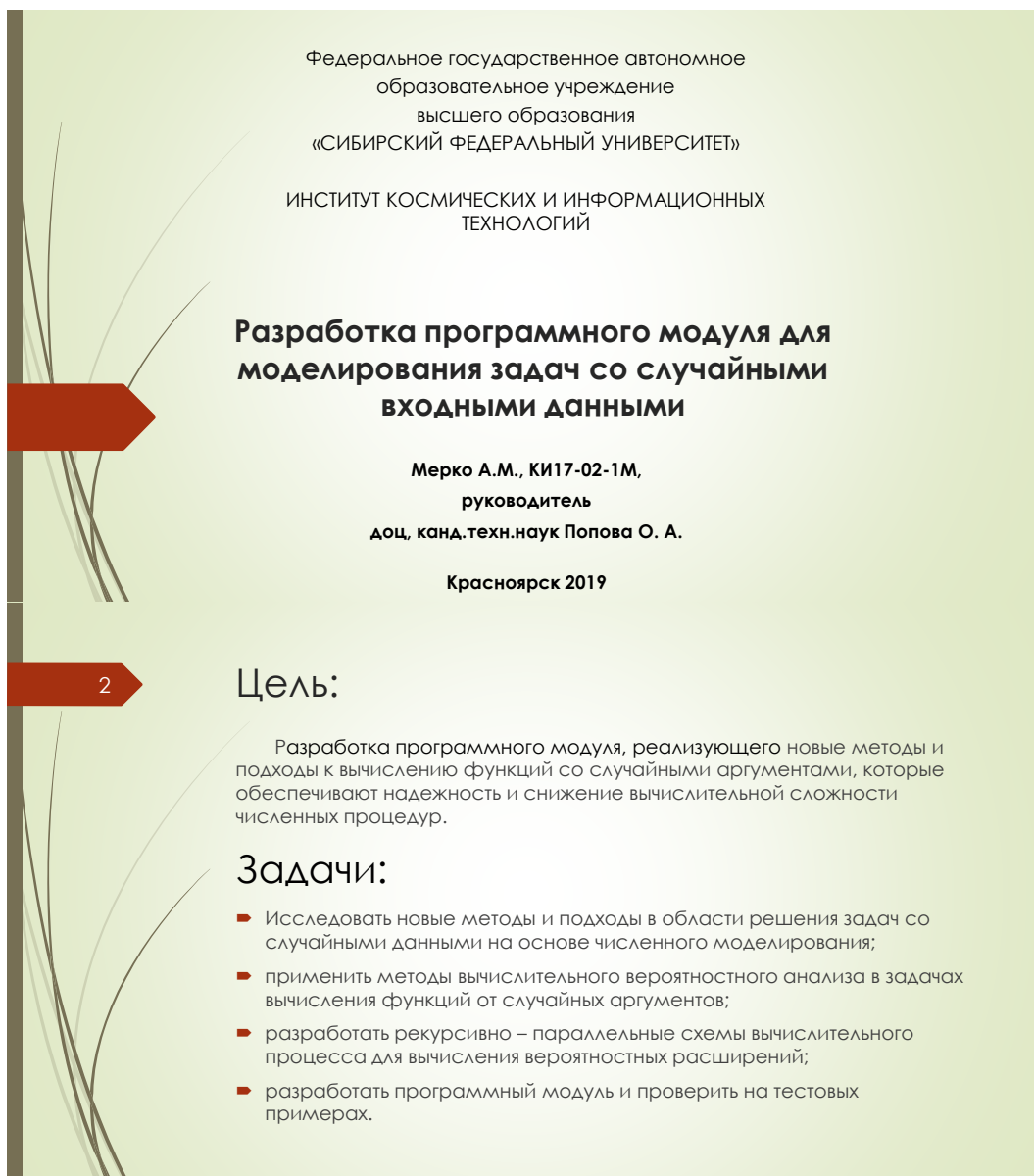
39. Ужга-Ребров О. И. Управление неопределенностями. Часть 1. // Современные концепции и приложения теории вероятностей. – Rezekne: RAIZdevnieciba, 2004. – 292 с.

40. Шарый С. П. Конечномерный интервальный анализ. // Институт вычислительных технологий СО РАН. Г. Новосибирск. 2017. 617с.



# ПРИЛОЖЕНИЕ А

## Слайды презентации



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

ИНСТИТУТ КОСМИЧЕСКИХ И ИНФОРМАЦИОННЫХ  
ТЕХНОЛОГИЙ

**Разработка программного модуля для  
моделирования задач со случайными  
ВХОДНЫМИ ДАННЫМИ**

Мерко А.М., КИ17-02-1М,  
руководитель  
доц, канд.техн.наук Попова О. А.

Красноярск 2019

2

**Цель:**

Разработка программного модуля, реализующего новые методы и подходы к вычислению функций со случайными аргументами, которые обеспечивают надежность и снижение вычислительной сложности численных процедур.

**Задачи:**

- Исследовать новые методы и подходы в области решения задач со случайными данными на основе численного моделирования;
- применить методы вычислительного вероятностного анализа в задачах вычисления функций от случайных аргументов;
- разработать рекурсивно – параллельные схемы вычислительного процесса для вычисления вероятностных расширений;
- разработать программный модуль и проверить на тестовых примерах.

3

## Задачи со случайными входными данными

Общая поставка задач:  $y = f(x_1, \dots, x_n)$

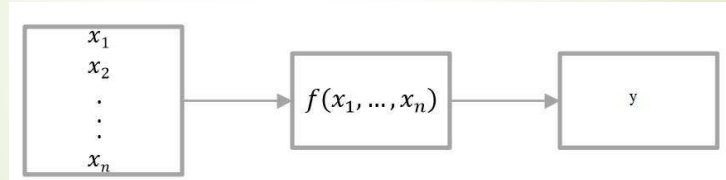


Рисунок 1 - Задачи со случайными входными данными

4

## Пример задачи со случайными входными данными

- NPV – это сумма приведенных к текущему моменту времени чистых денежных потоков по инвестиционному проекту. Данный показатель определяется по следующей формуле:

$$NPV = \sum_{z=1}^T \frac{CF_z}{(1+d)^z}$$

- IRR – расчетная ставка дисконтирования, при которой чистый дисконтированный доход (NPV) равен нулю. IRR определяется из уравнения

$$\sum_{z=1}^T \frac{CF_z}{(1+IRR)^z} = 0.$$

5

## Метод Монте-Карло.

Универсальным подходом к решению таких задач является метод Монте-Карло, который давно и успешно используется для решения разнообразных задач в условиях неопределенности.

При всех его положительных качествах этот метод обладает рядом недостатков. Один из самых существенных это низкая скорость сходимости и большой объем вычислений, что представляет дополнительные сложности при работе с данными большого объема.

## Численный вероятностный анализ

Численный вероятностный анализ - это раздел вычислительной математики, предназначенный для решения различных задач со случайными входными данными.

Одним из основных понятий численного вероятностного анализа является понятие вероятностного расширения.

## Вероятностные расширения

- Рассмотрим задачу определения закона распределения функции нескольких случайных аргументов. Пусть имеется система непрерывных случайных величин  $(x_1, \dots, x_n)$  с плотностью распределения  $p(x_1, \dots, x_n)$ . Случайная величина  $z$  связана с  $(x_1, \dots, x_n)$  функциональной зависимостью:
 
$$z = f(x_1, \dots, x_n)$$

Тогда плотность вероятности случайной величины  $z$ , будем называть вероятностным расширением функции  $f$ .

## Основная теорема вероятностных расширений

■

**Теорема.** Пусть  $f(z, x_1, x_2, \dots, x_n)$  вероятностное расширение  $f(x_1, x_2, \dots, x_n)$  и для всех вещественных  $t$  функция  $f(\cdot, t, x_2, \dots, x_n)$  есть вероятностное расширение функции  $f(t, x_2, \dots, x_n)$ . Тогда

$$f(z, x_1, x_2, \dots, x_n) = \int x_1(t) f(z, t, x_2, \dots, x_n) dt$$

## Дерево параллельной рекурсии

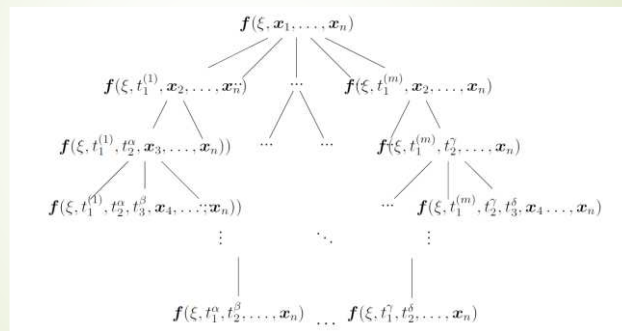


Рисунок 2 - Дерево параллельной рекурсии

## Общая схема модуля

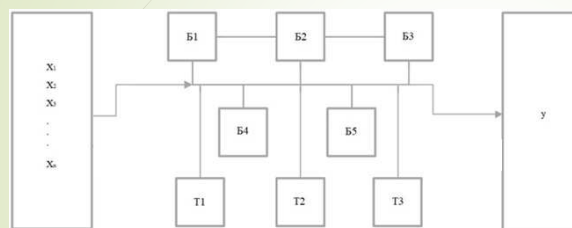


Рисунок 3 - Общая схема модуля

Основные блоки:

- B1 – Построение кусочно-полиномиальной модели.
- B2 – Построение вероятностных расширений.
- B3 – Организация и вычисления рекурсий.
- B4 – Анализ результатов.
- B5 – Блок визуализации модуля.

Дополнительные блоки:

- T1 – Распределение Ирвина-Холла.
- T2 – Численное интегрирование. Квадратурные формулы.
- T3 – Вероятностная арифметика.

## Пример 1

- Рассмотрим пример для вычисления суммы равномерно распределенных случайных величин на интервале  $[0, 1]$ .

$$y = f(x_1, x_2, x_3, x_4).$$

тогда

$$f(x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4,$$

где

$x_i$  - равномерно распределенная случайная величина на интервале  $[0, 1]$

Формула Ирвин-Холла для  $n = 4$ :

$$f_x(x) = \begin{cases} \frac{1}{6}x^3 & 0 \leq x \leq 1 \\ \frac{1}{6}(-3x^3 + 12x^2 - 12x + 4) & 1 \leq x \leq 2 \\ \frac{1}{6}(3x^3 - 24x^2 + 60x - 44) & 2 \leq x \leq 3 \\ \frac{1}{6}(x-4)^3 & 3 \leq x \leq 4 \end{cases}$$

12

## Организация вычислительного процесса

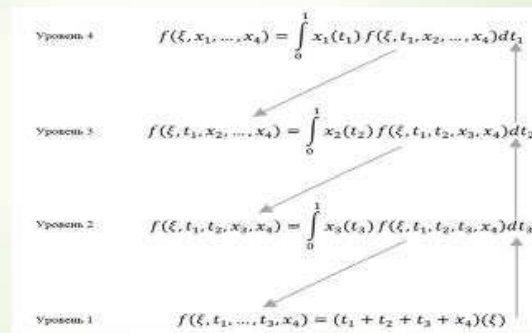
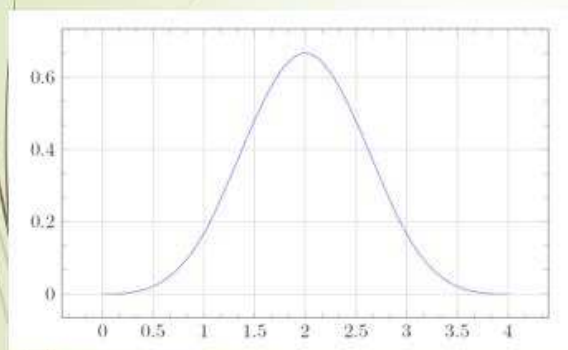


Рисунок 4 - Организация вычислительного процесса для тестового примера

13

## Результаты тестирования



- При  $h = 0.1$  значение вероятностного расширения равно 0.6700. Ошибка  $\varepsilon_1 = 0.00333$
- При  $h = 0.05 - 0.66750$ . Ошибка  $\varepsilon_2 = 0.00083$ .

Рисунок 5 - Результат вычислений тестового примера

14

## Пример 2. Расчет NPV и IRR

Таблица 1 – Исходные данные примера

	Год 0	Год 1	Год 2	Год 3
Цена упаковки		\$6,00	\$6,05	\$6,10
Количество проданных, шт.		802000	967000	1132000
Выручка		\$4 812 000	\$5 850 350	\$6 905 200
Себестоимость		\$2 646 600	\$3 217 693	\$3 797 860
Валовая прибыль		\$2 165 400	\$2 632 658	\$3 107 340
Операционные издержки		\$324 810	\$394 899	\$466 101
Чистый доход до налогов		\$1 840 590	\$2 237 759	\$2 641 239
Налоги		\$588 989	\$716 083	\$845 196
Стартовые инвестиции	-\$3 400 000			
Чистый доход	-\$3 400 000	\$1 251 601	\$1 521 676	\$1 796 043

15

## Результаты решения примера. Расчет NPV

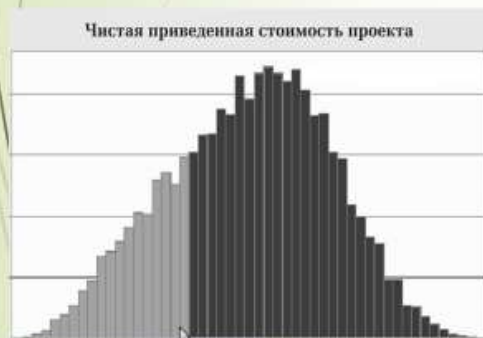


Рисунок 6 – Результат анализа Монте-карло из статьи Лукашов А.В.

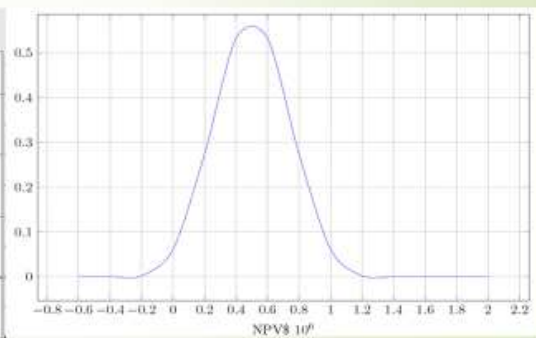


Рисунок 7 - Функция плотности вероятности NPV

16

## Заключение

Разработанный программный модуль реализует новый подход к вычислению функций от случайных аргументов, представляющий собой технику быстрых вычислений, основанную на идее параллельных вычислений и применении численного вероятностного анализа.

Для вычисления функции от случайных аргументов используется одно из основных понятий численного вероятностного анализа – вероятностное расширение.

17

По результатам исследования темы магистерской диссертации была опубликована статья:  
«АГРЕГАЦИЯ И ТЕХНИКА БЫСТРЫХ ВЫЧИСЛЕНИЙ ДЛЯ ДАННЫХ БОЛЬШИХ ОБЪЕМОВ»

Данная статья была представлена на X Всероссийской научно-технической конференции с международным участием.

г. Железногорск, 8 декабря 2018 г

