

На правах рукописи

Соловьев Александр Анатольевич

**АНАЛИЗ СЛУЧАЙНЫХ ДИСКРЕТНО-ТОЧЕЧНЫХ ПОЛЕЙ С
ИСПОЛЬЗОВАНИЕМ АНАЛИТИЧЕСКИХ ПРЕОБРАЗОВАНИЙ
НА ЭВМ**

05.13.18. “Математическое моделирование, численные методы и
комплексы программ”

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Новосибирск – 2013

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте автоматики и электрометрии Сибирского отделения Российской академии наук (ИАиЭ СО РАН)

Научный руководитель: **Резник Александр Львович**
доктор технических наук, заведующий лабораторией

Официальные оппоненты: **Нежевенко Евгений Семенович**
доктор технических наук,
Федеральное государственное бюджетное учреждение науки Институт автоматики и электрометрии Сибирского отделения Российской академии наук,
ведущий научный сотрудник

Курносков Михаил Георгиевич
кандидат технических наук,
Федеральное государственное образовательное бюджетное учреждение высшего профессионального образования «Сибирский государственный университет телекоммуникаций и информатики»,
доцент

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук

Защита состоится 1 ноября 2013 г. в 10-00 на заседании диссертационного совета Д 003.005.01 при Федеральном государственном бюджетном учреждении науки Институте автоматики и электрометрии СО РАН по адресу: 630090, Новосибирск, проспект Академика Коптюга, 1.

С диссертацией можно ознакомиться в библиотеке ИАиЭ СО РАН.

Автореферат разослан «27» сентября 2013 г.

Ученый секретарь диссертационного совета

д.ф.-м.н.

Насыров К.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

В XXI веке компьютеры стали привычной частью жизни. За относительно короткий отрезок времени мощность обычного настольного компьютера выросла многократно – он намного превосходит лучшие ЭВМ тридцатилетней давности по своим характеристикам, а современные суперкомпьютеры могут совершать миллиарды операций в секунду.

Именно это сделало возможным получение тех результатов, о которых пойдет речь в представленной диссертационной работе, которая описывает исследования в области обработки случайных полей и изображений. Изучение этих задач было начато в Институте автоматизации и электрометрии СО РАН более 30 лет назад, но основные результаты, представленные в работе, получены в течение последних нескольких лет.

Главной особенностью представляемой диссертационной работы является то, что в ней для решения сложных проблем фундаментального характера, возникающих при изучении случайных дискретных полей и цифровых изображений, разработаны уникальные методы, позволяющие проводить на ЭВМ не *численные* расчеты, а *аналитические* преобразования. При этом компьютер используется не просто как мощный вычислитель, а как интеллектуальное средство для проведения эквивалентных математических выкладок, выполнить которые вручную исследователь не в состоянии из-за их гигантского объема.

Программные системы, использующие аналитическое манипулирование на ЭВМ, были успешно применены для научных исследований многими российскими и советскими учеными, такими как В.М.Глушков, В.А.Брумберг, Л.В. Канторович, И.В. Поттосин, Н.Н. Яненко, Д.В. Ширков и другими. Разработанные алгоритмы были предназначены для решения проблемных и трудоемких задач математики, теоретической физики, механики и других дисциплин.

Второй отличительной особенностью описываемых в диссертации методов является то, что с их помощью удалось в чистом виде реализовать идею, высказанную в свое время Дж. фон Нейманом: исследователь сталкивается с задачей, которую не в состоянии решить, привлекает ЭВМ для проведения трудоемких расчетов, которые способны натолкнуть его на «правильный» ответ, и в случае удачи (т.е. зная «подсказанное» компьютером решение) проводит строгое и конструктивное доказательство. В диссертации удалось, используя компьютерные расчеты, сначала установить, а затем математически строго доказать ряд точных аналитических формул, описывающих

вероятность безошибочного считывания случайных дискретно-точечных изображений, что сделать без серьезной программной поддержки было бы невозможно либо весьма затруднительно. Дополнительная сложность в данном случае заключалась в том, что компьютерные вычисления были связаны не с численными расчетами, а со специально разработанными программными системами для проведения аналитических выкладок.

Третьей отличительной особенностью диссертационной работы является то, что для строгого математического доказательства полученных в ней вероятностных соотношений было введено понятие и найдены явные выражения для обобщенных чисел Каталана, являющихся естественным трехмерным расширением классической числовой последовательности Каталана, которая известна еще со времен Леонарда Эйлера и возникает во многих приложениях теории вероятностей и математической статистики. Не вызывает сомнений, что введенное понятие обобщенной последовательности Каталана является серьезным вкладом в развитие перечислительной комбинаторики и окажется полезным инструментом при решении многих теоретических и прикладных вероятностно-комбинаторных задач.

Цель и задачи работы

Целью диссертационной работы было нахождение с помощью аналитических преобразований на ЭВМ формул, описывающих вероятность безошибочного считывания случайных дискретных полей и изображений интеграторами с несколькими пороговыми уровнями, и проведение строгого математического доказательства полученных соотношений.

Для достижения поставленной цели потребовалось решить следующие задачи:

1. Разработать специализированные алгоритмы и на их основе создать программные системы, ориентированные на проведение аналитических преобразований, требующихся при обработке и анализе случайных дискретно-точечных изображений.
2. Ввести новое понятие обобщенных трехмерных чисел Каталана и найти их явный аналитический вид.
3. С помощью построенных программно-аналитических систем и с применением обобщенных чисел Каталана последовательно выполнить трехэтапную доказательную схему: на первом этапе рассчитать максимально широкий набор частных решений задачи многопорогового считывания; на втором этапе на основе системного анализа полученных частных решений сформулировать непротиворечивую гипотезу-

предположение об аналитическом виде общего решения; на третьем этапе математически строго доказать выдвинутую гипотезу.

Научная новизна

1. Предложена комплексная методика решения трудоемких математических задач с использованием аналитических преобразований на ЭВМ. Разработанные на ее основе три программно-алгоритмические системы успешно применены для изучения процесса считывания дискретно - точечных полей.
2. Найдено новое трехмерное обобщение числовой последовательности Каталана, успешно примененное для расчета надежности различных методов считывания случайных дискретных изображений.
3. Создан новый программный алгоритм параллельных вычислений для аналитического расчета многомерных интегральных выражений, позволивший на порядок увеличить скорость нахождения частных решений при расчете надежности многопорогового считывания случайных изображений.
4. Найдены замкнутые аналитические соотношения для вероятности безошибочного считывания дискретных изображений, осуществляемого двухпороговыми интеграторами.

Практическая значимость

К числу практически важных результатов работы относится создание программной системы «АПП-МНИТ», предназначенной для вычисления вероятности безошибочного считывания случайных дискретно-точечных полей и изображений с использованием параллельных вычислений. Данная система основана на методе аналитического вычисления многомерных интегралов (зарегистрирована в Роспатенте под номером 2013612764).

Другим важным практическим результатом является нахождение нового трехмерного обобщения классической последовательности Каталана, которое является вкладом в развитие перечислительной комбинаторики и может быть полезно при решении многих прикладных задач вероятностного характера.

В диссертации найдены новые математические формулы, которые могут быть непосредственно использованы в ряде других научных областей (теория массового обслуживания, биометрия, математическая эпидемиология и другие).

Еще одним существенным практически значимым результатом является развитый в диссертации подход к использованию

предварительных компьютерных расчетов для отыскания частных решений задачи (в частности, создание методов и алгоритмов для проведения аналитических преобразований на ЭВМ, способ их практической реализации и использование «компьютерных» подсказок для нахождения общего доказательства). Этот подход может быть успешно перенесен на решение других задач, причем не только вероятностно-комбинаторных.

Значительная часть вошедших в диссертацию исследований была проведена в рамках Государственной программы IV.29.1. «Теоретические основы и методы информационных и вычислительных технологий проектирования и принятия решений». Исследования были поддержаны грантами РФФИ (№ 13-01-00361 «Интеллектуальный анализ случайных дискретных полей на основе компьютерных аналитических вычислений», № 10-01-00458 «Цифровые методы оптимального анализа случайных полей дискретной структуры»); программой Президиума РАН № 15/2012-2014 г.г. (проект «Интеллектуальная программная поддержка в задачах оптимальной цифровой обработки случайных полей и изображений дискретной структуры»); программой совместных фундаментальных исследований СО РАН с НАН Беларуси 2012-2014 г.г. (проект «Методы, алгоритмы и программно-аппаратные системы реконструкции, улучшения качества и повышения разрешающей способности сигналов и изображений видимого и ИК диапазонов»).

Основные положения, выносимые на защиту

1. Разработка трех эффективных алгоритмов для определения вероятности безошибочного считывания дискретных изображений, осуществляемого интеграторами, имеющими ограниченное количество пороговых уровней.
2. Создание и использование программной системы «АПП-МНИТ» (с применением параллельных вычислений) для нахождения частных решений при фиксированных значениях параметров сканирования.
3. Введение в научную практику, нахождение явного аналитического вида и применение в задачах анализа случайных дискретно-точечных полей трехмерной обобщенной последовательности Каталана.
4. Проведение строгого математического доказательства рассчитанных с помощью ЭВМ замкнутых аналитических соотношений, описывающих надежность двухпорогового считывания случайных дискретных изображений.

Методы исследований

В диссертационной работе используются методы теории вероятностей, математической статистики, математического моделирования, комбинаторики, прикладного программирования и ряда других дисциплин.

Апробация работы

Основные научные и практические результаты работы докладывались и обсуждались на следующих международных конференциях и семинарах: «Распознавание образов и анализ изображений: новые информационные технологии – РОАИ-10-2010», Санкт-Петербург; «Automation, Control, and Information Technology (ACIT 2010)», Новосибирск; «Современные проблемы математики, информатики и биоинформатики», посвященная 100-летию со дня рождения А.А.Ляпунова, Новосибирск, 2011 г.; «7 VIII German-Russian Workshop “Pattern Recognition and Image Understanding”», Нижний Новгород, 2011 г.; «Neural Networks and Artificial Intelligence (ICNNAI - 2012)», Минск, 2012 г.; «Euro-American Conference for Academic Disciplines», Париж, 2013 г.

Публикации

По материалам диссертации опубликовано 12 научных работ [1-12], из них 6 статей в рецензируемых журналах, входящих в перечень ВАК.

Структура и объем диссертации

Диссертационная работа состоит из введения, трех глав и заключения. Работа изложена на 136 страницах машинописного текста, содержит 8 рисунков, 2 таблицы, список литературы (95 наименований) и приложение.

Личный вклад автора

Создание систем аналитических преобразований на современных вычислительных системах (в том числе алгоритмов для параллельных вычислений на специализированных кластерах) выполнено автором лично. Участие автора в разработке математических методов, использованных в диссертации для доказательства новых соотношений, описывающих надежность многопорогового считывания случайных дискретных изображений, полностью отражает приведенный в работе список литературы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** описана область, в которой проводились исследования, сформулированы актуальность темы, научная новизна, защищаемые положения работы, личный вклад автора.

Глава 1 диссертационной работы является обзором литературы по теме работы и посвящена целям и основным направлениям развития систем аналитических преобразований на ЭВМ. В начале главы уделено внимание истории развития таких систем - начиная с самых первых специализированных программ для больших ЭВМ до современных прикладных пакетов общего назначения для персональных компьютеров. Кратко описаны такие проекты, как «Schoonchip» (система расчета задач физики быстрых частиц Мартина Велтмана), «Reduce» (программная среда для аналитических вычислений, разработанная Энтони Херном), «АНАЛИТИК» (система для аналитических вычислений в области математического анализа, созданная под руководством В.М.Глушкова), и ряд других.

Далее в главе описано современное состояние систем аналитического манипулирования. Отмечено, что процесс программирования на современных языках и системах существенно изменился – многие трудности прошлого были преодолены (малое количество памяти, отсутствие «длинной арифметики» и т.д.). Другим серьезным продвижением сегодня является возможность использования вычислительных кластеров («суперкомпьютеров»), которые имеют огромную вычислительную мощность и возможность использования параллельных вычислений.

В заключительной части главы описаны созданные в лаборатории №12 Института автоматизации и электротехники СО РАН уникальные программные системы аналитических преобразований для изучения различных научных проблем (в частности, для нахождения асимптотически оптимальных декоррелирующих преобразований сигналов различной степени гладкости и для построения эффективных алгоритмов локализации импульсно-точечных объектов со случайным временем генерации бесконечно коротких дельта-импульсов).

Глава 2 диссертации содержит постановку задачи, а также описание разработанных для ее решения специализированных алгоритмов аналитического манипулирования на ЭВМ.

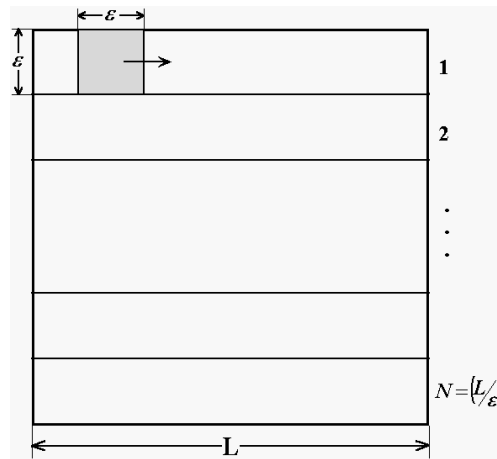


Рис.1. Схема телевизионного считывания случайного “пуассоновского” поля размером $L \times L$.

Задача, которая стимулировала проведение данной работы, формулируется следующим образом: необходимо определить вероятность безошибочного считывания случайного дискретного поля (изображения), когда регистрация координат точечных объектов ведется телевизионным методом с помощью интегратора, обладающего несколькими пороговыми уровнями (рис 1.). При этом считывающая апертура, имеющая размер $\varepsilon \times \varepsilon$, осуществляет последовательное сканирование по горизонтали всех полос изображения, имеющих ширину ε . Апертура непрерывно перемещается внутри каждой из полос слева направо, двигаясь по оси абсцисс.

При попадании точечного объекта в пределы считывающей апертуры суммарный сигнал интегратора увеличивается на единицу. При выбывании какого-либо объекта из окна апертуры суммарный уровень сигнала интегратора уменьшается на единицу. При телевизионном считывании дискретно-точечных изображений интеграторами, имеющими ограниченное число пороговых уровней, важнейшей характеристикой является вероятность безошибочного считывания изображения, т.е. вероятность того, что за весь период сканирования в окне считывающей апертуры ни разу не будет находиться более k объектов (здесь k – число пороговых уровней интегратора).

На начальном этапе решения двумерная задача определения вероятности безошибочного считывания сводится к следующей эквивалентной одномерной задаче: "Пусть n точек x_1, x_2, \dots, x_n случайно брошены на интервал $(0, L)$, т.е. имеется n независимых испытаний случайной величины, равномерно распределенной на интервале $(0, L)$. Необходимо найти вероятность $P_{n,k}(\varepsilon, L)$ того, что внутри интервала $(0, L)$

не содержится ни одного подынтервала Ω_ε длиной ε , содержащего более k точек."

Аналитическое решение этой задачи на данный момент известно лишь для случая, когда $k=1$:

$$P_{n,1}(\varepsilon, L) = (L - (n-1)\varepsilon)^n / L^n, \quad (0 \leq \varepsilon \leq L / (n-1)).$$

Существующие методы решения поставленной задачи не позволяют вычислить вероятности $P_{n,k}(\varepsilon, L)$ при произвольных значениях n и k , поэтому в диссертации были разработаны несколько алгоритмов, в основе которых лежат аналитические преобразования на ЭВМ.

Первый алгоритм поиска частных решений, который был реализован, основан на методе аналитического вычисления на ЭВМ многомерных интегралов по выпуклым многогранникам в n -мерном пространстве.

Искомая вероятность представляется в следующем виде

$$P_{n,k}(\varepsilon) = n! \int \dots \int_{D_{n,k}(\varepsilon)} dx_1 \dots dx_n,$$

где область интегрирования $D_{n,k}(\varepsilon) \subset R^n$ описывается системой линейных неравенств

$$\begin{cases} 0 < x_1 < x_2 < \dots < x_{n-1} < x_n < 1, \\ x_{k+1} - x_1 > \varepsilon, \\ x_{k+2} - x_2 > \varepsilon, \\ \vdots \\ x_n - x_{n-k} > \varepsilon. \end{cases}$$

Далее интеграл по области $D_{n,k}(\varepsilon)$ записывается в особой эквивалентной форме, где интегрирование ведется по всему пространству, а индикаторная функция области интегрирования трансформируется в произведение единичных функций Хевисайда. Затем n -мерный интеграл с помощью процедуры «последовательного расщепления» сводится к набору повторных интегралов с уже расставленными пределами интегрирования.

На заключительном этапе сначала проводится последовательное интегрирование каждого из N повторных интегралов, на которые распался исходный интеграл, а затем объединение полученных результатов с учетом границ изменения свободного параметра ε . Этот алгоритм позволяет конструктивно вычислять формулы $P_{n,k}(\varepsilon)$, что полностью решает задачу нахождения вероятности $P_{n,k}(\varepsilon)$ для конкретных значений n и k .

Вышеописанный алгоритм, основанный на методе аналитического вычисления многомерных интегралов, лег в основу разработки программной системы «АПП-МНИТ», использующей параллельные вычисления (версия для кластера ИВЦ НГУ). С помощью этой системы удалось значительно продвинуться в нахождении частных решений и получить формулы $P_{n,k}(\varepsilon)$ до $n=14$ включительно.

Параллельный алгоритм в процессе расчета *одновременно* проверяет альтернативные пределы в системе линейных неравенств (рис. 2), посылая данные на разные процессоры (за счет этого получается существенное ускорение процесса вычисления).

$$\left\{ \begin{array}{l} x_2 > \alpha, \\ x_2 > \beta, \\ \vdots \\ \vdots \\ \vdots \\ x_2 < x_3. \end{array} \right. \Rightarrow \left\{ \begin{array}{l} x_2 > \alpha, \\ \vdots \\ x_2 < x_3 \\ \alpha > \beta. \end{array} \right. \Rightarrow \text{CPU - 1}$$

$$\Rightarrow \left\{ \begin{array}{l} x_2 > \beta, \\ \vdots \\ x_2 < x_3 \\ \beta > \alpha. \end{array} \right. \Rightarrow \text{CPU - 2}$$

Рис. 2. Параллельная проверка системы неравенств

Второй программно реализованный алгоритм основан на рекурсивных комбинаторно-аналитических вычислениях. Использовалась следующая комбинаторная модель. Интервал $(0,L)$ интерпретировался как совокупность r равных дискретов. Случайное бросание n точек на интервал $(0,L)$ интерпретировалось как случайное бросание n неразличимых шаров по r ящикам. Множество из l смежных ячеек служило аналогом подынтервала длиной ε . Исход бросания, когда ни один из таких l - подынтервалов, содержащихся внутри исходного r - интервала $(0,L)$, не имел более 2 точек, считался “успешным”, а отношение общего числа “успешных” бросаний $Q(r,n,l)$ к общему числу исходов опыта $Q(r,n)$ принималось в качестве аналога вероятности $P_{n,2}(\varepsilon)$. На этом пути удалось существенно продвинуться в вычислении вероятностей $P_{n,2}(\varepsilon)$, т.е. для случая $k=2$.

Третьим реализованным алгоритмом аналитического манипулирования является метод дифференцирования многомерных интегралов по параметру. Алгоритм основан на том, что любой полином $P(\varepsilon)$ степени n может быть однозначно восстановлен по своему значению и значению всех своих производных по параметру ε в некоторой определенной точке.

Также в главе 2 приведен перспективный метод многомерного интегрирования. Его основная идея состоит в том, что с помощью прямого интегрирования непосредственно в общей аналитической форме последовательно отыскиваются формулы для вероятностей $P_{n,n-1}(\varepsilon)$, $P_{n,n-2}(\varepsilon)$, $P_{n,n-3}(\varepsilon)$ и т.д. В отличие от всех алгоритмов, приводившихся ранее, в этой схеме свободными остаются уже две переменные – не только непрерывный параметр ε , но также целочисленный параметр n .

В результате работы всех вышеперечисленных алгоритмов был получен широкий набор частных решений $P_{n,k}(\varepsilon)$ при различных значениях параметров n и k .

В частности, для четных n были установлены следующие соотношения:

$$P_{4,2}(\varepsilon) = 2(1-\varepsilon)^4, \quad (1/2) < \varepsilon < 1)$$

$$P_{6,2}(\varepsilon) = 5(1-2\varepsilon)^6, \quad (1/3) < \varepsilon < (1/2)$$

$$P_{8,2}(\varepsilon) = 14(1-3\varepsilon)^8, \quad (1/4) < \varepsilon < (1/3)$$

$$P_{10,2}(\varepsilon) = 42(1-4\varepsilon)^{10}, \quad (1/5) < \varepsilon < (1/4)$$

$$P_{12,2}(\varepsilon) = 132(1-5\varepsilon)^{12}, \quad (1/6) < \varepsilon < (1/5)$$

Анализ этих «компьютерных» формул позволил усмотреть общую закономерность и высказать гипотезу, заключающуюся в том, что при двух пороговых уровнях (т.е. когда $k = 2$) для четных n на интервале $1/(n/2) < \varepsilon < 1/((n/2) - 1)$ справедлива формула

$$P_{n,2}(\varepsilon) = (2/n)C_n^{(n/2)-1} (1 - ((n/2) - 1)\varepsilon)^n.$$

Данная формула была «подсказана» компьютером и опубликована еще в 1981 году, а вот ее строгое доказательство удалось получить сравнительно недавно – в 2011 году [5].

Кроме того, в результате проведенных расчетов было установлено, что для случая телевизионного считывания, проводимого с двумя пороговыми уровнями, при произвольных n на участке $0 < \varepsilon < 1/(\text{entire}((n+1)/2))$ вероятность безошибочного считывания $P_{n,2}(\varepsilon)$ представляется в виде

$$\begin{aligned}
P_{n,2}(\varepsilon) = & C_n^0 + C_n^2(-n+2)\varepsilon^2 + C_n^3(4n-10)\varepsilon^3 + C_n^4(3n^2-37n+86)\varepsilon^4 + \\
& + C_n^5(-40n^2+394n-922)\varepsilon^5 + C_n^6(-15n^3+625n^2-5171n-12086)\varepsilon^6 + \\
& C_n^7(420n^3-10724n^2+79996n-187002)\varepsilon^7 + C_n^8(105n^4-10570n^3+ \\
& + 205499n^2-1426841n+3336406)\varepsilon^8 + C_n^9(5040n^4-155708n^3+ \\
& + 2267664n^2-17317506n+52315558)\varepsilon^9 + C_n^{10}(-945n^5+189000n^4- \\
& -15794625n^3+389687181n^2-3798029823n+1299896664)\varepsilon^{10} + o(\varepsilon^{10}).
\end{aligned}$$

Данное соотношение также является «компьютерной» гипотезой. Хотя ей не противоречит ни один из полученных компьютерных результатов, тем не менее, соотношение требует своего строгого математического доказательства. Компьютерные расчеты лишь помогают в поиске общей замкнутой аналитической зависимости, описывающей поведение функций $P_{n,2}(\varepsilon)$ в диапазоне изменения параметра ε , включающем точку $\varepsilon = 0$, но точный характер этой зависимости пока еще остается невыясненным.

Глава 3 посвящена математически строгому доказательству полученных с помощью ЭВМ новых замкнутых аналитических соотношений, описывающих вероятность $P_{n,2}(\varepsilon)$ безошибочного считывания случайного одномерного n -точечного изображения.

Первая часть главы посвящена доказательству полученной с помощью ЭВМ (см. главу 2) формулы-гипотезы для четных n на интервале $1/(n/2) < \varepsilon < 1/((n/2)-1)$

$$P_{n,2}(\varepsilon) = (2/n)C_n^{(n/2)-1} (1 - ((n/2)-1)\varepsilon)^n.$$

Формулировка задачи, для которой приведено строгое решение, такова:

«Имеется интервал $(0,1)$, на который случайно брошены $n = 2m$ точек. Требуется найти вероятность P события, заключающегося в том, что внутри интервала $(0,1)$ не существует ни одного подынтервала Ω_ε длины ε ($\frac{1}{m} < \varepsilon < \frac{1}{m-1}$), содержащего более 2 точек».

Схема размещения точек, удовлетворяющая формулировке задачи, показана на рис. 3.

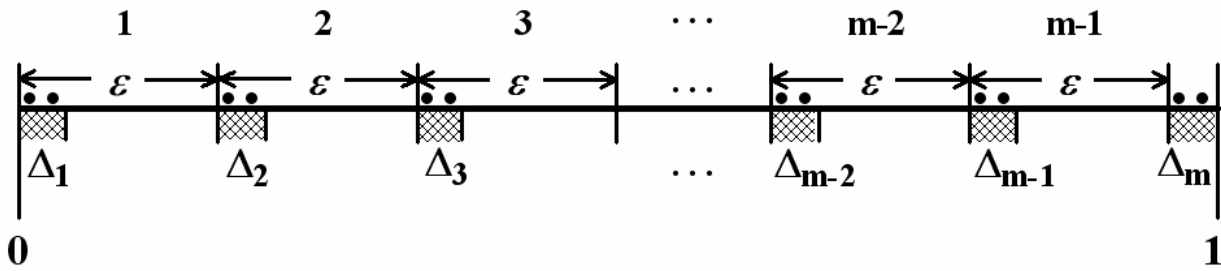


Рис.3. Схема размещения четного ($n = 2m$) числа случайных точек на интервале $(0,1)$, обеспечивающая возможность их безошибочного «одномерного» считывания интегратором, имеющим апертуру ε и два пороговых уровня.

Краткая схема доказательства такова:

1. На первом этапе определяется вероятность P_0 того, что все $n=2m$ точек при их случайном бросании на интервал $(0,1)$ попадут в заштрихованную зону, причем в каждый заштрихованный интервал попадет ровно 2 точки

$$P_0 = P_1 \times P_2 = \frac{(2m)!}{2^m} (1 - (m-1)\varepsilon)^{2m}.$$

2. Далее учитывается, что в рассматриваемом случае двухпорогового считывания ($k=2$) все ближайшие члены с четными номерами и все ближайшие члены с нечетными номерами в ранжированной последовательности $x_1 < x_2 < \dots < x_{n-1} < x_n$ должны быть удалены друг от друга на расстояние, большее ε

$$P = P_0 \times \frac{N_m}{A_m} = \frac{(2m)!}{2^m} (1 - (m-1)\varepsilon)^{2m} \times \frac{N_m}{(2m)! / 2^m} = N_m \times (1 - (m-1)\varepsilon)^{2m}.$$

3. Третьим этапом является решение задачи подсчета количества различных «правильных» LR-слов (Left-Right-слов), имеющих длину $2m$

$$N_{M+1} = C_{2M+2}^{M+1} - N_{M+1}^* = C_{2M+2}^{M+1} - C_{2M+2}^M = \frac{1}{M+2} C_{2M+2}^{M+1}.$$

На финальном этапе, делая подстановку и заменяя $2m$ на n , окончательно получаем требуемую формулу для вероятности P события, заключающегося в том, что внутри интервала $(0,1)$ не существует ни одного подынтервала ω_ε длиной ε ($\frac{1}{m} < \varepsilon < \frac{1}{m-1}$), содержащего более 2 точек:

$$\begin{aligned}
P &= N_m \times (1 - (m-1)\varepsilon)^{2m} = \frac{1}{m+1} C_{2m}^m (1 - (m-1)\varepsilon)^{2m} = \\
&= \frac{1}{m} C_{2m}^{m-1} (1 - (m-1)\varepsilon)^{2m} = (2/n) C_n^{(n/2)-1} (1 - ((n/2) - 1)\varepsilon)^n.
\end{aligned}$$

Анализ полученных результатов показывает, что в данном случае удалось полностью реализовать уже упоминавшуюся идею Дж. фон Неймана об использовании «компьютерной подсказки»: в нашем случае вышеупомянутая формула была «подсказана» компьютером, после чего удалось провести ее строгое доказательство.

Далее в главе 3 изложено расширение классической последовательности Каталана, которое потребовалось для доказательства формул, описывающих вероятность безошибочного двухпорогового считывания $P_{n,k}(\varepsilon)$ для нечетных значений $n = 2m + 1$ на участке $1/(m+1) < \varepsilon < 1/m$. Приведены некоторые задачи, приводящие к числам Каталана: задача расстановки скобок, задача разбиения выпуклого многоугольника, количество путей в квадрате, не пересекающих главную диагональ, задача о рукопожатиях и ряд других.

Трехмерное обобщение чисел Каталана и приводящая к ним комбинаторно-«лингвистическая» задача формулируется следующим образом: «Из N_a символов “a”, N_b символов “b” и N_c символов “c” составляются различные слова длиной $(N_a + N_b + N_c)$. Нужно определить, сколько среди всех этих слов таких, что при просмотре слова слева направо количество встреченных символов “b” никогда не превышает количества встреченных символов “a”, а при просмотре слова справа налево количество встреченных символов “c” никогда не превышает количества встреченных символов “a” (естественным образом предполагается, что $N_{b,c} \leq N_a$)».

Показано, что число таких слов

$$W(N_a, N_b, N_c) = \frac{(N_a + N_b + N_c)!}{N_a! N_b! N_c!} \left[1 - \frac{N_b + N_c}{N_a + 1} + \frac{N_b N_c}{(N_a + 1)(N_a + 2)} \right].$$

Числа $W(N_a, N_b, N_c)$ получили название «обобщенных трехмерных чисел Каталана». Представленная формула действительно обобщает известную по многим приложениям одномерную последовательность Каталана, которая получается из $W(N_a, N_b, N_c)$ при $N_c = 0$ и $N_b = N_a$.

Далее в главе 3 описано применение обобщенных чисел Каталана для доказательства замкнутого аналитического соотношения, описывающего надежность $P_{n,k}(\varepsilon)$ двухпорогового (т.е. для $k=2$) считывания случайных

дискретных изображений при нечетных значениях $n = 2m + 1$ на участке $1/(m+1) < \varepsilon < 1/m$.

Решаемая задача формулируется следующим образом:

“Пусть $n = 2m + 1$ точек случайно брошены на интервал $(0,1)$. Требуется указать вероятность $P_{2m+1,2}(\varepsilon)$ события, состоящего в том, что внутри интервала $(0,1)$ не существует ни одного подынтервала Ω_ε длиной ε , содержащего более 2 точек, если $1/(m+1) < \varepsilon < 1/m$.”

Представлено подробное последовательное решение этой задачи. Сначала кратко (без доказательства) приводятся основные утверждения, а потом по каждому из них даются необходимые пояснения и доказательные выкладки. Приводимые утверждения неравнозначны по степени сложности своего доказательства, а одно из утверждений является проблемной задачей, которая и потребовала для своего решения введения абсолютно нового понятия обобщенных чисел Каталана, нахождения их явного вида и построения на этой основе специальных вероятностно-геометрических схем. Но каждое из приведенных утверждений необходимо для нахождения замкнутой аналитической формулы, являющейся решением поставленной выше задачи.

Перемножая вероятности P_i , соответствующие этим утверждениям, и осуществляя суммирование

$$\sum_{p=0}^m \sum_{q=0}^{m-p} (P_1 \times P_2 \times P_3 \times P_4 \times P_5 \times P_6 \times P_7) ,$$

после всех преобразований получаем формулу для вероятности $P_{2m+1,2}(\varepsilon)$ на участке $1/(m+1) < \varepsilon < 1/m$:

$$\begin{aligned} P_{2m+1,2}(\varepsilon) = & C_{2m+1}^{m+1} (1 - m\varepsilon)^{m+1} (1 - (m-1)\varepsilon)^m - \\ & - 2C_{2m+1}^{m+2} (1 - m\varepsilon)^{m+2} (1 - (m-1)\varepsilon)^{m-1} + \\ & + C_{2m+1}^{m+3} (1 - m\varepsilon)^{m+3} (1 - (m-1)\varepsilon)^{m-2} \end{aligned}$$

В **заключении** диссертации отмечено, что многие задачи, связанные со случайным разбиением интервала, просты в постановке, но весьма трудны в решении. Представленная работа еще раз подтверждает, что зачастую нахождение даже частных решений совершенно простых по постановке задач выливается в серьезную научную проблему. Успехи в построении программных систем аналитического манипулирования и приобретенный в процессе их создания опыт позволяют надеяться, что в ближайшей перспективе удастся расширить список удачных применений компьютера для решения проблемных научных задач.

Основные результаты

1. Разработаны специализированные алгоритмы и на их основе созданы программные системы (в том числе система расчета многомерных интегралов в n -мерном пространстве на основе параллельных вычислений «АПП-МНИТ»), ориентированные на проведение аналитических преобразований, требующихся при обработке и анализе случайных дискретно-точечных изображений.
2. Введено новое понятие обобщенных трехмерных чисел Каталана и найден их явный аналитический вид.
3. С помощью построенных программно-аналитических систем и с применением обобщенных чисел Каталана найдены и доказаны формулы, описывающие вероятность безошибочного двухпорогового считывания дискретных полей и изображений.
4. Продемонстрирован пример эффективного применения компьютера для интеллектуальной программной поддержки при решении проблемных вероятностных задач, когда компьютер играет роль не только мощного вычислителя, но и интеллектуального помощника, оснащенного инструментом для проведения разветвленных и трудоемких аналитических выкладок.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

- [1] **Резник А.Л., Ефимов В.М., Соловьев А.А.** Оценивание надежности считывания случайных дискретных изображений с применением средств компьютерной аналитики // Вестник НГУ. Серия: Физика, 2010, т. 5, вып. 2. С. 104–110.
- [2] **Alexander L. Reznik, Vitaly M. Efimov, Alexander A. Soloviev.** Reliability of random discrete image reading estimated with computation analysis // Proceedings of the IASTED International Conferences on Automation, Control, and Information Technology. Novosibirsk, Russia, 2010, June 15-18, pp. 86-89.
- [3] **A.L. Reznik, V.M. Efimov, A.V. Torgov, A.A. Soloviev.** Computer analytical calculations for the random discrete structures analysis // 10th Int. Conf. Pattern Recognition and Image Analysis: New Information Technologies PRIA-10-2010, St.-Petersburg, December 5-12, 2010. Conference Proceedings, vol. 1, pp.251-254.
- [4] **Reznik A.L., Efimov V.M., Torgov A.V., and Solov'ev A.A.** Computer Analytics in Problems with a Random Partition of the Interval // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications, 2011, vol. 21, № 2. P. 202–205.

- [5] **А.Л.Резник, В.М.Ефимов, А.А.Соловьев.** Компьютерно-аналитический расчет вероятностных характеристик процесса считывания случайных точечных изображений // *Автометрия*, 2011, т.47, №1, с.10-16.
- [6] **Резник А.Л., Ефимов В.М., Соловьев А.А., Торгов А.В.** Обобщенные числа Каталана в задачах обработки случайных дискретных изображений // *Автометрия*, 2011, т. 47, № 6. С. 11–16.
- [7] **Reznik A., Efimov V., Soloviev A., Torgov A.** On reliability of random discrete images sensing // *Proc. of the VIII German-Russian Workshop “Pattern Recognition and Image Understanding”* (Nizhni Novgorod, Russia, November 21–26, 2011). P. 244–245.
- [8] **Соловьев А.А., Резник А.Л., Торгов А.В.** Специализированные программные системы для проведения аналитических выкладок как инструмент решения вероятностных задач со случайным разбиением интервала // *Международная конференция «Современные проблемы математики, информатики и биоинформатики», посвященная 100-летию со дня рождения А.А. Ляпунова* (г. Новосибирск, 2011). С. 63–64.
- [9] **A. L. Reznik, V. M. Efimov, A. V. Torgov, and A. A. Solov'ev.** Analytical Computer Calculations in Problems with Random Division of an Interval // *Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*, 2012, Vol.22, No.2, pp.354-359.
- [10] **Резник А.Л., Ефимов В.М., Соловьев А.А., Торгов А.В.** О безошибочном считывании случайных дискретно-точечных полей // *Автометрия*, 2012 т. 48, № 5. С. 93–103.
- [11] **A.L.Reznik, V.M. Efimov, A.V.Torgov, A.A.Soloviev.** Random Discrete Images Readout and Generalized Catalan Numbers // *Neural Networks and Artificial Intelligence (ICNNAI - 2012): Proceedings of th 7th International Conference* (Minsk, Belarus 10-12 October 2012) - Minsk: BSUIR, 2012. p.160-163.
- [12] **A.L.Reznik, V.M. Efimov, A.V.Torgov, A.A.Soloviev.** Reliability of Random Discrete Images Reading Estimation Based on Computer Analytical Calculations // *Neural Networks and Artificial Intelligence (ICNNAI - 2012): Proceedings of 7th International Conference* (Minsk, Belarus 10-12 October 2012) - Minsk: BSUIR, 2012. p.163-167.