Федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(национальный исследовательский университет)

На правах рукописи

Мхитарян Георгий Араикович

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И АЛГОРИТМЫ АДАПТИВНОГО ТЕСТИРОВАНИЯ В ПРОГРАММНОМ КОМПЛЕКСЕ МАТЕМАТИЧЕСКОЙ ПОДДЕРЖКИ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

05.13.18 – Математическое моделирование, численные методы и комплексы программ

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук,
доцент А.В. Наумов

Москва, 2021 год

ОГЛАВЛЕНИЕ

Введение

Теоретические исследования в области математических моделей, описывающих компьютерное тестирование, играют огромную роль в современных системах дистанционного компьютерного обучения. Данные модели позволяют оценить не только характеристики тестируемых, но и качество материала, предоставляемого для проектирования тестов.

Исследования в области оценки результатов тестов развиваются с начала XX века, опираясь на работы Ч. Спирмена [130], изучавшего когнитивные тесты. Развитием идей Ч. Спирмена стали работы U. Yule [153], T.L. Kelley [95], F. Kuder [97], M. Richardson[124 – 126]. В работах L. Guttman [92, 93], M. Novick [111, 118] была сформулирована классическая теория тестирования, позволяющая оценивать такие параметры тестирования как сложность заданий или способности тестируемых.

С развитием вычислительных систем и появлением возможности активного их применения для формирования моделей тестирования и получения оценок параметров теста основной концепцией проектирования актуальных методов тестирования стала Item Response Theory (IRT, современная теория тестирования). Идеи IRT в своих работах с середины XX века развивали F.M. Lord [110, 111], G. Rasch [123], P. Lazarsfeld [29, 108], B.D. Wright [109, 151, 152], D. Andrich [69, 70, 71]. Фундаментальной идеей современной теории тестирования является применение логистических функций для описания зависимости вероятности правильного ответа на тестовое задания от трудности тестового задания и способности испытуемого. В свою очередь IRT является основой парадигмы компьютерного адаптивного тестирования (CAT, computerized adaptive testing), которая предполагает адаптацию тестов под способности тестируемого. Основные принципы проектирования компьютерных адаптивных систем в конце XX века описаны в работе D. J. Weiss [96, 143 – 145], G. G Kingsbury [96, 143], а также в недавних работах W. J. Van der Linden [133 – 137], H. Wainer [139, 140].

Современные проблемы теории тестирования рассматриваются в работах W. J. Van der Linden [133 – 137], J. Piton-Gonçalves [120, 121], S. M. Aluísio [120, 121], H.Y. Chong [82], M. Linacre [109], J.-P. Fox [85], M. D. Zeigenfuse [112], W. H. Batchelder [73, 112], А.И. Кибзуна [12 – 20, 35, 36, 116], А.В. Наумова [2, 12, 16 – 20, 38, 43 – 48, 64, 116, 117], В.С. Аванесова [1], Л.С. Куравского [11, 23 – 27, 30, 98 - 106], П.Н. Думина [8 – 11, 24, 26, 27, 30, 99, 104, 105] и др.

В настоящее время существуют различные подходы к построению и оценке характеристик тестов и испытуемых. В работах А.И. Кибзуна, А.В. Наумова, С.И. Панарина [15] используются методы стохастического линейного и смешанного программирования. В работах Л.С. Куравского, П.Н. Думина, Г.А. Юрьева, рассмотрены модели, основанные на теории марковских процессов и стохастических дифференциальных уравнений.

Среди моделей, учитывающих временные показатели, выделяются многоуровневые модели G. Rasch [123] и W. J. Van der Linden [133 – 137], которые позволяют получить оценки времени, затрачиваемого пользователем на чтение (G. Rasch) и решение тестового задания (W. J Van der Linden), благодаря предположению о случайности времени и о законе распределения соответствующей случайной величины: гамма или логнормальном. Проблемой использования данных моделей в прикладных задачах проектирования тестов, использующих временные показатели как критерия или ограничений, в первом случае является недостаточное концептуальное соответствие, во втором случае неудобство исследования логнормальных случайных величин в случае рассмотрения множества заданий, т.к. сумма логнормальных случайных величин не имеет известного закона распределения.

В случае с применением данных моделей для адаптации системы компьютерного тестирования возникают проблемы вычислительного характера, что в итоге говорит о необходимости разработки:

- простых, но гибких в использовании моделей, описывающих или использующих время ответа на задания;
- эффективных численных методов и их реализаций в рамках систем компьютерного тестирования.

Таким образом, **актуальность темы диссертации** обусловлена необходимостью создания эффективных подходов к проектированию тестов с помощью адаптивных компьютерных процедур, учитывающих время выполнения тестовых заданий и изменения в состоянии испытуемых, т.е. адаптации контента систем компьютерного обучения в зависимости от изменяющихся характеристик пользователей системы.

Цель работы: разработка математических моделей и методов генерирования тестовых заданий, учитывающих время, затраченное на решение заданий пользователями в системах компьютерного обучения.

В процессе исследования решены **задачи**:

- создания модели времени ответа пользователя на задание, основанной на применении гамма-распределения;
- создания математической модели адаптивного тестирования, позволяющей генерировать индивидуальные тесты с ограничением на общее время выполнения;
- создания вероятностной модели генерации тестов для универсального испытуемого или группы испытуемых;
- создания модели генерации тестов с возможной приоритизацией временного параметра либо параметра сложности теста;
- разработки численных методов для оптимального решения задач, основанных на вышеперечисленных моделях;
- создания комплекса программ, реализующих данные численные методы;
- интеграции комплекса программ в архитектуру используемой системы компьютерного тестирования в продуктивной среде.

Объектом исследования является математическое и программное обеспечение систем дистанционного обучения.

Предметом исследования являются математические модели, используемые для адаптации систем дистанционного обучения под изменяющийся контингент пользователей, и программные средства, реализующие поддержку адаптивных свойств систем с применением данных моделей.

Область исследования формулируется согласно пунктам паспорта специальности 05.13.18:

- разработка новых математических методов моделирования объектов и явлений;
- реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента.
- комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента.

Методологические основы и методы исследования: для решения поставленных задач использовались методы математического моделирования, теории оптимизации, стохастического программирования, статистического анализа и численные методы.

На защиту выносятся следующие научные результаты:

- вероятностные модели времени ответа пользователя на задание системы, основанные на дискретном и гамма-распределении;
- математические модели генерации тестов с учетом вероятностных критериев, основанных на использовании случайного времени как параметра модели;
- численные методы определения оптимальных наборов тестовых заданий;
- комплекс программ, реализующий численные методы;
- архитектура модуля генерирования тестов в активно функционирующей среде компьютерного тестирования.

Научная новизна. Получены новые результаты:

- вероятностные модели описания времени ответа на задания системы компьютерного тестирования (гамма, дискретная);
- математические модели для создания теста с ограничением на время выполнения;
- численные методы решения задач подбора оптимального множества тестов;
- пакет прикладных программ для придания адаптивных свойств системам дистанционного обучения и компьютерного тестирования

Практическая значимость диссертационной работы заключается в возможности создания принципиально новых инструментов для проектирования тестов в системах компьютерного тестирования, адаптации наборов тестовых заданий под изменяющийся контингент пользователей с учётом реальных показателей их работы в системе, а также создания программного комплекса в виде отдельного универсального интегрируемого внешнего модуля.

Достоверность результатов исследования подтверждается:

- оценкой адекватности полученных результатов наблюдениям с помощью статистических критериев согласия,
- практической реализацией и успешным применением системы математической поддержки функционирования, созданной на основе разработанного подхода,
- вычислительными экспериментами, подтвердившими эффективность и преимущества созданных численных методов и применения вероятностных моделей.

Структура и объём диссертации. Диссертация содержит введение, 3 главы, заключение и список используемой литературы. Работа состоит из 95 страниц, включая 13 рисунков, 20 таблиц и список литературы, содержащий 153 наименования.

Во введении обоснована актуальность проблемы, сформулированы цель и задачи исследования, определены объект, предмет и методы исследования, дана общая характеристика работы. Далее проанализированы основные современные методы и модели тестирования и времени ответа пользователя, указаны их преимущества и недостатки, сделан вывод о необходимости развития и применения ряда новых технологий.

В первой главе «Математические модели времени ответа пользователя на задание» приводится описание моделей времени ответа пользователя на тестовые задания в системах дистанционного обучения. В разделе 1.1 рассматриваются классические подходы тестирования с использованием компьютерных тестов, а также тестирование с использованием сложных технических систем. В разделе 1.2 описаны основные математические модели времени ответа пользователя на задания, среди которых основной является логнормальная модель В. ван дер Линдена. Далее приводится описание построения иерархической модели времени и получения оценок для параметров логнормального распределения в случае, когда тестовые задания выполняются группой тестируемых. Так как в случае логнормального распределения времени ответа на задания для группы студентов сложно получить близкую к достоверной оценку суммарного времени выполнения всего теста, предлагаются дискретная модель времени ответа и модель с гамма-распределенным временем. В разделе 1.3 рассматривается использование дискретного распределения, которое предполагает построение модели случайного времени ответа некого универсального, т.е. не учитывающего индивидуальные особенности, конкретного пользователя на задание. В такой модели индивидуальные особенности всех пользователей тестируемой группы учитываются интегрально при формировании модели времени ответа некоторого универсального пользователя, что позволяет увеличить объем используемой статистической информации для оценки параметров модели и упрощает формализацию данной модели и её использование в прикладных задачах. Для проверки адекватности модели при участии реальных пользователей системы дистанционного обучения был проведён эксперимент, в рамках которого проверялась гипотеза о том, что для каждого задания, время, которое было затрачено пользователями на выполнение, является логнормальной случайной величиной. Приведены примеры дискретных распределений для некоторых заданий, а также гистограмма и график плотности соответствующего логнормального распределения. Далее в разделе 1.4 рассматривается модель с гамма-распределением времени, основными преимуществами которой являются плотность распределения близкая по своей структуре к плотности логнормального распределения, и то, что выполняется свойство суммируемости, т.е. сумма таких случайных величин с одинаковым параметром λ является гамма-распределенной. Эти особенности модели позволили во второй главе диссертации рассмотреть постановку задачи формирования ограниченного по времени теста с квантильным критерием качества и предложить эффективный алгоритм ее решения.

Во второй главе «Методы и алгоритмы формирования адаптивных тестов с учетом времени ответов пользователей» приведены постановки оптимизационных задач для адаптации функционирования систем дистанционного обучения и компьютерного тестирования на основе анализа времени ответа пользователя на задания системы, а также методы и алгоритмы их решения. В разделе 2.1 рассмотрена задача определения набора заданий приблизительно равных по суммарной сложности с ограничением на время выполнения. Вводятся основные обозначения, которые позволяют формализовать модель построения тестов при условии, что время на выполнение теста ограничено. Описываются требования к множеству заданий и множеству пользователей. Рассматриваются подробно критерии формирования тестов. В разделе 2.2 рассмотрена задача формирования ограниченного по времени теста с логнормальным распределением для группы пользователей. Приводится постановка задачи с учётом вероятностного ограничения с логнормальными случайными величинами. Сформулированную задачу возможно решить, используя доверительный метод, но из-за особенностей свойств логнормального распределения возникают сложности в применении известных методов, основанных, например, на использовании аппроксимации оптимального доверительного множества, поэтому простейшим методом поиска решения является полный перебор, т.к. количество всех возможных наборов тестовых заданий конечно. Для ускорения поиска решения предложен алгоритм поиска наборов тестовых заданий, удовлетворяющих ограничениям задачи. Алгоритм ускоренного поиска основан на идее метода ветвей и границ и позволяет значительно уменьшить количество рассматриваемых наборов. Далее рассмотрен результат вычислительного эксперимента, приведены данные, характеризующие работу алгоритма. В разделе 2.3 рассмотрена задача формирования ограниченного по времени теста для универсального пользователя с дискретными случайными величинами в вероятностном ограничении. Данная задача сводится к детерминированной задаче целочисленного программирования, а также приводится теорема об эквивалентности задач по набору оптимальных решений и подробное доказательство данной теоремы. Далее рассмотрены результаты вычислительного эксперимента, в рамках которого показано соответствие решений вероятностной задачи и задачи целочисленного программирования. В разделе 2.4 рассмотрена задача формирования теста с квантильным критерием. В отличие от модели формирования теста с логнормальным временем распределения общее время на выполнение теста неизвестно, поэтому для того, чтобы за некоторое оптимальное время все тестируемые могли выполнить выданный вариант теста с заданной вероятностью α, рассматривается функция квантили. Критериальная функция принимает вид суммы двух нормированных безразмерных величин, с помощью которых можно приоритизировать подбор теста по сложности либо времени выполнения. Использование логнормальной модели времени для решения сформулированной задачи не позволяет получить точное решение, так как для задач квантильной оптимизации известны лишь методы поиска гарантирующих решений, основанные на доверительном методе, или плохо сходящиеся стохастические квазиградиентные процедуры, поэтому для решения сформулированной задачи рассматривается дискретную модель распределения времени. Предложен метод дискретизации логнормального распределения, и в случае дискретной модели распределения времени ответа на задание исходная задача с квантильным критерием может быть сведена к детерминированной задаче частично целочисленного программирования. Сформулирована и доказана теорема об эквивалентности задач. Далее рассмотрены результаты вычислительного эксперимента для задачи подбора теста в случае одного пользователя. В разделе 2.5 рассмотрена задача формирования ограниченного по времени теста с гамма-распределением времени ответа пользователя на задания. Предложен алгоритм подбора параметров гамма-распределения для поиска суммируемых распределений случайного времени. Также предложен вариант алгоритма 2.1 для случая гамма-распределения в вероятностном ограничении. Далее рассмотрены результаты вычислительного эксперимента для задачи подбора теста в случае одного пользователя в случае критериальной функции, учитывающей приоритет между сложностью теста и временем выполнения.

В третьей главе «Программный комплекс адаптации системы дистанционного обучения» рассмотрен программный комплекс, включающий в себя модели и алгоритмы, описанные в главах 1 и 2. В разделе 3.1 рассматривается общая архитектура программного комплекса, как веб-приложения с двумя ключевыми модулями: моделирования времени и адаптации. Также рассмотрены основные принципы взаимодействия программного комплекса с внешними системами дистанционного обучения и компьютерного тестирования. Описана архитектура взаимодействия между СДО и программным комплексом и представлены два варианта реализации интеграции программного комплекса в функционирующую среду для тестирования. В разделе 3.2 рассмотрен модуль моделирования времени. Описаны основные возможности модуля, входные и выходные данные. Также рассмотрена архитектура модуля и описаны основные сущности и их атрибуты. Приведены примеры контрактов для клиент-серверного общения. Описана логика работы сервиса, отвечающего за работу с данными модуля, в случае обработки данных для одного пользователя или группы пользователей. В разделе 3.3 рассмотрен модуль адаптации контента системы компьютерного тестирования. Описаны основные возможности модуля, входные и выходные данные, а также рассмотрена архитектура модуля и описаны основные сущности и их атрибуты. Приведены примеры контрактов для клиент-серверного общения. Описана логика работы сервиса, отвечающего за работу с данными модуля, в случае обработки данных для одного пользователя или группы пользователей. Описаны пользовательские сценарии взаимодействия с модулем, интегрированным в систему дистанционного обучения для роли «студент» и «преподаватель». Описаны типы данных, используемые для взаимодействия внешних систем с модулем. В разделе 3.4 рассмотрен пример графического интерфейса веб-приложения, входящего в программный комплекс адаптации системы дистанционного обучения. Описаны основные технологи, использованные для реализации интерфейса. Рассмотрены возможности интеграции графического интерфейса в системы дистанционного обучения.

В заключении подведены итоги работы, отражены основные результаты диссертации.

Апробация. Теоретические основы и практические результаты работы представлены на следующих научных мероприятиях: VII Международная аэрокосмическая декада (Алушта, 2014); Региональный этап Всероссийского конкурса прорывных проектов в области IT-технологий «IT-Прорыв» в рамках Международной недели авиакосмических технологий «Aerospace Science Week» (Москва, 2014); Международная научная конференция «Образование, наука и экономика в вузах и школах. Интеграция в международное образовательное пространство». (Горис, Армения, 2015); XVI Международный форум «Формирование современного информационного общества – проблемы, перспективы, инновационные подходы», (Санкт-Петербург, 2015); X Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование» (Москва, 2015); 14-я Международная конференция "Авиация и космонавтика - 2015" (Москва, 2015); XLII Международная молодёжная научная конференция Гагаринские чтения – 2016 (Москва, 2016); VIII традиционная всероссийская молодежная летняя школа «Управление, информация и оптимизация» (Санкт-Петербург, 2016); XI Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование» (Москва, 2016); 4-я Всероссийская научно-техническая конференция Суперкомпьютерные Технологии (СКТ-2016) (Геленджик, 2016); VIII Международная научно-практическая конференция «Информационные технологии в образовании «ИТО-Саратов-2016» (Саратов, 2016); XII Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование» (Москва, 2017); XLIV Международная молодёжная научная конференция Гагаринские чтения – 2018 (Москва, 2018); XII Международная конференция по прикладной математике и механике в аэрокосмической отрасли (NPNJ'2018) (Алушта, 2018); V Международная конференция «Функциональные пространства. Дифференциальные операторы. Проблемы математического образования» (Москва, 2018); XIII Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование» (Москва, 2018); XXIV Международная научная конференция "Системный анализ, управление и навигация" (Евпатория, 2019).

Работа поддержана грантами РФФИ (проекты №18-07-00617, №17-07-00203, №15-07-02914, №14-07-00006).

Публикации. Основные результаты по теме диссертации изложены в 22 работах, из которых 8 опубликованы в журналах, рекомендованных ВАК [2, 16, 31, 37, 46, 47, 116, 117], в том числе 3 опубликованы в журналах, цитируемых международной базой Scopus [16, 116, 117], в том числе 2 опубликованы в журналах, цитируемых международной базой Web of Science [46, 116], и 13 из которых опубликованы в тезисах докладов [12, 17 – 19, 34 – 38, 40, 45, 47, 64].

Личный вклад. В совместных публикациях с научным руководителем Наумовым А.В. руководителю принадлежат постановки задачи.

Благодарности. Автор выражает глубокую признательность научному руководителю Наумову А.В., а также заведующему кафедры «Теория вероятностей и компьютерное моделирование» МАИ (НИУ) Кибзуну А.И. за ценные критические замечания, позволившие улучшить качество диссертационной работы.

1 Математические модели времени ответа пользователя на задание

1.1 Обзор основных математических моделей в теории тестирования

Теоретические исследования в области математических моделей, описывающих выполнение заданий, играют огромную роль в современных системах дистанционного компьютерного обучения. Современные модели позволяют оценить трудность заданий, способности испытуемых и другие характеристики. В данной работе подробно рассмотрены модели времени ответа на задания, которые позволят нам решить задачу формирования ограниченных по времени тестов

Исследования в области оценки результатов тестов развиваются с начала XX века, опираясь на работы Ч. Спирмена [130], изучавшего когнитивные тесты. Чарльз Спирмен пришёл к выводу, что результаты тестируемых школьников, предположительно не связанных друг с другом, имели корреляцию, причём положительную. Исходя из этого вывода была разработана модель, которая описывала различия в результатах тестов с помощью индивидуального и группового фактора, влияющих на успешность выполнения задания. Теория Спирмена игнорировала большую часть групповых факторов, что влияло на точность при вычислениях с использованием его модели и итоговую правдивость выводов. С развитием двухфакторная модель трансформировалась в двухуровневую иерархическую модель, которая учитывала больше факторов влияния на результаты тестов. В дальнейшем Л. Терстоун [69, 131] и Дж. Гилфорд [90] разработали многоуровневые иерархические модели, игнорирующие наличие общего единственного группового фактора, при этом данные модели учитывали множество других групповых и индивидуальных факторов, что в итоге стало основной многомерного факторного анализа.

Исследования Спирмена легли в основу классической теории тестирования, которая в дальнейшем развивалась его учениками R. Cattell [81], D.Wechsler [142], а также L. Guttman [91, 92], M. Novick [111, 118], J.P. Guilford [90], P. Vernon [138], C. Burt [80], A. Jensen [94], H. Gulliksen [91], U. Yule [153], T.L. Kelley [95], F. Kuder

[97], M. Richardson [124-126]. Классическая теория тестирования позволяла оценить такие параметры тестирования как сложность заданий или способности тестируемых.

Основными принципами классической теории тестирования являются [118]:

1. эмпирически полученный результат измерения X представляет собой сумму истинного результата измерения T и ошибки измерения E, которые неизвестны:

$$X = T + E;$$

2. Истинный результат измерения можно выразить как математическое ожидание $M[X]$ от эмпирически полученного результата:

$$T = M[X];$$

3. Корреляция истинных и ошибочных компонентов по множеству испытуемых равна нулю:

$$\rho_{TE} = 0;$$

4. Ошибочные компоненты двух любых тестов не коррелируют:

$$\rho_{E_1 E_2} = 0;$$

5. Ошибочные компоненты одного теста не коррелируют с истинными компонентами любого другого теста:

$$\rho_{E_1 T_2} = 0.$$

С развитием вычислительных систем и появлением возможности активного их применения для формирования моделей тестирования и получения оценок параметров теста основной концепцией проектирования актуальных методов тестирования стала Item Response Theory (IRT, современная теория тестирования). Идеи IRT в своих работах с середины XX века развивали F.M. Lord [110, 111], G. Rasch [123], P. Lazarsfeld [29, 108], B.D. Wright [109, 151, 152], D. Andrich [69, 70, 71]. Фундаментальной идеей современной теории тестирования является применение логистических функций для описания зависимости вероятности правильного ответа на тестовое задания от трудности тестового задания и способности испытуемого.

Базовыми предположениями IRT являются:

1. существование латентных личностных характеристик, недоступных для явного наблюдения;
2. существование явных параметров доступных для наблюдения и связанных с латентными характеристиками;
3. латентная характеристика, которую оценивают, является одномерной величиной.

В отличие от классической теории тестирования в моделях IRT параметры моделей оцениваются для каждого задания и испытуемого по результатам тестов, которые проходят испытуемые, причем для значений оцениваемых параметров необязательно нормальное распределение.

Одной из самых популярных логистических моделей является однопараметрическая модель Раша [123]:

$$P_j(\theta) = \frac{e^{1,7(\theta-\beta_j)}}{1+e^{1,7(\theta-\beta_j)}}$$

$$P_i(b) = \frac{e^{1,7(\theta_i-\beta)}}{1+e^{1,7(\theta_i-\beta)}}$$

где

- θ_i – уровень подготовленности испытуемого, i – номер испытуемого, $i = 1, \dots, I$, I – общее количество тестируемых;
- β_j – трудность j-го задания, $j = 1, \dots, J$, J – общее количество заданий в тесте;
- множитель 1,7 является масштабирующим, чтобы модель Раша была совместима с моделью Фергюсона, в которой вероятность правильного ответа на выбранное задание с известной трудностью β_j определяется следующим распределением:

$$P_j(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta-\beta_j} e^{\frac{1}{2}x^2}\, dx$$

И уровень подготовленности, и трудность задания являются латентными параметрами.

В модели Раша предполагается, что уровень подготовленности θ_i и уровень трудности β_j измеряются в логитах:

1) $\theta_i = \ln\left(\frac{p_i}{q_i}\right)$, где p_i – отношение количества верных ответов ко всему числу заданий в тесте, $q_i = 1 - p_i$ – отношение количества неверных ответов ко всему количеству заданий в тесте;

2) $\beta_i = \ln\left(\frac{v_j}{u_j}\right)$, где v_j – отношение количества верных ответов на задание к общему количеству ответов на задание; u_j – отношение количества неверных ответов на задание j к общему количеству ответов на это задание.

Пример характеристической кривой для модели Раша при $\beta_j = 0$ на рис.1.1.

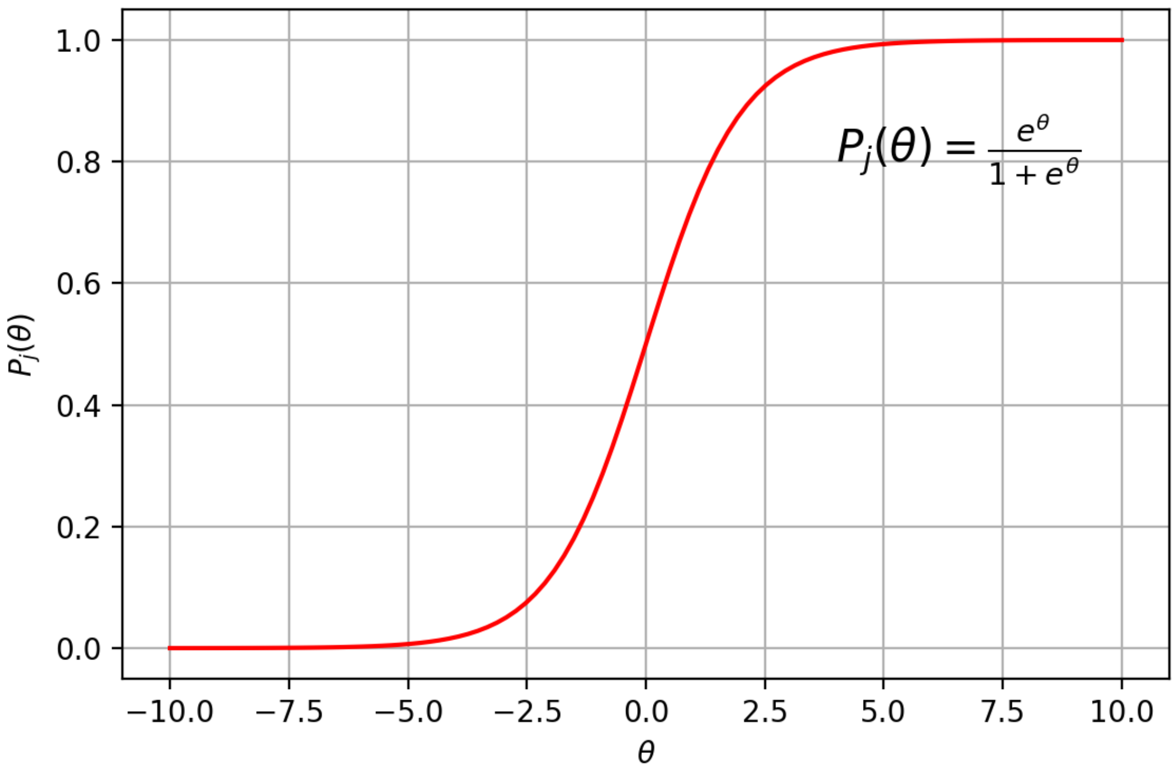

Рис. 1.1 Логистическая кривая модели Раша при $\beta_j = 0$

В свою очередь IRT является основой парадигмы компьютерного адаптивного тестирования (CAT, computerized adaptive testing), которая предполагает адаптацию тестов под способности тестируемого. Основные принципы проектирования компьютерных адаптивных систем в конце XX века описаны в работе D. J. Weiss [96, 143 – 145], G. G. Kingsbury [96, 143], а также в недавних работах W. J. Van der Linden [133 – 137] и H. Wainer [139, 140].

Основными особенностями компьютерного адаптивного тестирования являются:

– возможность получения индивидуального набора заданий каждым испытуемым, т.е. состав заданий в тесте и длительность его прохождения могут различаться для каждого испытуемого;
– возможность оценивать характеристики испытуемого индивидуально, т.е. независимо от остальных испытуемых, при этом допуская минимальную ошибку при оценке характеристик.

Компьютерное адаптивное тестирование также обладает следующими важными преимуществами по сравнению с классическим или очным тестированием:

– для получения оценки уровня подготовленности, при прохождении тестирования испытуемым, требуется значительно меньшее количество заданий;
– получаемые оценки подготовленности по итогам теста позволяют оценить характеристики испытуемого с минимальной ошибкой;
– подобранные задания в индивидуальных наборах соответствуют знаниям и способностям испытуемых;
– внешние факторы меньше влияют на результаты испытуемых, что позволяет повысить точность измерений и уменьшить значение ошибки.

К основным вопросам, исследуемым в области дистанционного обучения, относятся взаимосвязь между временем и корректностью ответа пользователя на задание системы.

Самые первые работы, посвященные взаимосвязи времени и правильности ответа на задание, принадлежат Максу А. Вудбери [149], который применил вероятностные модели к количеству верных ответов, предоставленных пользователем в течение прохождения теста. В дальнейшем эти работы легли в основу публикаций Lord и Novick [77, 110, 111], но в те же годы Gulliksen, [13] выдвинул другое предположение о том, что оценка студента должна быть составлена не только из количества правильных ответов, полученных при прохождении тестирования. Галексеном в его работах утверждалось, что при оценке стоит обязательно учитывать время, которое потребовалось студенту для решения тестовых заданий, поэтому учёным было предложено использовать различные виды тестов: тесты на скорость прохождения (speed tests) и тесты на сложность задач (power tests). Если студенту предлагали пройти первый вид теста, то, располагая ограниченным временем, тестируемому следовало решить максимально возможное число заданий самого низкого уровня сложности, что тем самым помогало оценить скорость, с которой студент принимает решение (благодаря минимизированному влиянию сложности задания на процесс). В случае, когда предлагался тест второго вида, сложности заданий были разных уровней, но при этом время на прохождение теста не ограничивалось, то есть целью прохождения теста являлась оценка уровня знаний независимо от времени.

Данный подход с разделением тестов на описанные подвиды обладает существенными недостатками. В первом случае, когда происходит прохождение теста на скорость, существует вероятность ошибки даже у самого подготовленного и способного студента, следствием чего может стать ошибочная оценка тестируемого при очевидности уровня его знаний. Во втором случае не учитывается скорость, требуемая для решения заданий, что может стать решающим фактором при оценивании студентов, имеющих одинаковые показатели по решенным заданиям, но требуется выделить кого-то из тестируемых.

Среди первых исследователей, пытавшихся решить данные недостатки, был Терстоун [138]. Терстоун представил графическую модель взаимосвязи скорости прохождения теста студентом и уровня знаний. Данная модель была названа «поверхностью ответа» (response surface), представленной на рис. 1.2. Для каждого конкретного студента и одного задания из теста поверхность ответа представляет собой график зависимости вероятности верного ответа от сложности задания и времени, затраченного на ответ.

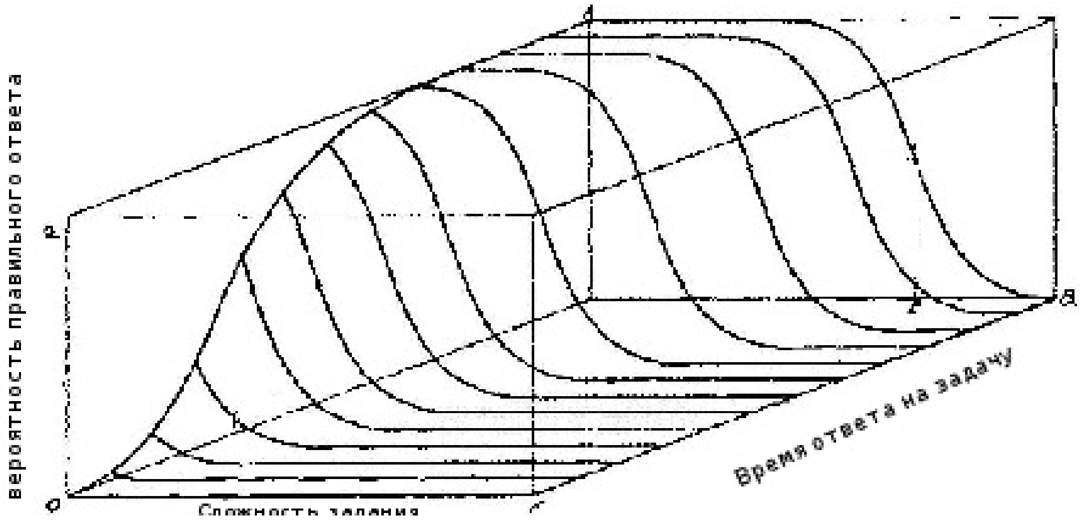

Рис. 1.2. Поверхность ответа

График поверхности в соответствии с рис. 1.2 отражает зависимость между вероятностью правильного ответа и временем, требуемым на решение задания одним студентом, что не подходит для применения модели к групповому тестированию. Закономерность, использованная Терстоуном для построения поверхности, следующая: вероятность правильного ответа, данного студентом, растёт, если им на решение задания затрачивается больше времени, но при этом, если увеличивается сложность задания, то вероятность правильного ответа уменьшается.

Терстоун также ввёл понятия скорости студента и способности студента. Скоростью тестируемого является число заданий, решенных за единицу времени. Способность студента была определена как сложность заданий, на которые тестируемый отвечает с вероятностью $P = 0.5$ при условии, что время ответа студента не ограничивается.

Данная закономерность проста и очевидна, что является преимуществом этой модели, но существуют также неочевидные недостатки. Одним из таких недостатков является то, что для ответов студентов на задания используется вероятностная (стохастическая) модель, но при этом время ответа принимается как детерминированная величина, что не соответствует действительности, то есть построение поверхности должно учитывать совместное распределение двух случайных параметров: времени ответа на задание и корректности ответа. Следующим недостатком является то, что в модели в качестве параметров вероятностной модели ответа принимаются параметры задачи (сложность) и студента (способность) и не принимается параметр, отвечающий за время, потраченное тестируемым на решение задания. Другим недостатком можно считать, что модель предполагает зависимость вероятности правильного ответа от времени ответа, но в таком случае не учитывается, что время, отведенное на решение каждой конкретной задачи, не зависит от предыдущей статистики по прохождению заданий из теста, то есть не зависит ни от правильности ответа на предыдущие задания, ни от времени, которое было затрачено на решение.

Существуют и прочие эвристические модели, учитывающие время ответа пользователей, проходящих тестирование. Такие методы были описаны в статье Вайса и Линлина Ма [147], а также в работах Гавирии [87], который разработал логнормальную модель времени ответа. Многие работы посвящены поведению студентов во время тестирования, например, работы Вайса и Конга [143-145]. Американскими исследователями Вангом и Хансоном [141] была рассмотрена модель, включающая время ответа в качестве параметра, и разработан алгоритм, позволяющий в режиме реального времени конструировать вариант из пула заданий с учётом индивидуальных особенностей студента.

При построении моделей, описывающих поведение пользователя в процессе ответа на задание, выделяются два различных основных принципа:

- время ответа и распределение вероятности правильного ответа тестируемого используются в одном и том же уравнении;
- используются несколько уравнений, одно из которых отвечает за время ответа.

Рассмотрим, например, модель Раша [123], которая основана на втором принципе. Модель Раша, называемая также моделью процесса чтения, состоит из нескольких уровней. К моделям низкого уровня относятся модель ошибок в чтении и модель скорости чтения. Модель ошибок описывается числом ошибок чтения a в тексте длиной N слов. Число ошибок при условии, что текст имеет длину N слов, имеет пуассоновское распределение, функция вероятности которой имеет вид

$$P(a|N) = e^{-\lambda}\frac{\lambda^a}{a!}, \tag{1.1}$$

где N – длина слов, a – количество ошибок чтения, $\lambda = N\rho$ – среднее число ошибок. Раш рассматривал коэффициент ρ как отношение

$$\rho = \frac{\delta_i}{\xi_j}, \tag{1.2}$$

где δ_i – сложность текста, ξ_j – способность студента j.

Время, за которое текст будет прочитан студентом, при этом имеет пуассоновское распределение с функцией плотности вероятности

$$P(t|N) = e^{-\lambda t}\frac{(\lambda t)^{N-1}}{(N-1)!}, \tag{1.3}$$

где λ – параметр интенсивности потока, который соответствует скорости чтения студента. Таким образом, обе модели являются пуассоновскими потоками разного происхождения.

Далее будет приведена более сложная двухуровневая модель, разработанной В. ван дер Линденом [133 – 137], которая использовалась в данной работе в качестве одной из моделей в задаче формирования ограниченного по времени теста.

1.2 Модель ван дер Линдена

1.2.1 Описание модели

В основе двухуровневой модели В. ван дер Линдена [133 – 137] были использованы следующие основополагающие принципы:

- время ответа t_{ij} студента j на задачу i является реализацией случайной величины T_{ij}
- правильность (корректность) ответа v_{ij} является реализацией случайной величины V_{ij}, принимающей значения:

$$V_{ij} = \begin{cases} 1, & \text{если студент } j \text{ ответил верно на задачу } i; \\ 0, & \text{если студент } j \text{ ответил неверно на задачу } i. \end{cases}$$

- время, затрачиваемое студентом на ответ на задание теста, и скорость, с которой выполняются студентом задания являются неэквивалентными понятиями: время меняется в зависимости от параметров задания, скорость студента неизменна, откуда следует следующее уравнение:

$$\tau_j^* = \frac{\beta_i^*}{t_{ij}}, \tag{1.4}$$

где β_i^* - трудозатраты, требуемые студенту j для решения задачи i, τ_j^* - скорость студента.

Для удобства в использовании модели распределение приводят к симметричному виду при помощи логарифмического преобразования над временем ответа:

$$\ln t_{ij} = \beta_i - \tau_j, \tag{1.5}$$

где $\beta_i = \ln \beta_i^*$, $\tau_j = \ln \tau_j^*$ - параметры в логарифмическом масштабе.

1.2.2 Иерархическая модель

Основываясь на представленных в разделе 1.2.1 принципах, ван дер Линден предложил использовать для моделирования процесса обучения студента двухуровневую иерархическую модель, представленную на рис. 1.3:

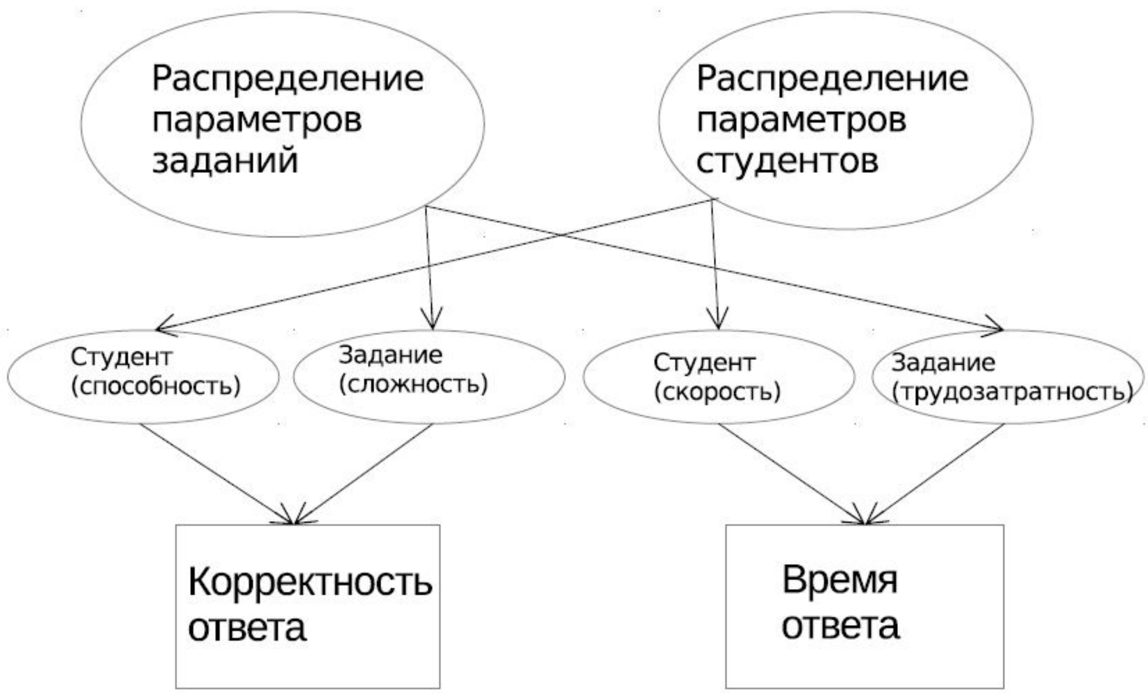

Рис. 1.3 Двухуровневая иерархическая модель

На первом уровне двухуровневой модели находятся вероятностные модели правильности (корректности) ответа и времени ответа студента. На следующем уровне вероятностная модель распределения параметров моделей первого уровня – способность студента и его скорость для студентов всех групп и вероятностная модель распределения параметров сложности и трудозатрат для каждой задачи из пула.

Время ответа студента в такой модели имеет логнормальное распределение:

$$\ln T_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0; \sigma^2). \tag{1.6}$$

Следовательно, плотность распределения времени ответа примет следующий вид:

$$f\left(t_{ij}; \tau_j, \beta_i, \sigma\right) = \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln t_{ij} - \left(\mu + \tau_j + \beta_i\right)}{\sigma}\right]^2\right\}. \tag{1.7}$$

В данной модели: β_i – временной параметр, индивидуальный для задания i; τ_j – временной параметр, индивидуальный для студента j; ε_{ij} – случайное отклонение; μ – параметр времени, общий для всего пула заданий и всех тестируемых студентов.

Для модели ван дер Линдена были получены оценки параметров распределения [134, 136]:

$$\hat{\mu} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I} \ln t_{ij}}{I \cdot J}, \tag{1.8}$$

$$\widehat{\beta_i} = \frac{\sum_{j=1}^{J} \ln t_{ij}}{J} - \hat{\mu}, \tag{1.9}$$

$$\widehat{\tau_j} = \frac{\sum_{i=1}^{I} \ln t_{ij}}{I} - \hat{\mu}, \tag{1.10}$$

$$\widehat{\sigma^2} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\ln t_{ij} - \widehat{\tau_j} - \widehat{\beta_i} - \hat{\mu}\right)^2}{I \cdot J}. \tag{1.11}$$

При использовании модели ван дер Линдена для описания времени ответа на весь тест, а не каждого задания отдельно, возникает сложность определения суммарного случайного времени из-за свойств логнормального распределения, поэтому в работе предлагается, наряду с моделью Ван дер Линдена, использовать также модели с дискретным распределением времени и гамма-распределением.

1.3 Дискретная модель времени ответа

Использование дискретного распределения предполагает, что рассматривается модель случайного времени ответа θ_i некого универсального пользователя на задание i. В такой модели индивидуальные особенности всех пользователей интегрируются в единую модель (универсального пользователя) отражающую особенности всей тестируемой группы, что позволяет предложить эффективные методы решения задачи формирования ограниченного по времени теста для этой группы.

Предположим, что $i -$ ое задание системы решало J_i пользователей. На основе времени, потребовавшегося им для ответа, были построены гистограммы (примеры гистограмм приведены на рис. 1.4 – 1.6).

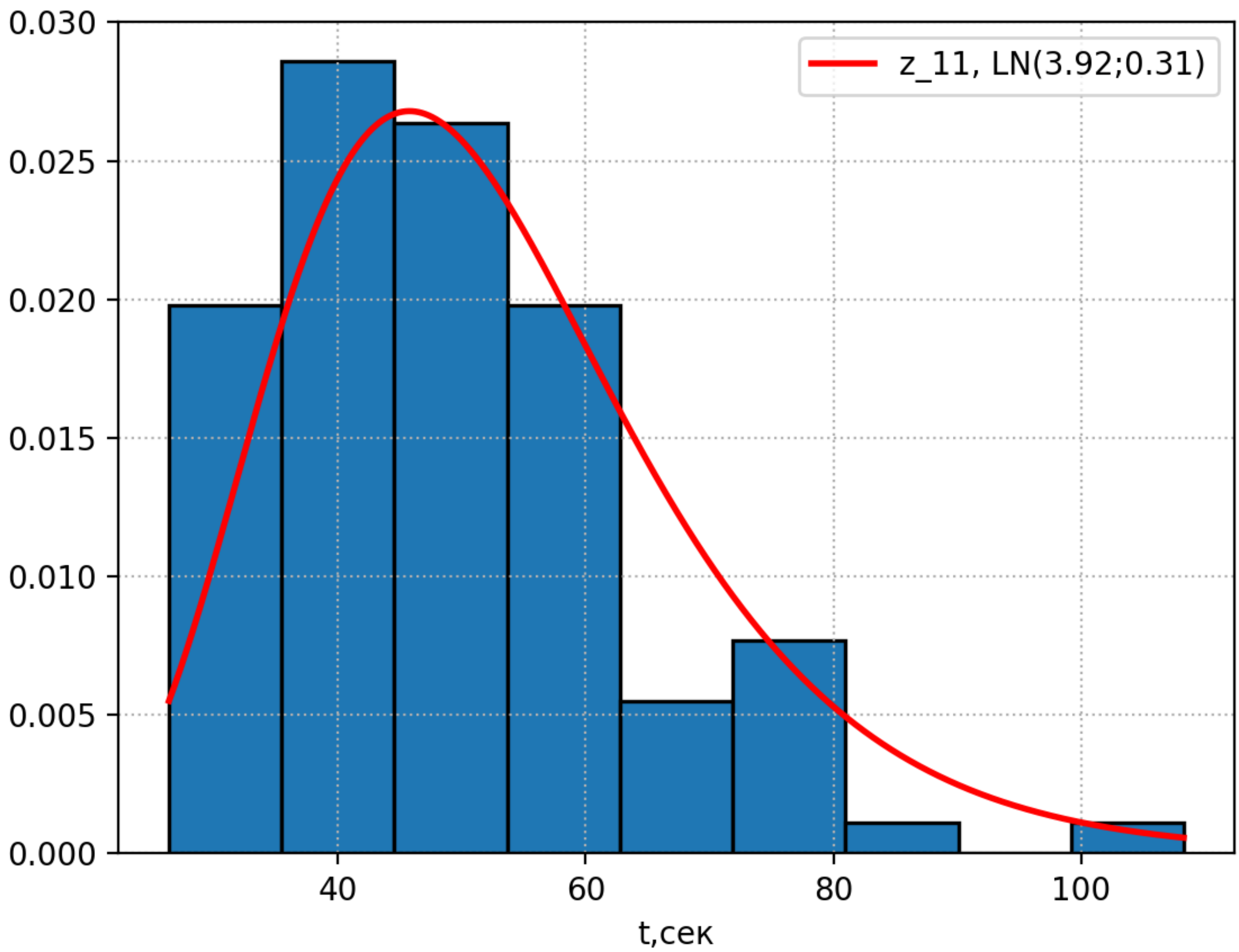

Рис. 1.4 Гистограмма времени выполнения задания низкого уровня сложности

Гистограммы соответствуют распределению времени ответа универсального пользователя на $i -$ ое задание, с помощью которых была выбрана дискретная модель распределения времени ответа на i – ое задание по имеющимся статистическим данным работы студентов в системе дистанционного обучения.

Рассмотрим следующий эксперимент. Выберем $J = 1200$ пользователей системы дистанционного обучения, которые будут выполнять один и тот же набор из $I = 10$ заданий.

С помощью критерия Пирсона на уровне доверия $\alpha = 0.95$ проверена гипотеза о соответствии выборки для каждой задачи логнормальному закону.

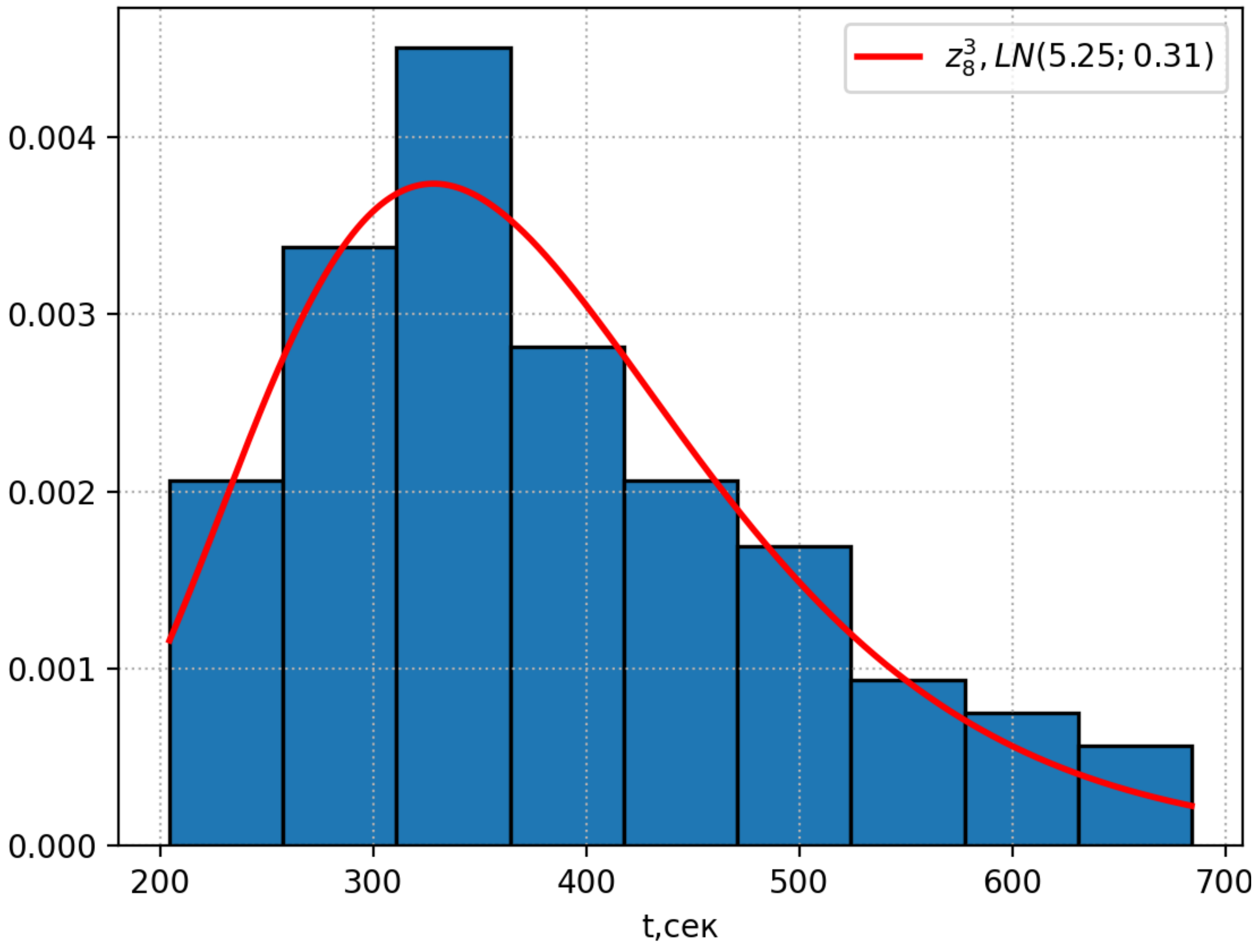

Рис. 1.5 Гистограмма времени выполнения задания среднего уровня сложности

Методом максимального правдоподобия были получены оценки параметров логнормального распределения для каждой задачи и затем были посчитаны значения статистики критерия χ^2 и построены доверительные области, позволившие сделать выводы о принятии гипотезы о логнормальности времени ответа на задания для всех заданий пула. Таким образом была обоснована согласованность выбираемых дискретных распределений, с известной ранее классической моделью времени ответа.

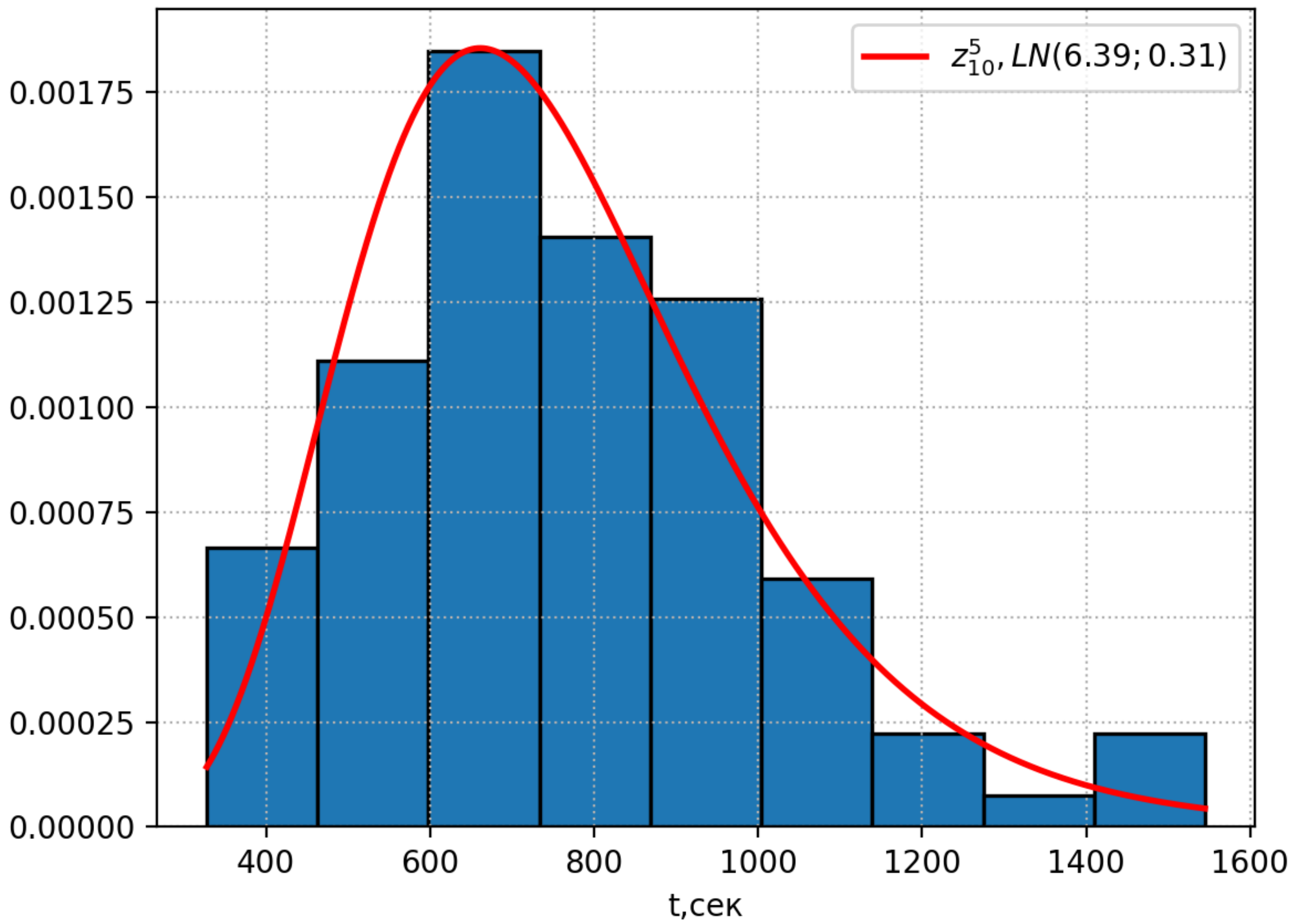

Рис. 1.5 Гистограмма времени выполнения задания высокого уровня сложности

На основе гистограмм, были выбраны дискретные модели времени ответа. Далее в диссертации и дискретная, и непрерывная рассмотренные модели времени ответа пользователя на задание будут использованы при формировании математической постановки задачи по выбору индивидуального набора заданий с ограничением на время выполнения пользователем этого набора заданий.

1.4 Модель времени с гамма-распределением

Оценивание параметров модели ван дер Линдена $\mu,\ \beta_i,\ \tau_j,\ i = 1, \dots, I, \quad j = 1, \dots, J$ не представляет существенных трудностей и эффективность их оценивания обеспечивается классическими методами максимального правдоподобия или наименьших квадратов/модулей. Выражения получаемых оценок приведены, например, в разделе 1.2.1.

Достоинством данной модели является возможность получения распределения времени ответа каждого пользователя на каждое задание системы по имеющейся разрозненной статистике, так как далеко не каждый пользователь решал каждое задание. Известным недостатком данной модели является отсутствие возможности получения точного значения квантили общего времени выполнения теста пользователем, которое представимо как сумма случайных величин, соответствующих времени ответа пользователя на задания теста. В дальнейшем в разделе 2.5 будет обоснована важность квантили как вероятностной характеристикой в задаче поиска оптимального набора тестов для систем компьютерного тестирования, т.к. квантиль дает понимание сколько гарантированно с заданным уровнем доверительной вероятности потребуется времени пользователю для выполнения теста.

Преодолеть указанный недостаток предлагается использованием для модели времени ответа пользователя на задание гамма-распределения с плотностью вероятности вида

$$f(t_i, \kappa_i, \lambda) = t_i^{\kappa_i - 1} \frac{e^{-t/\lambda}}{\lambda^{\kappa_i} \Gamma(\kappa_i)},$$

где $\Gamma(\kappa)$ – гамма-функция Эйлера.

При этом мы оцениваем параметры данной модели для каждого задания системы на основе выборки, состоящей из значений времени, затраченного на решения данного задания пользователями, решавшими его. Таким образом, мы получаем модель времени ответа универсального (усредненного) пользователя на каждое задание системы.

Плотности вероятности логнормального и гамма-распределений имеют схожие структуры (рис.1.7), однако известно, что сумма случайных величин, имеющих гамма-распределение, является гамма-распределенной случайной величиной, если эти случайные величины имеют одинаковый параметр λ.

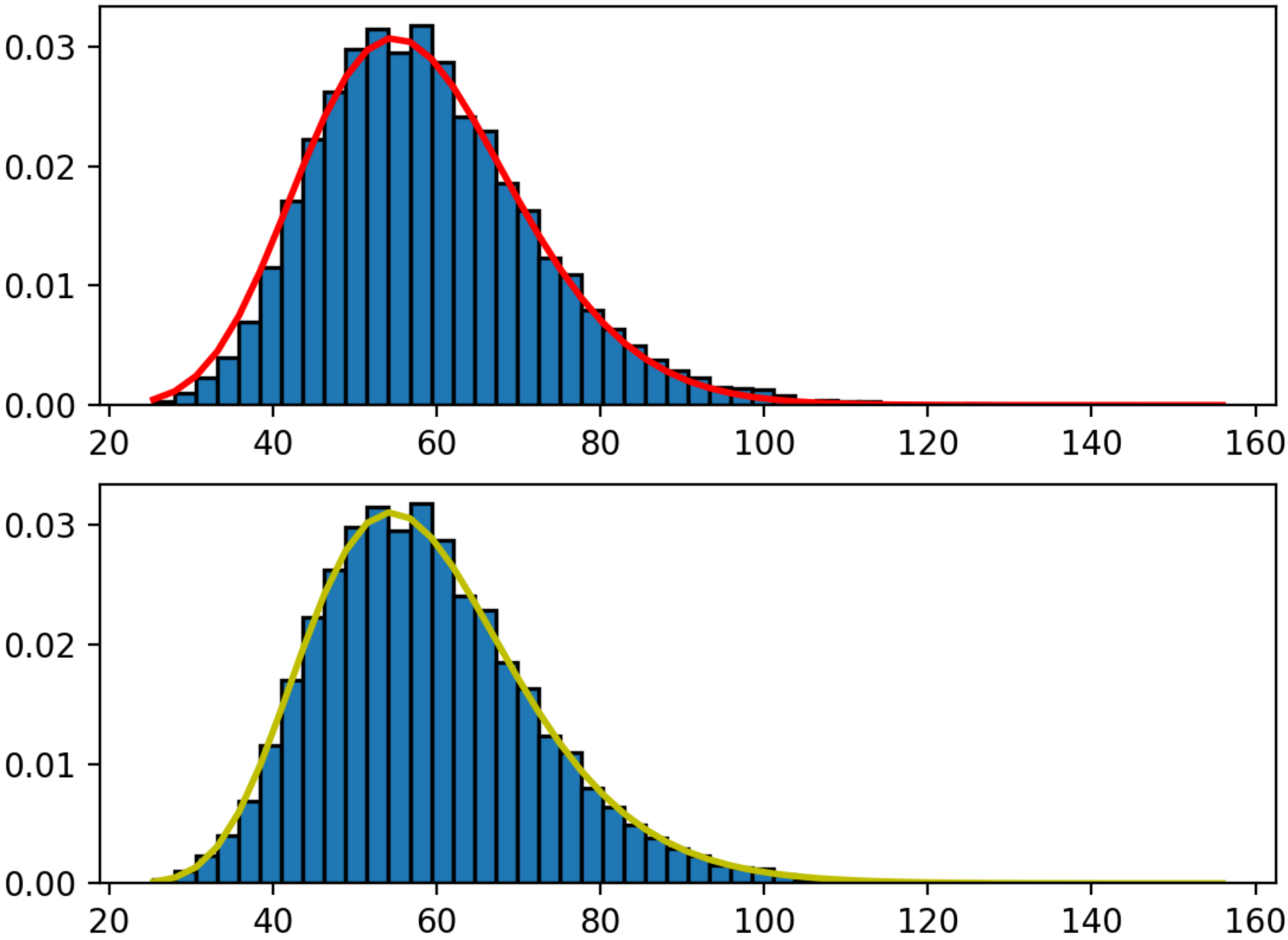

Рис. 1.7 Графики плотностей гамма-распределения (сверху) и логнормального (снизу) относительно гистограммы времени ответа на задание

Это обеспечивает возможность находить точные значения квантили общего времени, затрачиваемого пользователем на решение теста, так как оно будет иметь известное гамма-распределение. Таким образом, ключевое значение для предлагаемой модели обеспечит оригинальный алгоритм подбора параметров гамма-распределения времени ответа универсального пользователя на задание системы, рассматриваемый в главе 2. Этот алгоритм формулируется так, чтобы параметр λ распределения был бы одинаковым для всех заданий, а значение второго параметра κ определялось бы с помощью метода максимального правдоподобия.

1.5 Выводы по главе 1

1) Рассмотрены основные модели тестирования, учитывающие время как характеристику при описании тестового набора или испытуемого, обозначены основные недостатки данных моделей и обоснована потребность в разработке новых моделей и методов анализа времени, затрачиваемого на тестирование. В том числе логнормальная модель, которая позволяет учесть индивидуальные особенности испытуемого, при этом обосновано неудобство модели ван дер Линдена из-за отсутствия свойства суммируемости логнормальных величин и сложности в получении распределения их суммы.
2) Предложена дискретная модель времени ответа на задание для универсального пользователя, которая учитывает характеристики группы пользователей. Построены распределения для заданий и проведён эксперимент, подтверждающий адекватность модели
3) Предложена модель времени ответа с гамма-распределением. Данная модель предложена из-за схожести структуры плотностей распределения с логнормальным распределением, но гамма-распределение позволяет решить проблему суммируемости случайного времени и получения точного выражения для распределения времени, которое требуется для выполнения всего теста. Также описан принцип вычисления параметров гамма-распределения. Для этого нужно зафиксировать один из параметров гамма-распределения, и вычислять оценку максимального правдоподобия для второго параметра

2 Методы и алгоритмы формирования адаптивных тестов с учетом времени ответов пользователей

2.1 Задача определения набора заданий приблизительно равных по суммарной сложности с ограничением на время выполнения

Пусть существует множество $Z = (z_1, \dots, z_I)$ из I заданий, разделённых на M различных типов $Z_m, m = 1, \dots, M$. Для обозначения принадлежности задания к определенному типу введём матрицу A размерности $I \times M$:

$$A = \|a_{im}\|, a_{im} = \begin{cases} 1, & z_i \in Z_m, \\ 0, & z_i \notin Z_m. \end{cases}$$

Данная матрица определяет принадлежность задания z_i к типу $Z_m, m = 1, \dots, M$, если $a_{im} = 1$.

Каждое из заданий имеет определенную различную сложность, которую, например, можно определить с помощью метода максимального правдоподобия, примененного к модели Раша. Введем вектор $u \in \mathbb{R}^I$, координаты которого $u_i, i = 1, \dots, I$, обозначают принадлежность задания i к формируемому набору таким образом, что:

$$u_i = \begin{cases} 1, & \text{если задача } i \text{ попала в тестовый набор;} \\ 0, & \text{если задача } i \text{ не попала в тестовый набор.} \end{cases} \tag{2.1}$$

Тестовым набором будут считаться k заданий, для которых $u_i = 1$. Предположим, что для каждого задания известна её сложность, введём вектор $w \in \mathbb{R}^I$, i-ая координата которого является сложностью задания i и будет обозначена как w_i.

Требуется составить множество индивидуальных тестовых наборов из k задач, принадлежащих различным типам, учитывая, что $k \leq M$. При этом изначально задаётся суммарная сложность теста, обозначаемая c, которая вычисляется с помощью анализа сложности каждого задания на основе экспертной оценки. Предусмотрим, что возможно отклонение от данной требуемой суммарной сложности на какое-либо малое число в большую либо меньшую сторону.

Обозначим такое число ε. Основываясь на описанной модели и введенных обозначениях, сформулируем следующую задачу:

$$U_* = \underset{u}{\mathrm{Argmin}}|c - w^T u| \tag{2.2}$$

$$e_I^T u = k \tag{2.3}$$

$$c - w^T u \geq -\varepsilon \tag{2.4}$$

$$c - w^T u \leq \varepsilon \tag{2.5}$$

$$A^T u \geq e_m \tag{2.6}$$

где $e_I^{\mathrm{T}} \in \mathbb{R}^I, e_I^{\mathrm{T}} = (1, \dots, 1)$, $e_m \in \mathbb{R}^M$, $e_M^{\mathrm{T}} = (1, \dots, 1)$, $(\cdot)^{\mathrm{T}}$ - операция транспонирования.

Функция в (2.2) является критериальной в исследуемой задаче. В данной постановке целью исследования является поиск оптимальных наборов заданий, удовлетворяющих ограничениям (2.3 – 2.6) при условии, что задано фиксированное отклонения ε сложности $w^{\mathrm{T}} u$ подобранного набора заданий в тесте от требуемой суммарной сложности c. Ограничение (2.3) значит, что в наборе должно быть ровно k заданий. Ограничения (2.4) и (2.5) обозначают границы выбора набора заданий в тесте в зависимости отклонения от требуемой суммарной сложности. Ограничение (2.6) отвечает за то, чтобы среди всех заданий в тесте было хотя бы одно задание каждого типа, так как данная задача решается при условии, что $k \geq M$.

Количество наборов в решении задачи поиска допустимого множества наборов может сильно меняться в зависимости от значения ε отклонения от суммарной сложности: в случае, если значение ε близко к 0, решением задачи может стать пустое множество наборов, и, наоборот, при $\varepsilon \to +\infty$ количество подходящих наборов растёт. Значение параметра ε может определяться экспертным образом с учётом количества испытуемых в группе и требуемой сложности теста.

Данная задача решается стандартными пакетами решения задач линейного программирования, например IBM CPLEX.

В развитии этой модели предлагается учесть вероятностное ограничение на время выполнения пользователем заданий теста.

Пусть в тестировании участвуют J пользователей. Обозначим через T_{ji} случайное время, которое потребуется пользователю $j, j = 1, \dots, J$ на решение i-ой задачи, где $i = 1, \dots, I$. Рассмотрим матрицу T размерности $J \times I$:

$$T = \|T_{ji}\|.$$

Обозначим общее время, выделяемое пользователям на выполнение теста через t. Это время определятся экспертом или стандартными ограничениями, например, академическим часом. Тогда в случае непрерывной модели распределения модели получим следующее дополнительное вероятностное ограничение в задаче формирования тестов с заданной суммарной сложностью:

$$P(Tu \ \le \ t_f) \ge \alpha, \tag{2.7}$$

где $t_f \in \mathbb{R}^J, t_f = (t, \dots, t)^{\mathrm{T}}$.

В случае для универсального пользователя будет использовано следующее ограничение:

$$P(\Theta^{\mathrm{T}} u \le t) \ge \alpha,$$

где Θ – вектор из случайных времен ответа на задания с дискретным распределением Θ_i для задания i на интервалах времени $l = 1, \dots, L$ (табл. 2.1).

Таблица 2.1 Ряд распределения Θ_i

θ_{il}	θ_{i1}	θ_{i2}	$\cdots$	θ_{iL}
p_{il}	p_{i1}	p_{i2}	$\cdots$	p_{iL}

Условие (2.7) обозначает, что с заданной вероятностью α за требуемое время все тестируемые будут иметь возможность выполнить выданный вариант теста.

2.2 Задача формирования ограниченного по времени теста с логнормальным распределением для группы испытуемых

2.2.1 Постановка задачи

Пусть случайное время ответа пользователя на задание системы дистанционного обучения описывается логнормальным распределением, тогда в ограничении (2.7) задачи (2.2) - (2.7) элементы матрицы T являются логнормально распределёнными случайными величинами с плотностью распределения (1.2).

Рассмотрим постановку задачи с ограничением в виде (2.7) при условии, что $k \geq M$, то есть число заданий в тесте совпадает с количеством типов заданий или для некоторых типов количество заданий больше одного.

Сформулированную задачу возможно решить, используя доверительный метод [13], но из-за особенностей свойств логнормального распределения возникают сложности в применении известных методов, основанных, например, на использовании аппроксимации оптимального доверительного множества. Так как количество всех возможных наборов по k заданиям из общего числа I конечно и равно C_I^k, простейшим является решение задачи полным перебором всех возможных наборов заданий и последующая проверка выполнения ограничений (2.3) - (2.6). Условие (2.7) проверяется методом Монте-Карло. Данный способ требует большого количества вычислений. Для ускорения процесса перебора предлагается алгоритм, основанный на идее метода ветвей и границ.

Алгоритм 2.1

Пусть общее число заданий I разделено на M типов, I_m- число заданий m-ого типа, тогда

$$\sum_{m=1}^{M} I_m = I, \qquad m = 1, \dots, M.$$

Рассмотрим следующий *алгоритм ускоренного поиска* допустимых решений задачи:

1. Выбрать и зафиксировать задачу из I_1 задач первого типа; задать $s = 1$ счётчик уровня дерева, $s = 1, \dots, k$.
2. Построить дерево наборов заданий, проведя ветви к задачам следующего типа и суммируя сложности задач вдоль ветви;
3. Проверяя выполнение условия (2.5) на каждом уровне $s < M$ дерева, отсечь ветви, которые не удовлетворяют условиям. При удовлетворении условиям повторить п.2, $s := s + 1$. При достижении последнего уровня $s = M$ перейти к п.4.

4. Проверить на уровне $s = M$ выполнение условий (2.3) - (2.6), а также условия (2.7) методом Монте-Карло.
5. Если $s = M$, перейти к п.1 и повторить алгоритм для следующего задания первого типа. Если $s > M$, перейти к п.6.
6. Достроить дерево наборов заданий, проведя ветви к $C_{(I-M)}^{(k-M)}$ комбинациям, проверить выполнение условий (2.3) - (2.5) и (2.7). Перейти к п.1 и повторить алгоритм для следующего задания первого типа.

Данный алгоритм позволяет в случае, если $k = M$, уменьшить число рассматриваемых заданий, для которых необходимо проверить вероятностное ограничение, в зависимости от параметров сложности w и отклонения ε от требуемой суммарной сложности, поэтому особенностью алгоритма является зависимость скорости выполнения от параметра ε: чем меньше значение параметра отклонения от заданной суммарной сложности, тем быстрее происходит поиск.

Прокомментируем п.5-6 алгоритма: в случае, когда $k > M$, применение предложенного алгоритма ускоренного поиска позволяет гарантировать выполнение условия (2.6), которое обеспечивает присутствие в тесте хотя бы одной задачи каждого типа. При этом могут быть отсеяны заведомо неподходящие наборы тестовых заданий, суммарная сложность которых не удовлетворяет условию (2.5), что даст возможный выигрыш по времени поиска решений задачи. Так как тест к моменту выполнения п.5 уже содержит M заданий, выбранные задания следует исключить из общего числа I в п.6.

2.2.2 Результат вычислительного эксперимента

Рассмотрим задания $M = 5$ различных типов, в каждом из которых по 10 различных заданий $I_i = 10, i = 1, \dots, 5$. Обозначим задания в зависимости от типа $m = 1, \dots, M$ и номера $i = 1, \dots, I_m$ как z_i^m. Данные были получены из системы дистанционного обучения CLASS.NET МАИ (НИУ) [55].

Была проведена оценка сложности каждого задания из общего количества при помощи алгоритма, основанного на модели Раша и описанного в [123]. В результате были получены приведенные к десятибалльной шкале оценки значений сложностей w_i^m для каждого z_i^m, которые представлены в таблице 2.2.

Исходя из оценок значений сложности для каждого задания, выберем требуемую суммарную сложность $c = 29.5$ теста из $k = 5$, возможный критерий выбора которой был описан в [46], и параметр отклонений от требуемой суммарной сложности $\varepsilon = 0.001$. Параметр ε задаётся числом близким к 0 для выбора наиболее оптимальных наборов тестов.

Для каждого задания и каждого студента были получены оценки значений параметров логнормального распределения, которые приведены в таблицах 2.3-2.4

Общее время, отведенное на выполнение теста, определяется преподавателем либо экспертом на основе анализа смоделированного времени ответа на каждую задачу из общего числа в системе или прочих факторов. Примем $t = 45\text{мин} = 2700c$, что соответствует 1 академическому часу.

Таблица 2.2 Сложность w_i^m задания z_i^m

m / i	1	2	3	4	5
1	1,3113	4,132	2	1,9694	1,3131
2	3,2545	6,9023	2,4182	3,1841	1,3131
3	3,2545	2,1219	2,6666	3,1841	1,3131
4	3,2545	3,4366	3,6536	5,7378	4,2146
5	4,8738	2,4562	5,2425	5,7378	5,0101
6	5,3677	5,3587	5,5474	6,3872	5,3428
7	7,0109	6,9023	6,4528	6,9135	5,7649
8	7,2167	7,2833	7,1938	7,4123	6,2283
9	8,2444	7,8147	8,7945	9,6663	7,4139
10	9,6356	9,3991	3,6571	9,7777	9,6678

Таблица 2.3 Значения параметров логнормального распределения случайных величин $t_{ij} \sim LN(\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j; 0.31)$ для задач типа $m = 1, \ldots, 3$

z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$	z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$	z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$
z_1^1	3,51	z_1^2	4,72	z_1^3	3,65
z_2^1	3,92	z_2^2	5,87	z_2^3	3,73
z_3^1	3,89	z_3^2	3,83	z_3^3	3,87
z_4^1	3,91	z_4^2	3,91	z_4^3	3,96
z_5^1	4,22	z_5^2	3,87	z_5^3	4,84
z_6^1	4,63	z_6^2	5,13	z_6^3	4,95
z_7^1	5,67	z_7^2	5,25	z_7^3	5,53
z_8^1	5,71	z_8^2	5,71	z_8^3	5,89
z_9^1	6,13	z_9^2	5,94	z_9^3	6,18
z_{10}^1	6,39	z_{10}^2	6,27	z_{10}^3	3,88

Таблица 2.4 Значения параметров логнормального распределения случайных величин $t_{ij} \sim LN(\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j; 0.31)$ для задач типа $m = 4$ и $m = 5$

z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$	z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$
z_1^4	3,54	z_1^5	3,61
z_2^4	3,74	z_2^5	3,52
z_3^4	3,76	z_3^5	3,54
z_4^4	4,71	z_4^5	4,51
z_5^4	4,82	z_5^5	4,72
z_6^4	4,98	z_6^5	4,98
z_7^4	5,31	z_7^5	5,33
z_8^4	5,75	z_8^5	5,76
z_9^4	6,28	z_9^5	6,08
z_{10}^4	6,34	z_{10}^5	6,39

Задача решалась на уровне доверия $\alpha = 0.95$.

Требуется составить наборы тестовых заданий из 5 задач, которые с вероятностью 0.95 будут решены всеми пользователями СДО за 1 академический час.

Для решения задачи в случае с непрерывным распределением в вероятностном ограничении модифицированным методом ветвей и границ была разработана процедура на языке программирования MATLAB, использующая стандартные средства пакета. В главе 3 описана реализация процедуры на Python ввиду простоты интеграции в web-приложения библиотек на этом языке программирования. Расчёты проводились на компьютере MSI GE60 2PL Apache (Intel Core i5 2.5 GHz, 8Gb RAM). Решением задачи является множество из 16 индивидуальных наборов тестов по 5 заданий в каждом тесте. Полученные наборы представлены в таблице 2.5.

Среди множества тестовых наборов оптимальным является набор 8* (табл. 2.5).

В случае непрерывной модели времени решение задачи полным перебором было осуществлено за 106 секунд, а время, требуемое предложенным в данной работе алгоритмом, составило 39 секунд. Существенное различие возникает из-за количества перебираемых полных наборов заданий: в одном случае их 100000, в другом 16. Такое количество наборов обусловлено выбором значения параметра $\varepsilon = 0.001$.

При увеличении значения параметра ε количество полных наборов, которые перебирает алгоритм ускоренного поиска, значительно увеличивается. Пусть R – количество наборов, удовлетворяющих все ограничениям задачи, решаемой алгоритмом 1.

График, показывающий данную зависимость между количеством наборов тестовых заданий R и значением параметра ε отклонения от суммарной сложности представлен на рис. 2.1.

Таблица 2.5 Результат вычислительного эксперимента

Номер набора	Выбранные задания z_m^i	Суммарная сложность c
1	$(z_1^1, z_9^2, z_6^3, z_8^4, z_9^5)$	29.4996
2	$(z_2^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
3	$(z_3^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
4	$(z_4^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
5	$(z_6^1, z_9^2, z_4^3, z_{10}^4, z_6^5)$	29.5005
6	$(z_6^1, z_{10}^2, z_4^3, z_4^4, z_6^5)$	29.5010
7	$(z_6^1, z_{10}^2, z_5^3, z_4^4, z_6^5)$	29.5010
8*	$(z_7^1, z_9^2, z_2^3, z_7^4, z_6^5)$	29.5001
9	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_1^5)$	29.5008
10	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_2^5)$	29.5008
11	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_3^5)$	29.5008
12	$(z_9^1, z_9^2, z_5^3, z_1^4, z_8^5)$	29.4993
13	$(z_{10}^1, z_3^2, z_9^3, z_2^4, z_7^5)$	29.5010
14	$(z_{10}^1, z_3^2, z_9^3, z_3^4, z_7^5)$	29.5010
15	$(z_{10}^1, z_6^2, z_8^3, z_1^4, z_6^5)$	29.5003
16	$(z_{10}^1, z_9^2, z_3^3, z_1^4, z_9^5)$	29.5002

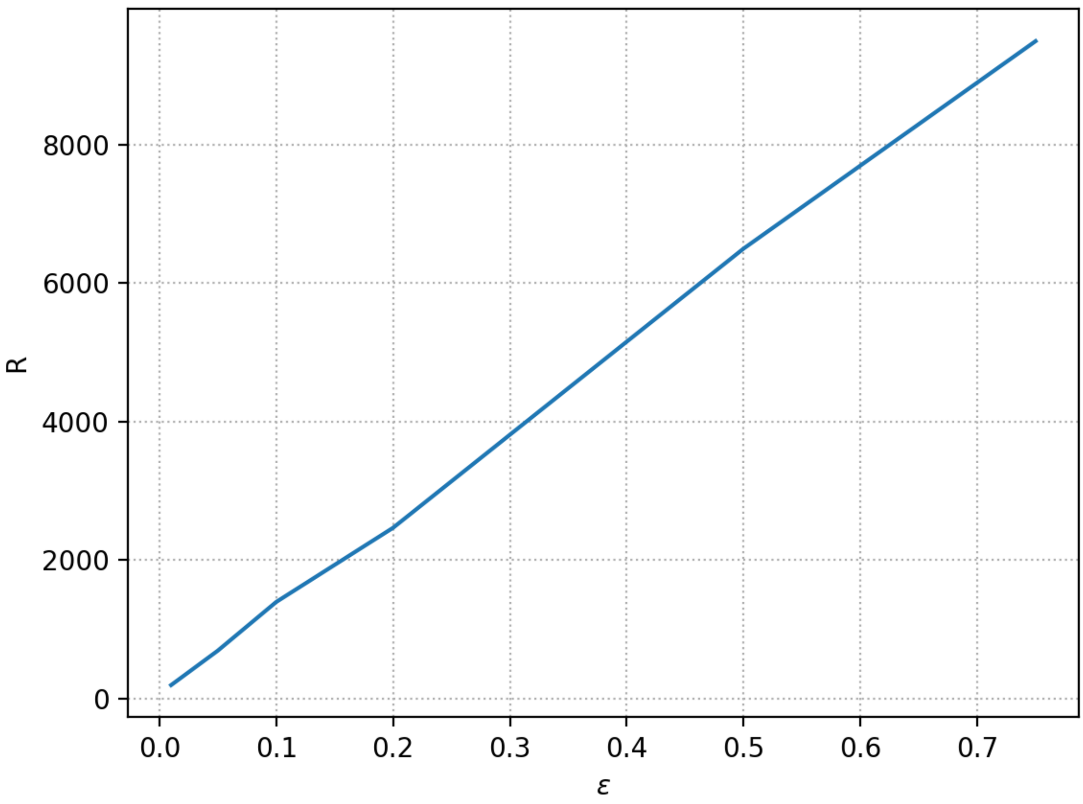

Рис.2.1 График зависимости количества наборов R от значения параметра ε отклонения от суммарной сложности

2.3 Задача формирования ограниченного по времени теста для универсального пользователя с дискретным распределением времени ответа

2.3.1 Постановка задачи

Особенностью использования дискретного распределения является то, что задача формирования тестов решается не для всей группы студентов, а для некоторого универсального пользователя. В такой модели индивидуальные особенности всех пользователей интегрируются в единую модель (универсального пользователя), отражающую особенности всей тестируемой группы, что позволяет предложить эффективные методы решения задачи формирования ограниченного по времени теста для этой группы.

Пусть $\Theta = (\Theta_1, \ldots, \Theta_i)^T$, $i = 1, \ldots, I$, где Θ_i - случайное время необходимое универсальному пользователю для решения i-ого задания. Тогда вместо ограничения (2.7) в задаче (2.2) - (2.7) в случае дискретного распределения будет использовано следующее ограничение:

$$P(\Theta^T u \leq t) \geq \alpha. \tag{2.8}$$

Случайные величины Θ_i, $i = 1, \dots, I$, являются независимыми и их распределения выбираются на основе построенных гистограмм. Типовые гистограммы для некоторых заданий СДО предложены на рис.1.3 - 1.5. Ниже приведены примеры рядов распределения случайного времени ответа универсального пользователя на простое задание Θ^{simple} и на сложное Θ^{hard} на $l = 1, \dots, L$ интервалах времени в таблицах 2.6 - 2.7.

Таблица 2.6 Ряд распределения Θ^{simple}

θ_l^{simple}	24	34	44	54	64	74	84	94	104	114
p_l^{simple}	0.09	0.25	0.32	0.17	0.1	0.01	0.02	0.01	0.01	0.01

Таблица 2.7 Ряд распределения Θ^{hard}

θ_l^{hard}	350	496	642	788	934	1080	1226	1372	1664
p_l^{hard}	0.15	0.33	0.26	0.14	0.06	0.03	0.01	0.01	0.01

В случае дискретной модели распределения времени ответа на задание исходная задача (2.2) – (2.7) с вероятностными ограничениями с помощью доверительного метода [21] сводится к детерминированной задаче целочисленного программирования:

$$U_* = \underset{u, \delta_1, \dots, \delta_D \in \{0,1\}}{\operatorname{Argmin}} |c - w^T u| \tag{2.9}$$

$$e_I^T u = k \tag{2.10}$$

$$c - w^T u \geq -\varepsilon \tag{2.11}$$

$$c - w^T u \leq \varepsilon \tag{2.12}$$

$$(\theta^d)^T u - t \leq (\nu(\theta^d) - t)\delta_d, \tag{2.13}$$

$$p^T \delta \leq 1 - \alpha \tag{2.14}$$

$$A^T u \geq e_m \tag{2.15}$$

где

$$d = 1, \dots, D, \qquad D = \prod_{i=1}^{I} L_i\,, \qquad i = 1, \dots, I;$$

$$\nu(\theta^d) = (\theta^d)^T e, \qquad e = (1, \dots, 1)^T, \qquad e \in \mathbb{R}^I;$$

$$p = (p_1, \dots, p_D), \qquad p_d = P(\Theta = \theta^d) = \prod_{i=1}^{I} P\big(\Theta_i = \theta_i^d\big),$$

$$\delta = (\delta_1, \dots, \delta_D);$$

$\theta^d \in \mathbb{R}^I, d = 1, \dots, D$ – реализация случайного вектора Θ; $p \in \mathbb{R}^D$ является вектором со значениями, равными вероятностям появления соответствующей реализации дискретной случайной величины Θ; $\delta \in \{0, 1\}^D$ – вектор булевых переменных, с помощью которых организуется перебор α-доверительных множеств; L_i – число возможных реализаций случайных величин $\Theta_i, i = 1, \dots, I$.

Рассмотрим определение эквивалентности задач, приведенное в [21].

Определение 1. Две задачи оптимизации вида

$$\Phi(u) \to \inf_{u \in U} \tag{2.16}$$

эквивалентны, если выполнены условия:

1. либо обе эти задачи имеют допустимые решения (с конечными значениями целевых функций), либо обе не имеют таких решений;
2. если эти задачи имеют допустимые решения, то оптимальные значения их целевых функций (конечные или бесконечные) совпадают;
3. если оптимальные значения их целевых функций конечны, то эти значения в обеих задачах либо достигаются, либо не достигаются;
4. если оптимальные значения достигаются, то по оптимальному решению одной задачи с помощью явно указанного алгоритма восстанавливается оптимальное решение другой.

Сформулируем теорему об эквивалентности.

Теорема 1. Задача (2.2) - (2.6) с вероятностным ограничением (2.8) эквивалентна задаче (2.9) - (2.15) целочисленного программирования в смысле определения 1.

Доказательство теоремы 1.

Рассматриваемая задача (2.2) - (2.6), (2.8) является частным случаем общего вида задачи с вероятностными ограничениями, рассмотренной в [21]. Воспользуемся теоремой об эквивалентности, доказанной в этой работе, и проверим выполнение условий теоремы.

Во-первых, требуется, чтобы случайный параметр нашей задачи являлся дискретной случайной величиной (допускается векторного типа) с конечным числом реализаций, что соответствует распределению параметра Θ.

Во-вторых, должны существовать функции:

$$-\infty < \mu_1(u,\theta) \le \inf_{\theta\in\{\theta^d, d=1,\dots,D\}} \Phi(u,\theta), \tag{2.16}$$

$$-\infty < \mu_2(u,\theta) \le \max\left\{0, \inf_{\theta\in\{\theta^d, d=1,\dots,D\}} Q(u,\theta)\right\}. \tag{2.17}$$

В нашей задаче функция

$$\Phi(u,\theta) = |c - w^T u| \tag{2.18}$$

не зависит от θ, поэтому первое условие выполняется автоматически, если выбрать $\mu_1(u,\theta) = |c - w^T u|$.

Функция $Q(u,\theta)$ задана следующим образом:

$$Q(u,\theta) = \theta^T u - t, \tag{2.19}$$

поэтому

$$\mu_2(u,t) = \max\{0, (\theta^1)^T u - t\}, \tag{2.20}$$

где $\theta^1 = (\theta_1^1, \dots, \theta_I^1)$, $\theta_i^1 -$ минимальная реализация СВ θ_i, т.к. все $\theta_i^d \ge 0, d = 1, .., D, i = 1, \dots, I, u_i \in \{0,1\}$.

Тогда из теоремы 1 [21] следует, что задача (2.2) – (2.6), (2.8) эквивалента задаче

$$U_* = \operatorname*{Argmin}_{u,\delta_1,\dots,\delta_D \in\{0,1\}} |c - w^T u| \tag{2.21}$$

$$e_I^T u = k \tag{2.22}$$

$$c - w^T u \ge -\varepsilon \tag{2.23}$$

$$c - w^T u \le \varepsilon \tag{2.24}$$

$$(\theta^d)^T u - t \le (\nu(\theta^d) - t)\delta_d, \tag{2.25}$$

$$p^T\delta \leq 1 - \alpha \tag{2.26}$$

$$A^T u \geq e_m \tag{2.27}$$

где

$$d = 1, \dots, D, \qquad D = \prod_{i=1}^{I} L_i\,, \qquad i = 1, \dots, I;$$

$$\nu(\theta^d) = (\theta^d)^T e, \qquad e = (1, \dots, 1)^T, \qquad e \in \mathbb{R}^I;$$

$$p = (p_1, \dots, p_D), \qquad p_d = P(\Theta = \theta^d) = \prod_{i=1}^{I} P\big(\Theta_i = \theta_i^d\big),$$

$$\delta = (\delta_1, \dots, \delta_D).$$

Случайный вектор Θ составлен из Θ_i дискретных случайных величин и принимает конечное число реализации $\theta^d \in \mathbb{R}^I, d = 1, \dots D,$; $p \in \mathbb{R}^D$ является вектором со значениями, равными вероятностям появления соответствующей реализации дискретного случайного вектора Θ.

Ограничения (2.25) и (2.26) обеспечивают выполнение ограничения под знаком вероятности в (2.8) для реализаций случайного вектора Θ, суммарная вероятность которых больше либо равна α. Если $\delta_d = 0$, то для реализации θ^d выполняется ограничение

$$(\theta^d)^T u \leq t. \tag{2.28}$$

Если $\delta_d = 1$, то ограничение

$$(\theta^d)^T u \leq \nu(\theta^d) \tag{2.29}$$

оказывается пассивным, так как

$$\nu(\theta^d) = \max_{u \in \{0;1\}^I} (\theta^d)^T u. \tag{2.30}$$

Выбор оптимального вектора $\delta \in \{0, 1\}^D$ обеспечивает выбор оптимального набора реализаций θ^d с суммарной вероятностной мерой, превышающей α, то есть оптимального α-доверительного множества.

Поэтому, согласно результатам, полученным в [21], задача стохастического программирования (2.2) - (2.6) с вероятностным ограничением (2.8) по набору оптимальных решений эквивалентна следующей задаче (2.9) - (2.15) целочисленного программирования. Что и требовалось доказать.■

2.3.2 Результат вычислительного эксперимента

В случае задачи с дискретной моделью времени были составлены ряды распределения времени ответа универсального пользователя на каждое задание системы. Примеры таких рядов приведены в таблицах 2.6 и 2.7.

В результате использования дискретной модели времени ответа пользователя были получены допустимые наборы тестовых заданий, совпадающие с допустимыми наборами задачи, основанной на непрерывной логнормальной модели (см. табл. 2.8).

Таблица 2.8 Результат вычислительного эксперимента

Номер набора	Выбранные задания z_m^i	Суммарная сложность c
1	$(z_1^1, z_9^2, z_6^3, z_8^4, z_9^5)$	29.4996
2	$(z_2^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
3	$(z_3^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
4	$(z_4^1, z_9^2, z_4^3, z_{10}^4, z_5^5)$	29.5005
5	$(z_6^1, z_9^2, z_4^3, z_{10}^4, z_6^5)$	29.5005
6	$(z_6^1, z_{10}^2, z_4^3, z_4^4, z_6^5)$	29.5010
7	$(z_6^1, z_{10}^2, z_5^3, z_4^4, z_6^5)$	29.5010
8 (оптимальное)	$(z_7^1, z_9^2, z_2^3, z_7^4, z_6^5)$	29.5001
9	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_1^5)$	29.5008
10	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_2^5)$	29.5008
11	$(z_7^1, z_{10}^2, z_1^3, z_{10}^4, z_3^5)$	29.5008
12	$(z_9^1, z_9^2, z_5^3, z_1^4, z_8^5)$	29.4993
13	$(z_{10}^1, z_3^2, z_9^3, z_2^4, z_7^5)$	29.5010
14	$(z_{10}^1, z_3^2, z_9^3, z_3^4, z_7^5)$	29.5010
15	$(z_{10}^1, z_6^2, z_8^3, z_1^4, z_6^5)$	29.5003
16	$(z_{10}^1, z_9^2, z_3^3, z_1^4, z_9^5)$	29.5002

В условиях задачи (2.2) – (2.6), (2.8) поиск оптимального решения с помощью средств вычислительного пакета IBM CPLEX (импортируется и в MATLAB, и в Python), потребовал 0.14с, что значительно меньше, чем время, которое было затрачено на поиск решений в случае логнормальной модели времени. Расчёты проводились на компьютере MSI GE60 2PL Apache (Intel Core i5 2.5 GHz, 8Gb RAM). В случае поиска множества допустимых решений время решения задачи составило 5с.

2.4 Задача формирования теста с квантильным критерием

2.4.1 Постановка задачи

Пусть в отличие от модели формирования теста с логнормальным временем распределения общее время на выполнение теста в вероятностном ограничении неизвестно. Обозначим его через φ. Тогда для того, чтобы за некоторое оптимальное время все тестируемые могли выполнить выданный вариант теста с заданной вероятностью α, рассмотрим функцию квантили:

$$\Phi_\alpha(u) \triangleq \min\{\varphi \in R^1 \colon \mathrm{P}\{\max_{j=\overline{1,J}} T_j u \le \varphi\} \ge \alpha\}, \tag{2.31}$$

где T_j – j-я строка матрицы T.

Основываясь на описанной модели в разделе 2.3 и введенных обозначениях, сформулируем задачу квантильной оптимизации:

$$u_\alpha = \mathrm{Arg}\min_{u\in\{0;1\}^I}\left(\frac{\gamma|c - w^T u|}{\varepsilon} + \frac{(1-\gamma)\Phi_\alpha(u)}{2700}\right), \tag{2.32}$$

$$\varphi_\alpha = \min_{u\in\{0;1\}^I}\left(\frac{\gamma|c - w^T u|}{\varepsilon} + \frac{(1-\gamma)\Phi_\alpha(u)}{2700}\right), \tag{2.33}$$

$$c - w^T u \le \varepsilon, \tag{2.34}$$

$$w^T u - c \le \varepsilon, \tag{2.35}$$

$$A^T u \ge e_M, \tag{2.36}$$

$$e_I{}^T u = k, \tag{2.37}$$

где $(\cdot)^T$ - операция транспонирования, $e_I \in R^I$, $e_I = (1, \dots, 1)^T$, $e_M \in R^M$, $e_M = (1, \dots, 1)^T$, $\alpha \in (0,1)$ - заданный уровень доверительной вероятности, $\gamma \in$

$[0; 1)$ – весовой коэффициент, $\frac{1}{2700}$ – коэффициент нормировки, величина обратная количеству секунд в 1 академическом часе.

Критериальная функция задачи в (2.32) представляет из себя сумму двух нормированных безразмерных величин. Первое слагаемое является отклонением сложности теста от заданного уровня c, нормированным максимально допустимым уровнем отклонения ε. Второе слагаемое представляет из себя время выполнения теста, которое не может быть превышено с заданным уровнем доверительной вероятности α. Это время нормируется максимально допустимым временем выполнения теста. Такой критерий представляется универсальным гибким инструментом формирования теста. С помощью весового коэффициента γ можно регулировать важность каждого слагаемого критерия. Ограничения (2.34) и (2.35) регламентируют выбор набора заданий в тесте, суммарная сложность которых должна отличаться от заданного экспертом уровня сложности не более чем на величину ε. Ограничение (2.36) отвечает за то, чтобы среди всех заданий в тесте было хотя бы одно задание каждого типа, так как данная задача решается при условии, что $k \geq M$. Ограничение (2.37) означает, что в наборе должно быть ровно k заданий.

Модель Ван дер Линдена, рассмотренная в разделе 1.2, ценна тем, что позволяет получить распределение времени ответа любого пользователя на любое задание системы по имеющейся неполной статистической информации (не все пользователи решали все задачи). Однако использование ее для решения сформулированной задачи не позволяет получить точное решение, так как для задач квантильной оптимизации известны лишь методы поиска гарантирующих решений, основанные на доверительном методе, или плохо сходящиеся стохастические квазиградиентные процедуры, поэтому для решения сформулированной задачи рассматривается дискретную модель распределения времени ответа j-го пользователя на i-е задание, для которой существует способ поиска точного решения путем перехода к задаче смешанного математического программирования.

Дискретизировать полученную ранее модель Ван дер Линдена можно следующим образом.

Зафиксируем номер пользователя и номер задания, а также зафиксируем интервал $(\underline{t}, \bar{t})$ действительной прямой $(-\infty, \infty)$ и назначим $L_{ji} - 1$ порогов дискретизации

$$0 = \underline{t} < t_1 < t_2 < \cdots < t_{l-1} < t_l < \cdots < t_{L_{ji}-1} < \bar{t} = +\infty,$$

разбивающих интервал $(\underline{t}, \bar{t})$ на L_{ji} подинтервалов (t_{l-1}, t_l), $l = 1, \dots, L_{ji}$. Полагаем $t_0 = \underline{t}, t_{L_{ji}} = \bar{t}$.

В этом случае непрерывной случайной величине T_{ji} может быть сопоставлена дискретная случайная величина Θ_{ji}, определяемая рядом распределения:

θ_{ji}^l	Θ_{ji}^1	Θ_{ji}^2	$\cdots$	$\Theta_{ji}^{L_{ji}}$
p_l	p_1	p_2	$\cdots$	$p_{L_{ji}}$

где θ_{ji}^l - середины интервалов (t_{l-1}, t_l), а $p_{jl} = \int_{t_{l-1}}^{t_1} f(t, \tau_j, \beta_i, \sigma)\, dt$ - соответствующие им вероятности $l = 2, \dots, L_{ji} - 1$, а Θ_{ji}^1 и $\Theta_{ji}^{L_{ji}}$ - квантили уровней 0.01 и 0.99.

Таким образом, вместо матрицы T будем использовать матрицу Θ размерности $J \times I$:

$$\Theta = \|\Theta_{ji}\|,$$

все элементы которой являются независимыми случайными величинами с заданными дискретными распределениями.

Тогда вместо функции квантили (2.31) в случае дискретного распределения в задаче (2.32) - (2.37) будет использована следующая функция квантили:

$$\Phi_\alpha(u) \triangleq \min\{\varphi : \mathrm{P}\{\max_{j=\overline{1,J}} \Theta_j u \le \varphi\} \ge \alpha\} \tag{2.38}$$

где Θ_j - j-я строка матрицы Θ, $j = 1, \dots, J$.

В случае дискретной модели распределения времени ответа на задание исходная задача с квантильным критерием может быть сведена к следующей детерминированной задаче смешанного целочисленного программирования:

$$u^* = \underset{u\in\{0;1\}^I, \delta\in\{0;1\}^D, \varphi\geq 0}{\operatorname{Arg\,min}} \left(\gamma \frac{|c-w^T u|}{\varepsilon} + (1-\gamma)\frac{\varphi}{2700}\right), \tag{2.39}$$

$$\theta^d u - \bar{\varphi} \leq (\theta^d e_I)\delta_d, \tag{2.40}$$

$$c - w^T u \leq \varepsilon, \tag{2.41}$$

$$w^T u - c \leq \varepsilon, \tag{2.42}$$

$$A^T u \geq e_M, \tag{2.43}$$

$${e_I}^T u = k, \tag{2.44}$$

$$p^T \delta \leq 1 - \alpha, \tag{2.45}$$

где

$$d = 1, \ldots, D, D = \prod_{j=1}^{J}\prod_{i=1}^{I} L_{ji},$$

$$p = (p_1, \ldots, p_D), p_d = P(\Theta = \theta^d) = \prod_{j=1}^{J}\prod_{i=1}^{I} P\left(\Theta_{ji} = \theta_{ji}^d\right),$$

$$\delta = (\delta_1, \ldots, \delta_D),$$

$$e_I \in \mathrm{R}^{\mathrm{I}},\ e_I = (1, \ldots, 1)^{\mathrm{T}},\ e_M \in R^M,\ e_M = (1, \ldots, 1)^T, \bar{\varphi} = (\varphi, \ldots, \varphi)^T,$$

$(\cdot)^T$ - транспонирование; θ^d, $d = 1, \ldots, D$ - реализация случайной матрицы Θ, $p \in R^D$ - вектор со значениями, равными вероятностям появления соответствующей реализации θ^d дискретной случайной матрицы Θ, $\delta \in \{0,1\}^D$ - вектор булевых переменных, с помощью которых организуется перебор α-доверительных множеств, L_{ji} - число возможных реализаций случайной величины Θ_{ji}.

Теорема 2. Задача (2.32) – (2.37) с функцией квантили (2.38) эквивалентна в смысле определения 1 задаче (2.39) – (2.45) смешанного целочисленного программирования.

Доказательство теоремы 2.

Случайный параметр в задаче (2.39) – (2.45) является случайной матрицей Θ с дискретным распределением и конечным числом реализаций. Воспользуемся теоремой об эквивалентности предложенной в [21]. Для этого запишем задачу

(2.32) – (2.37) в следующем эквивалентном виде задачи с вероятностным ограничением:

$$u^* = \operatorname*{Arg\,min}_{u\in\{0;1\}^I,\varphi\geq 0} (\gamma \frac{|c-w^T u|}{\varepsilon} + (1-\gamma)\frac{\varphi}{2700}),$$

$$c - w^T u \leq \varepsilon,$$

$$w^T u - c \leq \varepsilon,$$

$$A^T u \geq e_M,$$

$$e_I{}^T u = k,$$

$$\mathrm{P}\{\max_{j=\overline{1,J}} \Theta_j\, u \ \leq \varphi\} \geq \alpha\}$$

Проверим теперь выполнение условий теоремы [21]. В данном случае должны существовать функции:

$$-\infty < \mu_1(u,\varphi,\theta) \leq \inf_{\theta\in\{\theta^d, d=1,\dots,D\}} \Phi(u,\varphi,\theta), \tag{2.46}$$

$$-\infty < \mu_2(u,\varphi,\theta) \leq \max\left\{0, \inf_{\theta\in\{\theta^d, d=1,\dots,D\}} Q(u,\varphi,\theta)\right\}, \tag{2.47}$$

где:

$$\Phi(u,\varphi,\theta) = \frac{\gamma|c-w^T u|}{\varepsilon} + \frac{(1-\gamma)\varphi}{2700} \tag{2.48}$$

не зависит от θ, поэтому первое неравенство выполняется автоматически, если положить $\mu_1(u,\varphi,\theta) = \frac{\gamma|c-w^T u|}{\varepsilon} + \frac{(1-\gamma)\varphi}{2700}$.

Функция $Q(u,\varphi,\theta)$ задана следующим образом:

$$Q(u,\theta) = \max_{j=\overline{1,J}} \theta_j\, u - \varphi\ , \tag{2.49}$$

где θ_j - j-я строка реализации θ случайной матрицы Θ, поэтому

$$\mu_2(u,t) = \max\left\{0, \max_{j=\overline{1,J}} \theta_j^1\, u - \varphi\ \right\}, \tag{2.50}$$

где $\theta^1 = \left\|\theta_{ji}^1\right\|$, $\theta_{ji}^1 -$ минимальная реализация случайной величины Θ_{ji}, т.к.

$$\theta_{ji}^d \geq 0, d = 1,..,D, i = 1,\dots,I, u_i \in \{0,1\}.$$

Тогда из теоремы 1 [21] следует, что задача (2.32) – (2.37) эквивалента задаче

$$u^* = \operatorname*{Arg\,min}_{u\in\{0;1\}^I,\delta\in\{0;1\}^D,\varphi\geq 0} (\gamma \frac{|c-w^T u|}{\varepsilon} + (1-\gamma)\frac{\varphi}{2700}), \tag{2.51}$$

$$\theta^d u - \bar{\varphi} \le (\theta^d e_I)\delta_d, \tag{2.52}$$

$$c - w^T u \le \varepsilon, \tag{2.53}$$

$$w^T u - c \le \varepsilon, \tag{2.54}$$

$$A^T u \ge e_M, \tag{2.55}$$

$$e_I{}^T u = k, \tag{2.56}$$

$$p^T \delta \le 1 - \alpha, \tag{2.57}$$

где

$$d = 1, \dots, D, D = \prod_{j=1}^{J} \prod_{i=1}^{I} L_{ji},$$

$$p = (p_1, \dots, p_D), p_d = P(\Theta = \theta^d) = \prod_{j=1}^{J} \prod_{i=1}^{I} P\big(\Theta_{ji} = \theta_{ji}^d\big),$$

$$\delta = (\delta_1, \dots, \delta_D),$$

$$e_I \in \mathrm{R}^{\mathrm{I}},\ e_I = (1, \dots, 1)^{\mathrm{T}},\ e_M \in R^M,\ e_M = (1, \dots, 1)^T, \bar{\varphi} = (\varphi, \dots, \varphi)^T,$$

Ограничения (2.52) и (2.57) обеспечивают выполнение ограничения под знаком вероятности в (2.38) для реализаций случайной матрицы Θ: выбор оптимального вектора $\delta \in \{0, 1\}^D$ обеспечивает выбор оптимального набора реализаций θ^d с суммарной вероятностной мерой, превышающей α, то есть оптимального α-доверительного множества.

Поэтому, согласно результатам, полученным в [21], задача (2.32) – (2.38) эквивалентна в смысле определения 1 задаче (2.39) – (2.45) смешанного целочисленного программирования, что и требовалось доказать. ■

Сформулированная задача (2.39) - (2.45) требует перебрать 2^I комбинаций вектора u при достаточно большом количестве ограничений, зависящем от частоты дискретизации случайного времени, количества заданий и студентов. В рамках упрощения модели предлагается рассмотреть решение задачи для одного пользователя, которая также является актуальной. Это позволит значительно уменьшить размерность задачи за счёт меньшего числа ограничений.

2.4.2 Результаты вычислительного эксперимента

Рассмотрим решение задачи для одного пользователя с номером $j = 1$, при этом размерность вектора δ сократится до величины

$$D = \prod_{i=1}^{I} L_{1i}.$$

Для проведения анализа сформулируем задачу, исходные данные которой получены при обработке статистической информации системы дистанционного обучения в главе 1. Рассмотрим задания $M = 3$ различных типов, относящихся к основным изучаемым в течение первого семестра тематическим разделам курса по «Математическому анализу» СДО CLASS.NET МАИ НИУ [55], в каждом из которых по 10 различных заданий $I_M = 10,\ m = 1,\ \dots, 3$. Обозначим задания в зависимости от типа $m = 1,\ \dots, M$ и номера $i = 1,\ \dots, I_m$ как z_i^m. В статье [14] проведена оценка сложности каждого задания из общего числа при помощи алгоритма, основанного на модели Раша [123]. В результате получены приведенные к десятибалльной шкале оценки значений сложностей w_i^m для каждого задания w_i^m, которые представлены в табл. 2.9.

Таблица 2.9 Сложность w_i^m задания z_i^m

m / i	1	2	3
1	1,3113	4,132	2
2	3,2545	6,9023	2,4182
3	3,2545	2,1219	2,6666
4	3,2545	3,4366	3,6536
5	4,8738	2,4562	5,2425
6	5,3677	5,3587	5,5474
7	7,0109	6,9023	6,4528
8	7,2167	7,2833	7,1938
9	8,2444	7,8147	8,7945
10	9,6356	9,3991	3,6571

Исходя из оценок значений сложности для каждого задания, выберем требуемую суммарную сложность $c = 29{,}46$ теста из $k = 5$, возможный критерий выбора которой приведен в [42], и параметр отклонений от требуемой суммарной сложности ε, задаваемый числом, близким к нулю, для выбора наиболее оптимальных наборов тестов. Будем варьировать его от 0,0004 до 0,004 с шагом 0,0001. Оценки значений параметров логнормального распределения случайных величин $t_{ij} \sim LN(\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j; 0.31)$ получены в [42] и приведены в табл. 2.10.

Составлены ряды распределения времени ответа универсального пользователя на каждое задание системы. Задача решалась на уровне доверия $\alpha = 0{,}95$. Требовалось составить набор тестовых заданий из пяти задач, которые с вероятностью 0,95 могут быть решены пользователем за некоторое оптимальное время φ.

Таблица 2.10 Значения параметров логнормального распределения случайных величин

z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$	z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$	z_i^m	$\hat{\mu} + \hat{\beta}_i + \hat{\tau}_j$
z_1^1	3,51	z_1^2	4,72	z_1^3	3,65
z_2^1	3,92	z_2^2	5,87	z_2^3	3,73
z_3^1	3,89	z_3^2	3,83	z_3^3	3,87
z_4^1	3,91	z_4^2	3,91	z_4^3	3,96
z_5^1	4,22	z_5^2	3,87	z_5^3	4,84
z_6^1	4,63	z_6^2	5,13	z_6^3	4,95
z_7^1	5,67	z_7^2	5,25	z_7^3	5,53
z_8^1	5,71	z_8^2	5,71	z_8^3	5,89
z_9^1	6,13	z_9^2	5,94	z_9^3	6,18
z_{10}^1	6,39	z_{10}^2	6,27	z_{10}^3	3,88

Для решения задачи использованы средства вычислительной системы IBM CPLEX. Расчеты проводились на компьютере ASUS X550LC (Intel Corei5 2.3 GHz, 8 Gb RAM).

Для каждого ε получено число R наборов заданий, удовлетворяющих детерминированным ограничениям, а также при $\gamma = 0$ и $\gamma = 0{,}5$ для данных

наборов (табл. 2.11 - 2.12) получены оптимальные значения критерия ψ^*:

$$\psi^* = \frac{\gamma|c - w^T u^*|}{\varepsilon} + \frac{(1-\gamma)\varphi^*}{2700})$$

где u^*, φ^* - оптимальные значения u, φ решения задачи (2.39) – (2.45).

Очевидно, что для случая, когда минимизация отклонения сложности набора тестовых заданий и минимизация оптимального времени ответа имеют равный вес, приоритет все же отдается минимизации отклонения, так как с увеличением ε оптимальный набор заданий не изменяется.

Наибольшему числу наборов заданий соответствует $\varepsilon = 0{,}004$, для которого проведено исследование зависимости оптимального значения критериальной функции ψ^* от γ (рис. 2.2).

Таблица 2.11 Число наборов заданий при фиксированном ε и $\gamma = 0{,}5$

ε	Число решений, удовлетворяющих детерминированным ограничениям R	$\gamma = 0{,}5$	
		ψ^*	z_i^m
0,0007	3	0,5050	$z_5^1, z_8^1, z_3^2, z_7^3, z_9^3$
0,0009	4	0,4574	
0,0030	21	0,3408	
0,0040	35	0,3283	$z_1^5, z_1^8, z_2^3, z_3^7, z_9^3$

При $\gamma \geq 0{,}5$ оптимальный набор заданий не изменяется с увеличением ε и представляет собой набор с минимальным отклонением от заданной сложности c.

Для случая $0 \leq \gamma < 0{,}5$, т.е. когда минимизация наиболее приоритетна относительно времени выполнения, оптимальный набор тестовых заданий меняется несколько раз с увеличением ε. По этой причине отдельно рассмотрена задача для $\gamma = 0$.

Таблица 2.12 Число наборов заданий при фиксированном ε и $\gamma = 0$

ε	Число решений, удовлетворяющих детерминированным ограничениям K_Z	$\gamma = 0$	
		ψ^*	z_i^m
0,0007	3	0,5815	$z_5^1, z_8^1, z_3^2, z_7^3, z_9^3$
0,0009	4	0,5023	$z_6^1, z_1^2, z_7^2, z_9^2, z_5^3$
0,0030	21	0,4710	$z_6^1, z_8^1, z_1^2, z_6^3, z_8^3$
0,0040	35	0,4710	

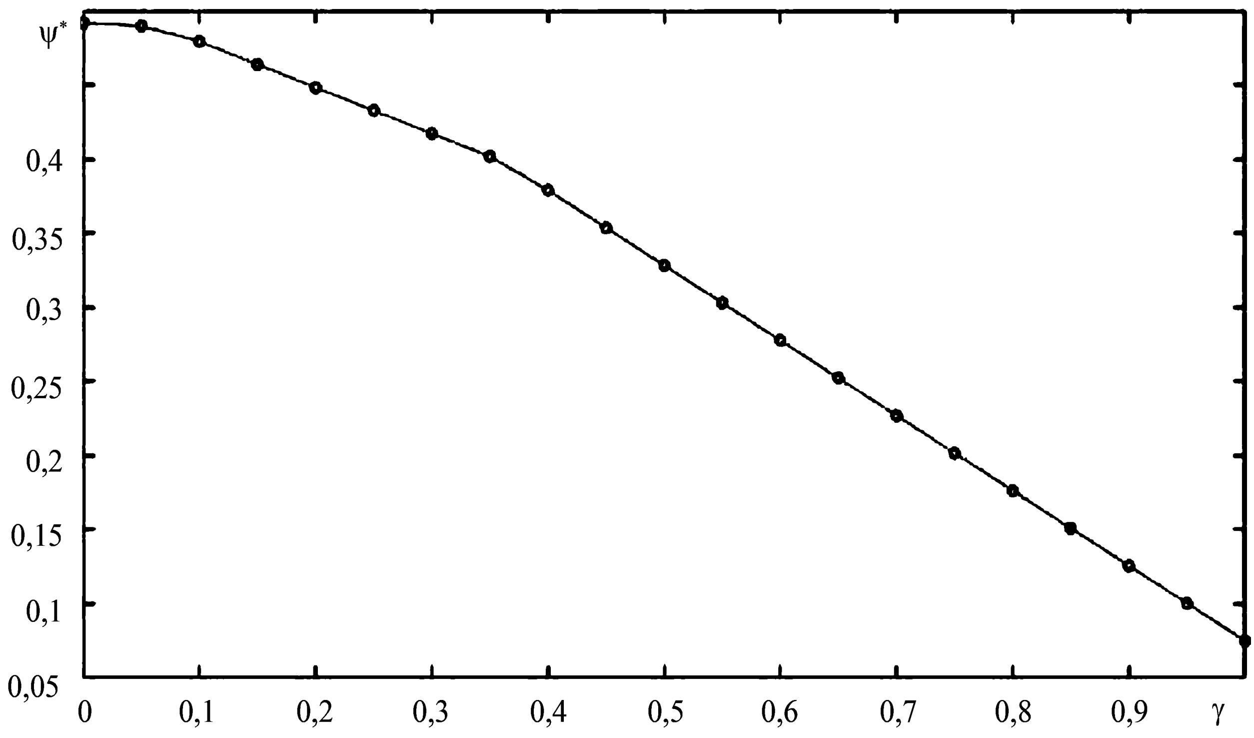

Рис. 2.2 Зависимость значения критериальной функции ψ^* от весового коэффициента γ

Результат вычислительного эксперимента для $\gamma = 0$ приведен в табл. 2.13.

В случае, когда приоритет полностью отдан нахождению минимального времени выполнения, для выбранных значений ε оптимальный тестовый набор меняется 5 раз и столько же раз, соответственно, меняется оптимальное время выполнения тестового набора (см. табл. 2.13). Время выполнения предложенного алгоритма приведено в табл. 2.14.

Таблица 2.13 Результат вычислительного эксперимента

ε	Оптимальный набор заданий Z^*	φ^*		ψ^*
		с	мин	
0,0004	$z_5^1, z_8^1, z_3^2, z_7^3, z_9^3$	1570,0833	26,1681	0,5815
0,0008	$z_5^1, z_8^2, z_7^3, z_8^3, z_{10}^3$	1377,6667	22,9611	0,5102
0,0010	$z_6^1, z_1^2, z_7^2, z_9^3, z_5^3$	1356,2000	22,6033	0,5023
0,0020	$z_6^1, z_7^1, z_4^2, z_7^3, z_8^3$	1328,0250	22,1337	0,4919
0,0030	$z_6^1, z_8^1, z_1^2, z_6^3, z_8^3$	1271,7750	21,1963	0,4710

Таблица 2.14 Время выполнения алгоритма

ε	Затраченное время, с	ε	Затраченное время, с
0,0004	4,189767	0,0009	7,3653710
0,0005	4,648206	0,0010	8,7168120
0,0006	4,521878	0,0020	23,158336
0,0007	4,178228	0,0030	34,178394
0,0008	5,498839	0,0040	39,744152

2.5 Задача формирования ограниченного по времени теста с гамма-распределением времени ответа пользователя на задания.

2.5.1 Постановка задачи

Основной проблемой использования логнормальной модели времени ответа пользователя на задания является сложность определения суммарного времени выполнения теста в ограничении (2.7), т.к. вычисление свёртки для получения совместного распределения является трудоёмкой задачей из-за отсутствия информации о суммарном распределении логнормальных случайных величин, и соответственно проблемой становится получение явного выражения для детерминированного эквивалента, поэтому в алгоритме 2.1 используется метод Монте-Карло для проверки ограничения (2.7). Далее решена задача формирования наборов тестов в случае универсального пользователя.

Для использования гамма-распределения сначала требуется решить задачу подбора распределения для каждого задания в тесте так, чтобы параметр распределения λ был общим, что позволит в явном виде получить распределение времени ответа пользователя на весь тест.

Алгоритм 2.2

Целью работы алгоритма является подбор параметров гамма-распределений времени ответа универсального пользователя для каждого задания так, чтобы параметр λ был бы общим для всех заданий, и при этом для максимального количества заданий при найденных оценках параметров распределений принималась бы гипотеза о гамма-распределении времени ответа пользователя на это задание.

Пусть $t_i,\ i = 1, \dots, I$ – время ответа универсального пользователя на задание i, где I – число заданий, из которых формируется тест; $t_{ji},\ i = 1, \dots, I, j = 1, \dots, I_i$ – реализация времени ответа пользователя j, затраченное им на решение задачи i, где I_i – число пользователей, решавших задачу i.

Опишем по шагам алгоритм подбора параметров гамма-распределений случайных величин $t_i,\ i = 1, \dots, I$.

0) Обнулим значения искомых параметров и некоторых счетчиков алгоритма. Положим $\lambda^* = 0$, где λ^* – искомое значение параметра гамма-распределения, который одинаков для всех задач; $\kappa_i^* = 0$, где κ_i^* – искомое значение второго параметра распределения для i -го задания; $S = 0$, где S – число задач, для которых принимается гипотеза о гамма-распределении времени ответа пользователя; $m = 0$, где m – счетчик. Выберем уровень доверительной вероятности $1 - \alpha$ для проверки статистических гипотез.

1) Для всех $i = 1, \dots, I$ по выборке объема I_i методом максимального правдоподобия находим оценки $\hat{\lambda}_i$ параметра λ. Среди полученных значений находим минимальное $\hat{\lambda}_{min}$ и максимальное $\hat{\lambda}_{max}$ значения. Для варьирования параметра λ выберем шаг

$$h = \frac{\hat{\lambda}_{max} - \hat{\lambda}_{min}}{L},$$

где L – выбранное заранее число шагов дискретизации по λ. Положим $\lambda_m = 0$.

2) Положим $m := m + 1$, а $\lambda_m = \lambda_{m-1} + h$. Для каждого $i = 1, \dots, I$ по выборке t_{ij}, $i = 1, \dots, I, j = 1, \dots, I_i$ определяем оценку второго параметра гамма-распределения

$$\widehat{\kappa_\iota} = \frac{\overline{t_{\iota J}}}{\lambda_m},$$

где $\overline{t_{\iota J}}$ – выборочное математическое ожидание.

3) Для всех $i = 1, \dots, I$ на выбранном уровне доверительной вероятности $1 - \alpha$ проверяем с помощью критерия Пирсона гипотезу H_0: $t_i \sim \Gamma(\widehat{\kappa_\iota^*}, \lambda_m)$. Если число принятых гипотез S' больше S, то полагаем $S = S'$, $\lambda^* = \lambda_m$, $\kappa_i^* = \widehat{\kappa_\iota}$ $i = 1, \dots, I$.

4) Если $m \le L - 1$, то перейти к шагу 2. Иначе завершить работу алгоритма.

Полученная модель распределения позволяет предложить эффективный алгоритм решения актуальной задачи формирования теста для универсального пользователя так, чтобы его сложность минимально отличалась от заданного экспертом уровня сложности и при этом минимизировалось время выполнения теста, которое гарантировано не будет превышено с заданным уровнем доверительной вероятности. Предполагается при этом, что сложности каждого задания оцениваются на основе обработки статистических данных о работе пользователей с помощью модели Раша.

В случае с применением гамма-распределения для случайного времени в ограничении (2.7) решение задачи упрощается за счёт свойства суммируемости независимых случайных гамма-распределенных величин с одинаковым параметром λ в распределении, т.е. если $t_1, \dots, t_I$ – независимые случайные величины, то:

$$\sum_{i}^{I} t_i \sim \Gamma\left(\sum_{i}^{I} \kappa_i, \lambda\right)$$

где $t_i \sim \Gamma(\kappa_i, \lambda), i = 1, \dots, I$.

2.5.2 Результаты вычислительного эксперимента

Для проведения анализа результатов решения сформулированной задачи воспользуемся исходными данными, полученными при обработке статистической информации о работе пользователей системы дистанционного обучения CLASS.NET МАИ (НИУ) [55].

Рассмотрим задания $M = 3$ различных типов, относящихся к основным, изучаемым в течение одного семестра, тематическим разделам курса по «Математическому анализу», в каждом из которых по 10 различных заданий $I^m = 10, m = 1,\dots,3$. Обозначим задания в зависимости от типа $m = 1,\dots,M$ и номера $i = 1,\dots,I^m$ как z_i^m.

С использованием методов, рассмотренных в [14], была получена оценка сложности каждого задания из общего количества при помощи алгоритма, основанного на модели Раша [123]. В результате были получены оценки значений сложностей w_i^m для каждого z_i^m, нормированные по десятибалльной шкале, которые представлены в таблице 2.14.

Исходя из оценок значений сложности для каждого задания, выберем требуемую суммарную сложность $c = 29.46$ теста из $l = 5$ заданий. Один из основных критериев подбора величины c был описан в [42]. Значение параметра ε задается числом, близким к нулю, для выбора наиболее оптимальных наборов тестов. Будем варьировать его от 0.0004 до 0.004 с шагом 0.0001.

Описанным в разделе 2.5.1 алгоритмом для каждого задания из выбранного пула были получены значения параметров гамма-распределения. Оценка параметра λ для всех задач одинакова и равна 2.29. Значения оценки второго параметра $\hat{\kappa}_i^m$ для каждого задания z_i^m приведены в таблице 2.15.

Таблица 2.14. Сложность w_i^m задания z_i^m

i \ m	1	2	3
1	1,3113	4,132	2
2	3,2545	6,9023	2,4182
3	3,2545	2,1219	2,6666
4	3,2545	3,4366	3,6536
5	4,8738	2,4562	5,2425
6	5,3677	5,3587	5,5474
7	7,0109	6,9023	6,4528
8	7,2167	7,2833	7,1938
9	8,2444	7,8147	8,7945
10	9,6356	9,3991	3,6571

На уровне доверительной вероятности 0.95 критерий Пирсона показал, что все гипотезы о том, что время ответа универсального пользователя на соответствующее задание системы имеет гамма-распределение с указанными параметрами, принимаются.

Далее решалась задача формирования теста, уровень доверительной вероятности также выбран равным 0.95. Требовалось составить наборы тестовых заданий из 5 задач, которые с вероятностью 0.95 могут быть решены пользователем за некоторое оптимальное время φ.

Для проверки адекватности модели времени ответа пользователя в случае гамма-распределения сформулированная выше задача была решена при тех же значениях ε и γ, что и в разделе 2.4, где в качестве модели времени ответа пользователя использовалась модель дискретного случайного времени. Сравним полученные результаты.

Таблица 2.15 Значения параметров гамма-распределения времени ответа универсального пользователя на задание z_i^m.

z_i^m	$\hat{\kappa}_i^m$	z_i^m	$\hat{\kappa}_i^m$	z_i^m	$\hat{\kappa}_i^m$
z_1^1	17.94	z_1^2	60.25	z_1^3	19.95
z_2^1	27.37	z_2^2	191.99	z_2^3	22.54
z_3^1	26.44	z_3^2	25.80	z_3^3	26.56
z_4^1	26.58	z_4^2	26.52	z_4^3	26.79
z_5^1	34.51	z_5^2	25.02	z_5^3	67.62
z_6^1	54.25	z_6^2	89.31	z_6^3	74.86
z_7^1	150.77	z_7^2	101.18	z_7^3	137.56
z_8^1	159.78	z_8^2	154.15	z_8^3	189.29
z_9^1	237.92	z_9^2	208.68	z_9^3	250.41
z_{10}^1	304.03	z_{10}^2	281.26	z_{10}^3	25.96

Для каждого ε было получено количество наборов заданий, удовлетворяющих детерминированным ограничениям, а также при $\gamma = 0$ и $\gamma = 0.5$ для этих наборов (таблицы 2.16 – 2.17) получены оптимальные значения критерия ψ^*.

Таблица 2.16 Решение задачи при $\gamma = 0.5$ и фиксированном ε

ε	Кол-во решений, удовлетворяющих детерминированному. ограничению	Оптимальное решение ψ^* при $\gamma =$ $= 0.5$	Оптимальный набор заданий при $\gamma = 0.5$
0.0007	3	0.4838	$z_6^1, z_{10}^1, z_5^2, z_6^3, z_7^3$
0.0009	4	0.4362	$z_6^1, z_{10}^1, z_5^2, z_6^3, z_7^3$
0.003	21	0.3195	$z_6^1, z_{10}^1, z_5^2, z_6^3, z_7^3$

Как видно из результата вычислений, для случая, когда минимизация отклонения сложности набора тестовых заданий и минимизация оптимального времени ответа имеют равный вес, приоритет всё же отдается минимизации отклонения, так как с увеличением ε оптимальный набор заданий не изменяется.

Таблица 2.17 Решение задачи при $\gamma = 0$ и фиксированном ε

ε	Кол-во решений, удовлетворяющих детерминированному. ограничению	Оптимальное решение ψ^* при $\gamma = 0$	Оптимальный набор заданий при $\gamma = 0$
0.0007	3	0.5322	$z_6^1, z_{10}^1, z_5^2, z_6^3, z_7^3$
0.0009	4	0.4816	$z_6^1, z_1^2, z_7^2, z_9^2, z_5^3$
0.003	21	0.4428	$z_6^1, z_8^1, z_1^2, z_6^3, z_8^3$

Наибольшему числу наборов заданий соответствует ε = 0.004. При $\gamma \geq 0.5$ оптимальный набор заданий не изменяется с увеличением ε и представляет собой набор с минимальным отклонением от заданной сложности c.

Полученные оптимальные значения критерия близки к значениям, полученным в разделе 2.2. Составы оптимальных тестовых наборов при некоторых ε полностью совпадают, при остальных ε отличаются на 1-2 задания.

2.6 Выводы по главе 2

1) Описана модель формирования тестов для систем компьютерного тестирования. Рассмотрена вероятностная постановка задачи определения набора приблизительно равных по суммарной сложности заданий с ограничением на время выполнения теста с непрерывным либо дискретным распределением случайного времени в виде задачи стохастического программирования. Решение вероятностной постановки задачи в общем виде является множеством индивидуальных наборов тестовых заданий.
2) Рассмотрена задача формирования ограниченного по времени теста в случае логнормального распределения времени для группы испытуемых. Так при

логнормальном распределении в вероятностном ограничении возникает проблема получения суммарного распределения, то нельзя получить выражение для детерминированного эквивалента. Так как число наборов конечное, то задачу можно решить полным перебором и проверкой вероятностного ограничения методом Монте-Карло. Для ускоренного поиска решения задачи предложен алгоритм поиска множества индивидуальных наборов тестовых заданий для группы студентов, основанный на идее метода ветвей и границ и позволяющий значительно сократить количество перебираемых полных наборов заданий. Также получен результат вычислительного эксперимента, показывающий преимущество предложенного алгоритма ускоренного поиска.

3) Рассмотрена задача формирования ограниченного по времени теста для пользователя дискретным распределением времени ответа пользователя на задания в вероятностном ограничении. Получена целочисленная задача эквивалентная задаче с логнормальным распределением в вероятностном ограничении для одного испытуемого. Доказана эквивалентность данных задач по набору оптимальных решений. Поиск оптимального решения с помощью средств вычислительного пакета IBM CPLEX потребовал значительно меньше времени, чем было затрачено на поиск решений в случае логнормальной модели времени. Это обусловлено тем, что поиск решения задачи осуществляется для одного пользователя, а не группы, а также не требуется использование метода Монте-Карло для проверки вероятностного ограничения.
4) Если в модели формирования теста с логнормальным временем распределения общее время на выполнение теста неизвестно, то для того, чтобы за некоторое оптимальное время все тестируемые могли выполнить выданный вариант теста с заданной вероятностью α, рассматривается функция квантили и формулируется задача квантильной оптимизации с критериальной функцией, содержащей нормированные безразмерные величины, которые позволяют приоритизировать подбор теста относительно отклонения от суммарной сложности теста или времени выполнения. Для

получения задачи квантильной оптимизации предлагается дискретизировать логнормальную модель времени, тем самым задачу квантильной оптимизации можно свести к детерминированной задаче частично целочисленного программирования. Сформулирована теорема 2 об эквивалентности этих задач. Проведён вычислительный эксперимент для подбора индивидуального набора заданий в случае одного пользователя для удобства вычислений. Проанализировано влияние параметров модели на результат решения задачи. Для случая, когда минимизация отклонения сложности набора тестовых заданий и минимизация оптимального времени ответа имеют равный вес, приоритет отдается минимизации отклонения, так как с увеличением ε оптимальный набор заданий не изменяется. В условиях проведения эксперимента наибольшему числу наборов заданий соответствует $\varepsilon = 0{,}004$, для которого проведено исследование зависимости оптимального значения критериальной функции ψ^* от γ. В случае, когда приоритет полностью отдан нахождению минимального времени выполнения, для выбранных значений ε оптимальный тестовый набор меняется 5 раз и столько же раз, соответственно, меняется оптимальное время выполнения тестового набора.

5) Рассмотрена задача формирования ограниченного по времени теста с гамма-распределением времени ответа пользователя на задания. Использование гамма-распределения позволяет получить детерминированный эквивалент для вероятностного ограничения и сформулировать задачу квантильной оптимизации. Для использования гамма-распределения решена задача поиска значений параметров распределения времени для каждого задания в тесте. Алгоритм позволяет получить параметры гамма-распределения с общим параметром λ для каждой задачи, что позволяет выполнить условие суммируемости гамма-распределенных величин в случае универсального пользователя. Полученная модель распределения позволяет предложить эффективный алгоритм решения актуальной задачи формирования теста для универсального пользователя так, чтобы его сложность минимально отличалась от заданного экспертом уровня сложности и при этом

минимизировалось время выполнения теста, которое гарантировано не будет превышено с заданным уровнем доверительной вероятности. В результате проведения вычислительного эксперимента получены оптимальные тестовые наборы для задачи в квантильной постановке с различными значениями параметров задачи.

3 Программный комплекс адаптации системы дистанционного обучения

3.1 Архитектура программного комплекса

Программный комплекс реализован как веб-приложение на языке программирования Python с использованием БД PostgreSQL и состоит из 2 программных модулей (сервисов):

1. моделирования времени: логнормального, гамма, дискретного;
2. адаптации: генерирования тестов, проверки компрометации заданий

Приложение имеет клиент-серверную архитектуру, т.е. в программном комплексе также доступен графический интерфейс, который можно интегрировать в системы компьютерного тестирования как виджет или через iframe внутри html-страницы (рис.3.1), либо можно взаимодействовать с сервисами при помощи https-запросов к Rest API программного комплекса напрямую (рис. 3.2).

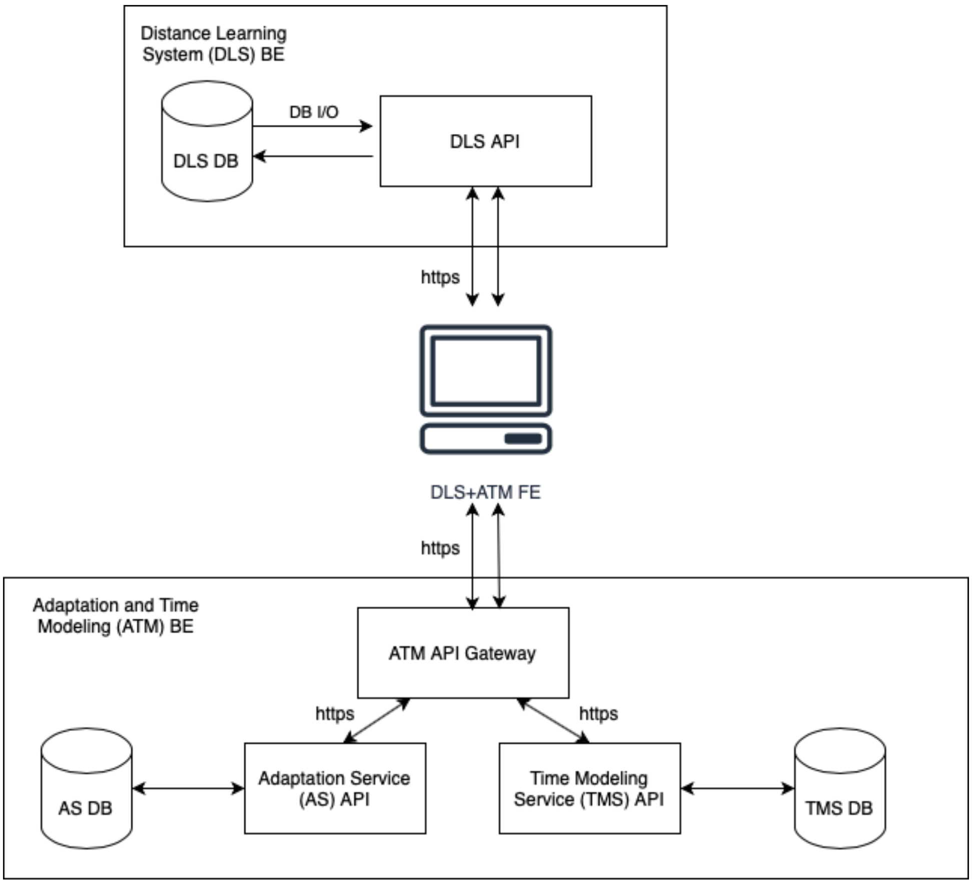

Рис. 3.1 Взаимодействие СДО и программного комплекса через общий клиентский интерфейс

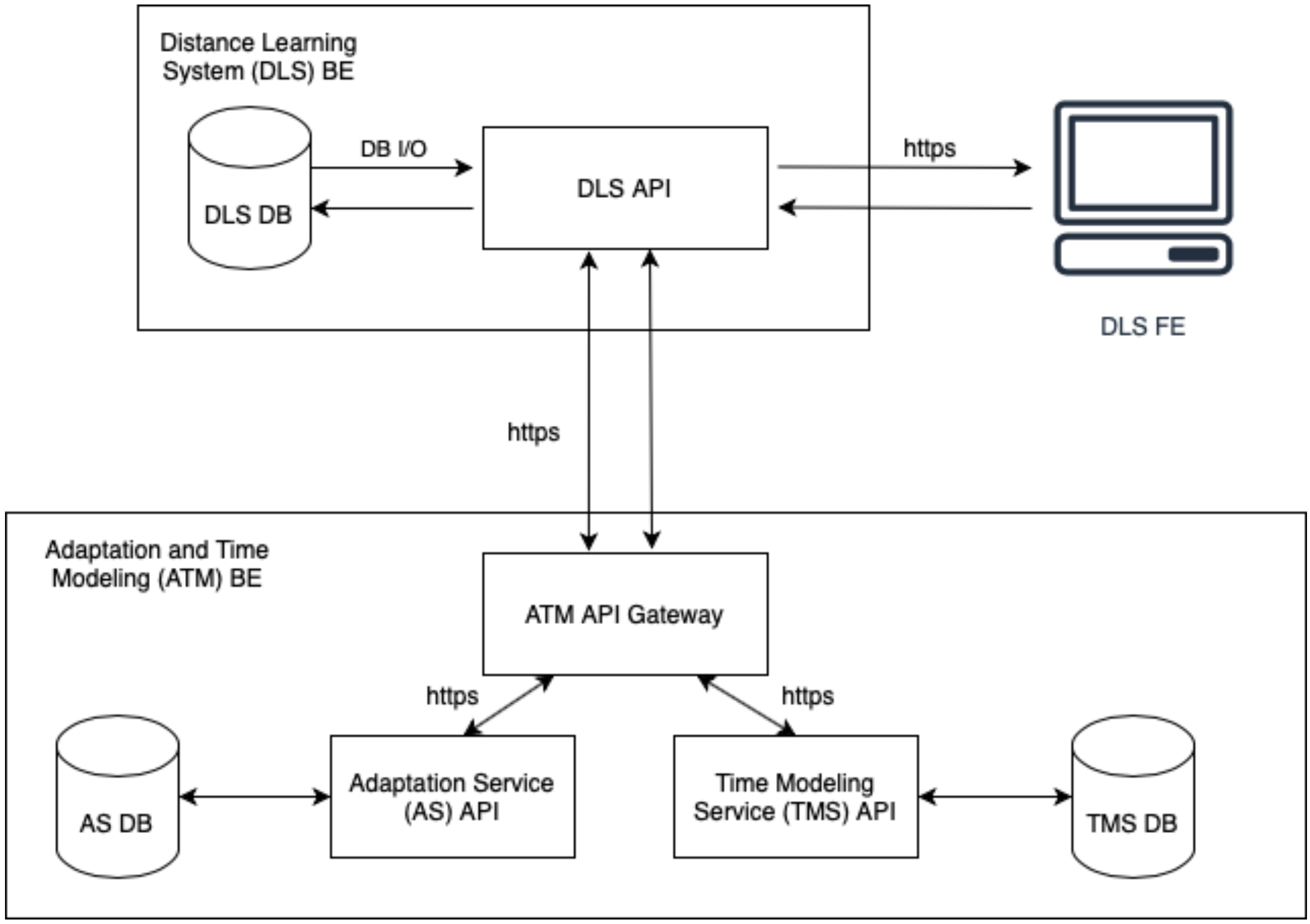

Рис. 3.2 Взаимодействие СДО и программного комплекса через серверный API

Выбор языка программирования Python обусловлен большим количеством библиотек, реализующих удобную работу с массивами данных и различными форматами их хранения. Также часть библиотек реализованы как предварительно скомпилированные библиотеки на C++ с Python API интерфейсом, что позволяет использовать уже оптимизированные методы с высокой скоростью работы.

Алгоритмы, используемые в каждом из модулей соответственно, для удобства при разработке сервисов вынесены в отдельные библиотеки. Данный подход позволяет использовать разработанные в данной диссертации алгоритмы как составную часть более сложных алгоритмов. Структура библиотек представлена в таблице 3.1.

Далее в главах 3.2 и 3.3 рассмотрены подробнее модули программного комплекса.

Таблица 3.1 Библиотеки и основные методы, используемые в модулях программного комплекса

Название библиотеки	Основные методы
dlsUserTimeModeling	dlsUserTimeModeling.lognormal dlsUserTimeModeling.gamma dlsUserTimeModeling.discrete
dlsTestGenerator	dlsTestGenerator.classicTest dlsTestGenerator.probabilityTimeLimitedTest dlsTestGenerator.quantileTimeLimitedTest

Для интеграции программного модуля должны выполняться следующие технические требования со стороны системы дистанционного обучения:

1. возможность осуществлять запросы к внешнему Rest API;
2. возможность фиксирования времени выполнения заданий для каждого студента;
3. возможность формирования csv или json, содержащего данные о времени выполнения.

Данные передаваемые через API должны быть анонимизированы для безопасности хранения и передачи данных испытуемых.

3.2 Модуль моделирования времени

Модуль моделирования времени отвечает за обработку данных и прогнозирование времени, затрачиваемом студентами на задания. Здесь реализованы модели, описанные в 1 главе, с логнормальным распределением для группы пользователей, дискретным и гамма-распределением для универсального.

Модуль получает на вход следующие данные:

1. список пользователей/список групп;
2. список заданий;
3. начало выполнения задания;
4. конец выполнения задания;
5. время выполнения задания.

После получения этих данных модуль проводит вычисления оценок параметров логнормального распределения при помощи оценок, описанных в 1 главе. В случае с дискретным распределением формируется гистограмма, и в явном виде вычисляется ряд распределения. Для гамма-распределения параметры вычисляются с помощью алгоритма 2.5.

Модуль моделирования времени после обработки данных формирует список пользователей, заданий и оценок параметров распределений.

В таблице 3.2 представлена основная таблица для хранения данных модуля моделирования времени.

Таблица 3.2 Таблица для хранения данных модуля моделирования времени

Атрибут	Тип	Ключ
id	integer	private key
user_id	integer	foreign key
group_id	integer	foreign key
task_id	integer	foreign key
task_start_date	timestamp	-
task_finish_date	timestamp	-
create_date	timestamp	-
update_date	timestamp	-
gamma_k	numeric	-
gamma_theta	numeric	-
ln_mu	numeric	-
ln_sigma	numeric	-
discrete	text	-

Рассмотрим подробнее атрибуты таблицы:

- user_id и group_id отвечают за идентификацию пользователя либо группы пользователей;
- task_id позволяет идентифицировать задание системы;

- task_start_date и task_finish_date позволяют рассчитать длительность выполнения задания определенным пользователем в системе дистанционного обучения;
- create_date определяет дату создания записи о задаче, которую решил определенный пользователь, в таблице;
- update_date определяет дату изменения данных в записи о задаче, которую решил определенный пользователь, в таблице, например, после пересчёта оценок для распределений случайного времени;
- gamma_k и gamma_theta являются значениями соответствующих параметров гамма-распределения;
- ln_mu и ln_sigma являются значениями соответствующих параметров логнормального распределения, полученных при оценке;
- discrete хранит значения времени и соответствующие вероятности для дискретного распределения.

Запрос к API модуля для создания записи о задании выглядит следующим образом:

```
POST /timeSimulation HTTP/1.1
{
  user_id: 123,
  task_id: 354,
  task_start_date: "2019-12-13T15:40:48.573Z",
  task_finish_date: "2019-12-13T15:56:34.849Z"
}
```

Примером ответа на запрос может быть следующее:

```
HTTP/1.1 201 CREATED
{
  id: 567,
  user_id: 123,
  task_id: 354,
  start_date: "2019-12-13T15:40:48.573Z",
  finish_date: "2019-12-13T15:56:34.849Z"
  create_date: "2019-12-13T15:57:00.377Z"
}
```

Сервис после получения данных вычисляет параметры распределений гамма (gamma_k, gamma_theta), логнормальное (ln_mu, ln_sigma), дискретное (discrete, пара «интервал – значение» в формате text/json). Гамма и логнормальное распределение для каждого пользователя уникальное, дискретное распределение конкретного задания одинаковое для всех пользователей и регулярно пересчитывается.

Для получения результатов вычислений система дистанционного обучения отправлением GET-запроса запрашивает параметры распределений и интервал времени, за который пользователь с вероятностью success_probability = 0,95 выполнит задание в результате моделирования для группы пользователей либо для пользователя, передавая user_id, task_id, distribution_type (gamma, ln, discrete). Также можно запросить данные для группы пользователей или для всех пользователей, отправив идентификатор учебной группы group_id, идентификатор задания task_id, либо не передавая параметры, тогда на выходе будет объект со всеми значениями для всех групп.

Пример запроса для получения результатов вычисления:

```
GET /timeDistribution?user_id=1&task_id=234&distribution_type=gamma HTTP/1.1
```

Ответ на такой запрос будет следующим:

```
HTTP/1.1 200 OK
Content-Type: application/json
{
  user_id: 1,
  task_id: 234,
  gamma_k: 3,
  gamma_theta: 4
}
```

Также модуль умеет работать с csv-файлами, т.е. внешняя система дистанционного обучения может отправить напрямую в API модуля файл с данными или пользователь системы может загрузить данные с помощью графического интерфейса.

3.3 Модуль адаптации

Модуль адаптации отвечает за изменение контента системы дистанционного обучения или компьютерного тестирования.

Основной задачей модуля является при изменении характеристик пользователей системы пересобрать задания и тесты.

При реализации модуля были учтены основные пользовательские сценарии взаимодействия с системой дистанционного обучения для преподавателя и студента (рис.3.3).

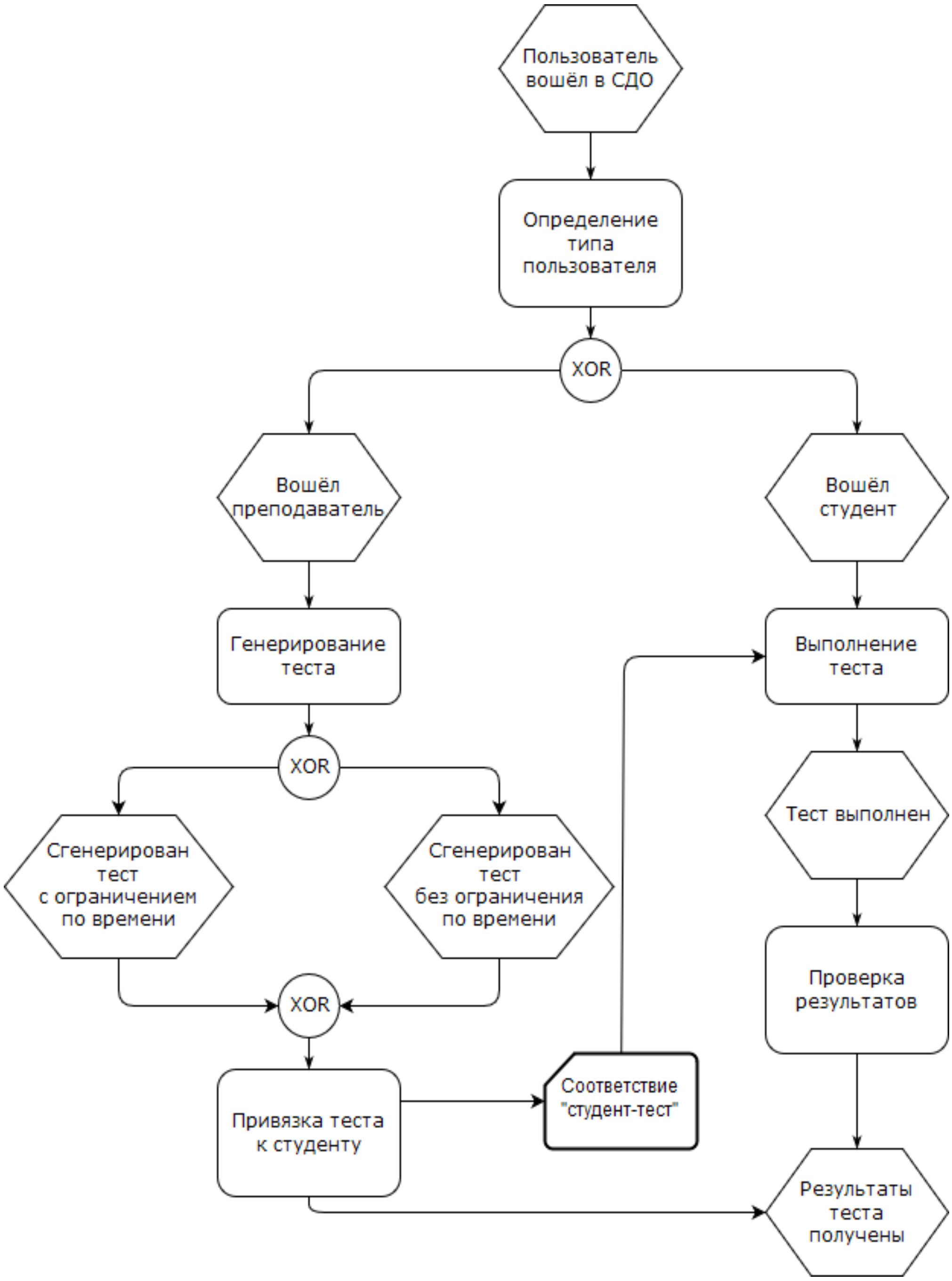

Рис.3.3 Пользовательские сценарии генерирования тестов в СДО

На вход модуль получает следующие данные от СДО:

1. Список групп пользователей или список пользователей
2. Тип теста (с ограничением или без ограничения по времени)
3. Пул заданий: разделы заданий или список заданий
4. Общее время на выполнение теста (при условии наличия ограничения)
5. Количество заданий в тесте
6. Сложность теста по выбранной нормированной шкале

В результате модуль формирует наборы тестов, в зависимости от входных параметров.

В таблице 3.3 представлена основная таблица для хранения данных модуля адаптации.

Таблица 3.3 Таблица для хранения данных модуля адаптации

Атрибут	Тип	Ключ
id	integer	private key
user_id	integer	foreign key
group_id	integer	foreign key
test_type	varchar	-
task_set	text	-
time_limit	numeric	-
task_number	int	-
difficulty	numeric	-
custom_test	text	-
create_date	timestamp	-

Рассмотрим подробнее атрибуты:

– user_id является идентификатором тестируемого пользователя, для которого формируется тестовый набор заданий;

– group_id является идентификатором тестируемой группы;
– test_type определяет наличие или отсутствие ограничения на время выполнения теста;
– task_set задаёт множество заданий, из которых будет формироваться тест;
– time_limit является ограничением на время выполнения теста;
– task_number определяет количество заданий в тесте;
– difficulty является параметром суммарной сложности теста;
– custom_test определяет полученный набор заданий в тесте;
– create_date является датой создания теста .

При взаимодействии с внешней системой дистанцинного обучения выполняется GET-запрос со следующими параметрами:

1. Группы тестируемых или тестируемые (group_id, user_id);
2. Тип теста с ограничением или без ограничения по времени (test_type, time_limited = true/false)
3. Пул заданий (task_set)
4. Общее время на выполнение теста при условии наличия ограничения (time_limit)
5. Количество заданий в тесте (task_number)
6. Сложность теста по выбранной нормированной шкале **(difficulty)**

Рассмотрим пример GET-запроса на получение тестового набора:

```
GET /customTest?user_id=1&test_time_limited=true&task_set=[1289,123455,378]&
        time_limit=45&task_number=5&difficulty=30 HTTP/1.1
```

Ответом на такой запрос является:

```
HTTP/1.1 200 OK
Content-Type: application/json
{
 user_id: 123,
 group_id: 56,
 custom_test: [task_id_1, task_id_2, task_id_3]
}
```

Модуль адаптации также, как и модуль моделирования времени, может обрабатывать csv-файлы с данными и формировать список тестовых наборов заданий в таком же формате.

Работоспособность модуля не зависит от состояния других модулей, внешних систем или графического интерфейса. В случае потребности в большом количестве вычислений допускается горизонтальное масштабирование, что позволит производить параллельные вычисления.

3.4 Графический интерфейс

Для простейшей интеграции в СДО был разработан графический интерфейс.

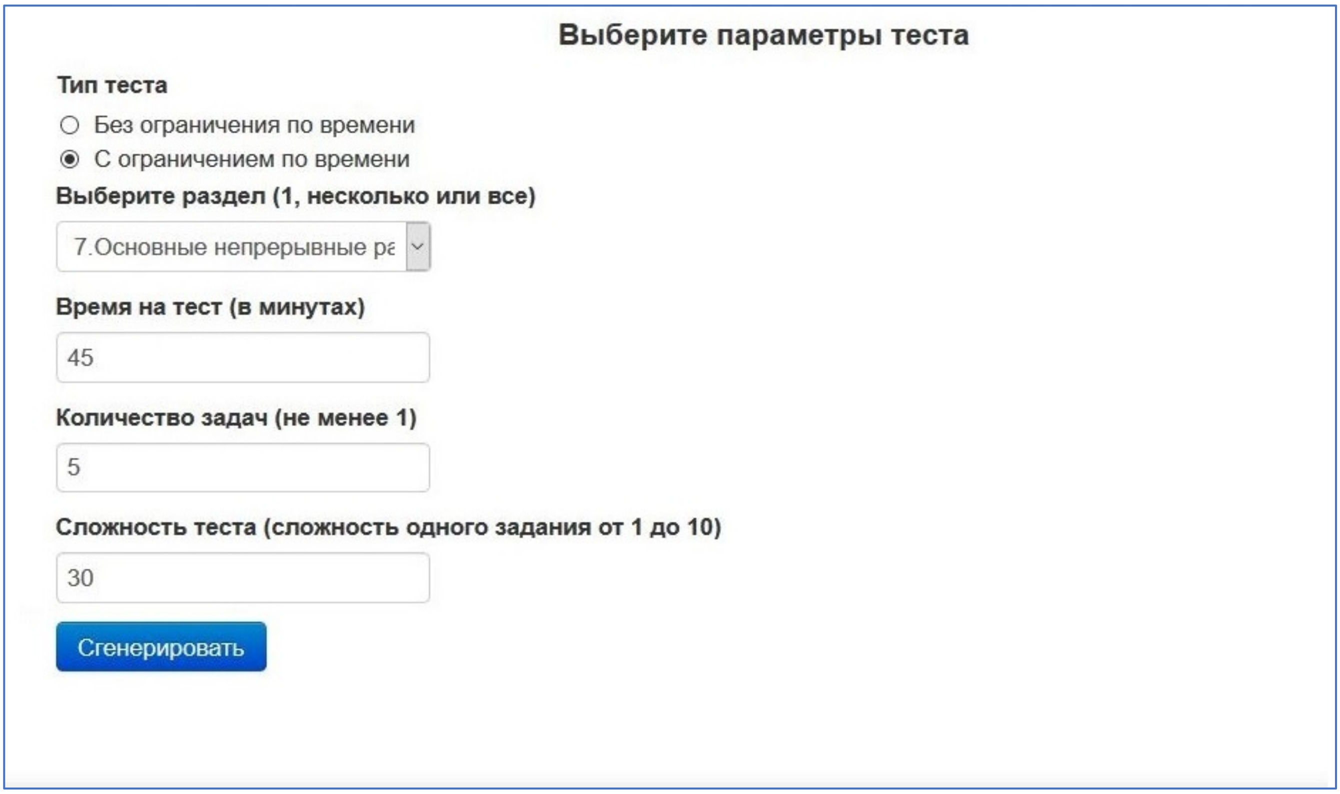

Рис.3.4 Интерфейс генератора тестов

Данный интерфейс позволяет пользователям сформировать тест для пользователей, указав параметры теста.

Интерфейс реализован стандартными средствами html и javascript. Для стилей элементов дизайна интерфейса был использован фреймворк Bootstrap.

Доступ к графическому интерфейсу может осуществляться прямым обращением по URL либо интеграцией интерфейса как виджета через iframe внутри какой-либо из страниц системы дистанционного обучения.

Графический интерфейс находится на уровне frontend и изолирован от backend сервисов и баз данных, взаимодействуя с серверным Rest API с помощью https-запросов, что позволяет освободить браузер клиента от сложной вычислительной логики, используемой в сервисах для моделирования времени и адаптации контента.

3.5 Выводы по главе 3

1) Разработан программный комплекс адаптации контента системы дистанционного обучения под изменяющийся контингент пользователей. Описана архитектура программного комплекса, состоящего из двух модулей, и возможности интеграции с системами дистанционного обучения через открытый Rest API, получающий анонимизированные исторические данные о пользователях и выполненных имя заданиях. Разработаны библиотеки, в которых собраны реализации алгоритмов, описанных в главах 1 и 2.
2) Модуль моделирования времени позволяет рассчитать оценки параметров логнормального, дискретного и гамма-распределения, получая на вход обезличенную информацию о пользователях и заданиях. Разработана архитектура модуля, позволяющая поддерживать его работоспособность независимо от других модулей, внешних систем и графического клиентского интерфейса. Общение с другими модулями и системами дистанционного обучения осуществляется посредством http-запросов к Rest API в формате json.
3) Модуль адаптации отвечает за изменение контента системы дистанционного обучения или компьютерного тестирования в зависимости от изменений контингента пользователей, т.е. при изменениях пользовательских данных формируются новые наборы тестовых заданий. Модуль адаптации учитывает сценарии работы с модулем различных пользователей: методиста (автора теста или преподавателя) и испытуемого. Входными данными являются значения параметров задачи формирования ограниченного по времени теста в формулировках, рассмотренных в главе 2. В зависимости от параметров

запроса к API модуля генерируются соответствующие наборы тестовых заданий и отправляются в систему компьютерного тестирования. Работоспособность модуля также не зависит от состояния других модулей, внешних систем или графического интерфейса. В случае потребности в большом количестве вычислений допускается горизонтальное масштабирование модуля, что позволит производить параллельные вычисления

4) Разработан графический интерфейс для взаимодействия с вычислительными модулями через веб-браузер. Интерфейс позволяет вручную через форму ввода данных выбрать параметры, по которым будут формироваться наборы тестовых заданий, или ввести нужные значения. Графический интерфейс можно встраивать в интерфейсы систем компьютерного тестирования либо использовать отдельно. Взаимодействие с вычислительными модулями происходит через Rest API.

Заключение

Основные результаты работы, выносимые на защиту.

1) Дискретная вероятностная модель описания времени ответа на задание системы дистанционного обучения или компьютерного тестирования. Модель учитывает индивидуальные особенности всех пользователей, характеристики которой интегрируются в единую модель (универсального пользователя), отражающую особенности всей тестируемой группы. Проведён эксперимент для проверки адекватности модели. Предложена модель гамма-распределения для описания времени ответа на задание для пользователя системы дистанционного обучения с алгоритмом подбора параметров распределения с учётом условия суммируемости гамма-распределенных случайных величин. Модель позволяет получить распределение для суммарного времени выполнения заданий в тестовом наборе для универсального пользователя, который характеризует группу испытуемых. [2, 38, 41, 42, 47, 64]
2) Вероятностные постановки задач формирования ограниченных по времени тестов для универсального пользователя и группы пользователей. Сформулированы задачи формирования тестов с логнормальным распределением времени ответа пользователя на задания для группы пользователей; с дискретным распределением для универсального пользователя; задачи квантильной оптимизации для случаев с дискретным распределением и гамма-распределением. Для задач квантильной оптимизации доказаны теоремы об эквивалентности по множеству допустимых решений. Получены результаты вычислительных экспериментов для каждой из постановок. [2, 34, 38, 41, 42, 47, 64]
3) Алгоритмы поиска множества допустимых решений и оптимального решения задач в рассмотренных постановках, а также численные методы, реализующие поиск. В случае задачи с логнормальным распределением времени для группы пользователей предложен алгоритм ускоренного поиска решений. Также разработан алгоритм, позволяющий получить

оценки параметров гамма-распределения случайного времени ответа. [2, 38, 42]

4) Комплекс программ, реализующих численные методы и вычисления по предложенным моделям. Разработана библиотека с вычислительными алгоритмами. [12, 16-19, 20, 31, 35-37, 39, 40, 45, 116]
5) Архитектура модулей моделирования времени и генерирования тестов в активно функционирующей среде компьютерного тестирования. Разработаны два вычислительных модуля и графический интерфейс для пользователей. [12, 16-19, 20, 31, 35-37, 39, 40, 45, 116]

Список литературы

1. Аванесов В.С. Математические модели педагогического измерения. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. — 26 с.
2. Босов А. В., Мхитарян Г.А., Наумов А.В., А. П. Сапунова Использование модели гамма-распределения в задаче формирования ограниченного по времени теста в системе дистанционного обучения // Информатика и ее применения», Т. 13, № 4, 2019, С. 11-17
3. Ван дер Варден Б.Л. Математическая статистика. – М.: Издательство иностранной литературы, 1960. 435 с.: ил.
4. Глущенко А. Ю. Разработка метода адаптивного управления обучением по индивидуальной образовательной траектории: дис. канд. техн. наук. ФГОУ ВПО «Государственный технологический университет «Московский институт стали и сплавов» Москва, 2009.
5. Горюшкин Е.И. Использование нейросетевых технологий в адаптивном тестировании по информатике в вузе: дис. канд. пед. наук, ГОУ ВПО «Курский государственный университет», Москва, 2009.
6. Гублер Е. В., Генкин А. А. Применение непараметрических критериев статистики в медико-биологических исследованиях. — Л., 1973.
7. Данг Х. Ф., Камаев В. А., Шабалина О. А. Метод разработки алгоритмов адаптивного тестирования // Известия ВолгГТУ. 2012. №13. URL: http://cyberleninka.ru/article/n/metod-razrabotki-algoritmov-adaptivnogo-testirovaniya (дата обращения: 11.10.2020).
8. Думин П.Н. Концепция системы поддержки принятия решений // Нейрокомпьютеры: разработка, применение. 2016. №9. С. 14-21.
9. Думин П.Н. Решение многокритериальной задачи выбора тестового задания в системе поддержки принятия решений. // Нейрокомпьютеры: разработка, применение. 2014. №10. С. 47 – 51.

10. Думин П.Н., Куравский Л.С. Анализ психологических аспектов игры на основе модифицированной функции Раша. // Нейрокомпьютеры: разработка, применение. 2016. №4. С. 43-47.
11. Думин П. Н., Куравский Л. С., Мармалюк П. А., Юрьев Г. А. Численные методы идентификации марковских процессов с дискретными состояниями и непрерывным временем // Матем. моделирование, 2017, Т.29, №5, с. 133–146
12. Жарков Е.А., Кибзун А.И., Мхитарян Г.А., Наумов А.В. Технологии создания и эксплуатации системы дистанционного обучения МАИ CLASS.NET // Функциональные пространства. Дифференциальные операторы. Проблемы математического образования тезисы докладов Пятой Международной конференции, посвящённой 95-летию со дня рождения члена корреспондента РАН, академика Европейской академии наук Л.Д. Кудрявцева. Российский университет дружбы народов. 2018. С. 449-450.
13. Кан Ю.С., Кибзун А.И. Задачи стохастического программирования с вероятностными критериями. М.: Физматлит, 2009.
14. Кибзун А. И., Иноземцев А. О., «Оценивание уровней сложности тестов на основе метода максимального правдоподобия» // Автоматика и телемеханика, 2014, №4, с. 20–37
15. Кибзун А. И., Панарин С. И. Стохастическая модель модифицируемости системы дистанционного обучения // Вестник Московского авиационного института, 2009, Т. 16, №7, с. 76-79.
16. Кибзун А.И., Мартюшова Я.Г., Мхитарян Г.А., Наумов А.В., Рыбалко А.А., Архитектура и технологии адаптации СДО МАИ как комплекса электронных учебников по математическим дисциплинам // Современные информационные технологии и ИТ-образование. 2016. Т. 12. № 3-2. С. 68-74. Kibzun A., Martiushova I., Mkhitaryan G., Naumov A., Rybalko A. System architecture and technologies of adaptation of LMS MAI CLASS.NET as set of electronic math textbooks // CEUR Workshop Proceedings Selected Papers of the 11th International Scientific-Practical Conference Modern Information Technologies and IT-Education, SITITO 2016. 2016. С. 164-171.

17. Кибзун А.И., Мхитарян Г.А., Наумов А.В. Проблемы дистанционного электронного обучения в вузах по математическим курсам., XVI Международный форум «Формирование современного информационного общества – проблемы, перспективы, инновационные подходы», Санкт-Петербург, 1-5 июня, 2015 г., с.69-70
18. Кибзун А.И., Мхитарян Г.А., Наумов А.В. Проблемы дистанционного электронного обучения в вузах по математическим курсам // Труды Международной научной конференции «Образование, наука и экономика в вузах и школах. Интеграция в международное образовательное пространство. Горис. 28 сентября-2 октября 2015», с. 454-457
19. Кибзун А.И., Мхитарян Г.А., Наумов А.В., Современные компьютерные технологии в разработке и эксплуатации системы дистанционного обучения CLASS.NET по математическим дисциплинам // Суперкомпьютерные технологии (СКТ-2016) Материалы 4-й Всероссийской научно-технической конференции. В 2-х томах. 2016. С. 251-255.
20. Кибзун А.И., Наумов А.В., Мхитарян Г.А. Особенности и технологии разработки системы дистанционного обучения математическим дисциплинам CLASS.NET // «Современные информационные технологии и ИТ образование» Том 1 (№ 11). – 2015, С. 153-156.
21. Кибзун А.И., Наумов А.В., Норкин В.И. О сведении задачи квантильной оптимизации с дискретным распределением к задаче смешанного целочисленного программирования // АиТ, 2013, № 6, С. 66-86.
22. Крамер Г. Математические методы статистики. – М.: Мир, 1976. – 648 с.
23. Куравский Л.С., Баранов С.Н., Юрьев Г.А. Синтез и идентификация скрытых марковских моделей для диагностики усталостного разрушения. - Нейрокомпьютеры: разработка и применение, №12, 2010, с. 20-36.
24. Куравский Л.С., Марголис А.А., Мармалюк П.А., Юрьев Г.А., Думин П.Н. Обучаемые марковские модели в задачах оптимизации порядка предъявления психологических тестов. – Нейрокомпьютеры: разработка и применение. 2013. №4. С. 28–38.

25. Куравский Л.С., Мармалюк П.А., Алхимов В.И., Юрьев Г.А. Новый подход к построению интеллектуальных и компетентностных тестов. // Моделирование и анализ данных. 2013. №1. С. 4–28.
26. Куравский Л.С., Мармалюк П.А., Юрьев Г.А., Думин П.Н. Идентификация марковских процессов по статистическим данным. Нейрокомпьютеры: разработка, применение. 2015. №5. С. 42 – 47.
27. Куравский Л.С., Мармалюк П.А., Юрьев Г.А., Думин П.Н., Панфилова А.С. Вероятностное моделирование процесса выполнения тестовых заданий на основе модифицированной функции Раша. Вопросы психологии. 2015. №4. С. 109-118.
28. Лазарсфельд П.Ф. Измерение в социологии // Американская социология. М.: Прогресс, 1972.
29. Ларин Р.М., Плясунов А.В., Пяткин А.В. Методы оптимиазации. Примеры и задачи: Учеб. Пособие, Новосибирский университет, Новосибирск, 2003, 115 с.
30. Мармалюк П.А., Юрьев Г.А., Куравский Л.С., Думин П.Н. Результаты вычислительного эксперимента по сравнению методов идентификации марковских процессов. // Нейрокомпьютеры: разработка, применение. 2015. №9. С. 44-50.
31. Мартюшова Я.Г., Мещеряков Е.А., Мхитарян Г.А. Организация автоматизированной рейтинговой формы контроля в электронных учебниках СДО МАИ CLASS.NET // Современные информационные технологии и ИТ-образование Сборник научных трудов II Международной научной конференции и XII Международной научно-практической конференции. Под редакцией В.А. Сухомлина. 2017. С. 133-138.
32. Мельникова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 410 с.
33. Михеев О.В. Математические модели педагогических измерений // Педагогические измерения, 2004, No2. -С.75-88.
34. Мхитарян Г.А. О задаче вероятностной оптимизации для формирования ограниченного по времени теста и добавлении нового функционала в СДО

МАИ CLASS.NET // 14-я Международная конференция "Авиация и космонавтика - 2015". Тезисы, Московский авиационный институт (национальный исследовательский университет). 2015. С. 435-437.

35. Мхитарян Г.А., Жарков Е.А. Архитектура современных адаптивных систем дистанционного обучения на примере СДО МАИ CLASS.NET // Системный анализ, управление и навигация. Тезисы докладов. 2019. С. 124-125.

36. Мхитарян Г.А., Мартюшова Я.Г., Кибзун А.И., Жарков Е.А. Междисциплинарные аспекты разработки и программной реализации электронных учебников для студентов технических университетов // Современные информационные технологии и ИТ-образование, Россия, 2018. URL: http://it-edu.oit.cmc.msu.ru/index.php/SITITO/sitito2018/paper/view/787 (дата обращения: 11.10.2020).

37. Мхитарян Г.А., Мартюшова Я.Г., Кибзун А.И., Жарков Е.А. Основные междисциплинарные аспекты разработки и программной реализации электронных учебников для студентов технических университетов. // Современные информационные технологии и ИТ-образование. 2019. Т. 15. № 2. С. 507-515

38. Мхитарян Г.А., Наумов А.В., Черыгова Е.Е. О задаче квантильной оптимизаци ограниченного по времени тестирования для одного пользователя // Материалы XII Международной конференции по прикладной математике и механике в аэрокосмической отрасли (NPNJ'2018) 2018. С. 736-737.

39. Мхитарян Г.А., Рыбалко А.А., Принципы реализации программного комплекса интеллектуальной поддержки и обеспечения безопасности функционирования СДО МАИ // Сборник НИРС МАИ - 2016 Московский авиационный институт (национальный исследовательский университет). Москва, 2017. С. 161-171.

40. Мхитарян Г.А., Рыбалко А.А., Принципы реализации программного комплекса интеллектуальной поддержки и обеспечения безопасности функционирования СДО МАИ CLASS.NET // Гагаринские чтения - 2016

Сборник тезисов докладов XLII Международной молодёжной научной конференции. 2016., Т.3, С. 70-71.

41. Мхитарян Г.А., Черыгова Е.Е. О задаче квантильной оптимизации ограниченного по времени тестирования //Гагаринские чтения - 2018 Сборник тезисов докладов XLIV Международной молодёжной научной конференции. 2018. С. 375-376.

42. Наумов А.В., Мхитарян Г.А. О задаче вероятностной оптимизации для ограниченного по времени тестирования // Автоматика и телемеханика. – 2016. – №9. – С. 124-135.

    Naumov A.V., Mkhitaryan G.A., On the problem of probabilistic optimization of time-limited testing // Automation and Remote Control. 2016. Т. 77. № 9. pp. 1612-1621.

43. Наумов А.В., Иванов С.В. Исследование задачи стохастического линейного программирования с квантильным критерием // АиТ, 2011, № 2, С. 142-158.

44. Наумов А.В., Иноземцев А.О. Алгоритм формирования индивидуальных заданий в системах дистанционного обучения // Вестник компьютерных и информационных технологий. – 2013. – Т. 1, №6. – С. 35-42.

45. Наумов А.В., Кибзун А.И., Мартюшова Я.Г., Осокин А.В., Жарков Е.А., Мхитарян Г.А., Использование СДО МАИ CLASS.NET при проведении очных занятий по математическим дисциплинам с адаптированием системы под уровень знаний пользователей // Информационные технологии в образовании "ИТО-Саратов-2016" Материалы VIII Международной научно-практической конференции. 2016. С. 268-276.

46. Наумов А.В., Мхитарян Г.А. О задаче вероятностной оптимизации для ограниченного по времени тестирования // Автоматика и телемеханика. 2016. № 9. С. 124-135.

47. Наумов А.В., Мхитарян Г.А., Черыгова Е.Е. Стохастическая постановка задачи формирования теста заданного уровня сложности с минимизацией квантили времени выполнения // Вестник компьютерных и информационных технологий. 2019. № 2 (176). С. 37-46.

48. Наумов, А.В., Джумурат А.С., Иноземцев А.О. Система дистанционного обучения математическим дисциплинам CLASS.NET // Вестник компьютерных и информационных технологий. – 2014. - Т. 1, №10, С. 36-40.
49. Николенко С., Тулупьев А., Сироткин А. Байесовские сети. Логико-вероятностный подход.М.: Наука. 2006. 600 с.
50. Пантелеев Методы оптимизации в примерах и задачах: Учеб. пособие/А. В. Пантелеев, Т. А. Летова. — 2-е изд., исправл. - М.: Высш. шк., 2005. – 544 с.: ил.
51. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983.
52. Равен Д. Педагогическое тестирование: проблемы, заблуждения, перспективы // Пер.с англ. М.: «Когито-Центр», 2001. 144 с.
53. Савченко Е.Ю. Применение модифицированных алгоритмов обучения нейронных сетей в задачах адаптивного тестирования. // Научный аспект. 2012. №4.
54. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2000. – 350 с., ил.
55. Система дистанционного обучения CLASS.NET МАИ (НИУ) [Электронный ресурс]. 2020. URL: http://distance.kaf804.ru/matan (дата обращения: 20.10.2020).
56. Сологуб Г. Б. Построение и использование байесовской сети для моделирования знаний студента в интеллектуальной системе тестирования. // Компьютерные инструменты в образовании. 2012. №2.
57. Строганов Д. В., Свободин В. Ю., Ягудаев Г. Г., Сычева Н. В. Оценка эффективности процедур адаптивного тестового контроля. // Наука и образование. 2012, №11, С.40-48
58. Трусов П.В. Введение в математическое моделирование. М.: Логос, 2007. – 440 с.
59. Тюменева Ю.А. Психологическое измерение. – М.: Аспект-Пресс, 2007.
60. Фоминова Н.С. Разработка и анализ стохастической и аппроксимационной моделей адаптивного тестирования для информационноуправляющих

систем: дис. канд. тех. наук. Московский государственный институт электронной техники, Москва, 2010.

61. Формалев В. Ф., Ревизников Д. Л. Численные методы. – М.: ФИЗ- МАТЛИТ, 2004. – 400 с.
62. Хлопотов М.В. Применение байесовской сети при построении моделей обучающихся для оценки уровня сформированности компетенций // Науковедение. 2014. №5 (24).
63. Челышкова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. – 432 с.
64. Черыгова Е.Е., Наумов А.В., Мхитарян Г.А Задача формирования теста заданного уровня сложности с минимальным временем выполнения. // Авиация и космонавтика - 2018 Тезисы 17-ой Международной конференции. 2018. С. 480-481.
65. Шан М. А. Автоматизированная информационная система адаптивного обучения на основе компетентностного подхода: дис. канд. техн. наук. Кабардино-Балкарский государственный университет им. Х. М. Бербекова, Нальчик, 2014.
66. Шмелёв А. Г., Столин В. В. Практикум по психодиагностике. Дифференциальная психометрика. М.: Изд-во Моск. ун-та, 1984. – 151 с.
67. Юрьев Г.А. Математическая модель интерпретации результатов компьютерного тестирования с использованием марковских сетей: дис. канд. ф.-м. наук, ГБОУ ВПО МГППУ, Москва, 2013.
68. Amponsah S.K., Darkwah K. F., Inusah A. Logistic preference function for preference ranking organization method for enrichment evaluation (PROMETHEE) decision analysis. // African Journal of Mathematics and Computer Science Research Vol. 5(6), pp. 112-119, 15 March, 2012.
69. Andrich, D. Relationships between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement , 2, 451-462, 1978
70. Andrich, D. An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), Sociological methodology, San Francisco: Jossey-Bass, pp. 33-80, 1985a.

71. Andrich, D. The application of an unfolding model of the PIRT type to the measurement of attitude. Applied Psychological Measurement, №12, pp33-51, 1988.
72. Baker F.B. The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
73. Batchelder, W. H., Colonius, H., and Dzhafarov, E. N. (Eds.) (In Press). New Handbook of Mathematical Psychology, Vol. 2. Modeling and Measurement. Cambridge, U.K.: Cambridge University Press. 2018
74. Bergan J.R. Rasch Versus Birnbaum: New Arguments in an Old Debate // Assessment Technology, Incorporated. [Электронный ресурс] URL: http://www.ati-online.com/pdfs/researchK12/RaschVsBirnbaum.pdf (дата обращения: 02.10.2020).
75. Billingsley P. Probability and Measure. Wiley series in probability and mathematical statistics, 1995, 519 p.
76. Binet AND., Simon T.H.E. The Development of Intelligence in Young Children. Vineland, NJ: The Training School, 1916.
77. Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability / In: F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Reading, Mass: Addison - Wesly, 1968. - 568p.
78. Boatemaa B. Multi Criteria Ranking of Telecommunication Carriers using Promethee Method. Thesis submitted to the department of mathematics, Kwame Nkrumah University of science and technology, 2012, 120 p.
79. Bollen K.A. Structural Equations with Latent Variables. P. 182., N - Y, Wiley & Sons, 1989. - 514p.
80. Burt, C.L. "The Differentiation of Intellectual Ability". The British Journal of Educational Psychology. 24 (2): 76–90, 1954.
81. Cattell R. B. Structured personality-learning theory: A Wholistic Multivariate research approach. Centennial Psychology Series. New York: Praeger, 1983, 466 pp.

82. Chong, H. Y. A Simple guide to the Item Response Theory (IRT) and Rasch modeling. Retrieved from March, 2013 URL: http:// www.creativewisdom.com (дата обращения: 20.10.2020)
83. Downey A. B. Think Python. — O'Reilly Media, 2012, 300 p.
84. Fedyanin D. On Parameter Identification Methods for Markov Models Applied to Social Networks.
85. Fox J.P. Bayesian Item Response Modeling: Theory and Applications. Springer Verlag, 2010.
86. Fulop J. Introduction to Decision Making Methods. 2014.
87. Gaviria, J.-L., Increase in precision when estimating parameters in computer assisted testing using response times // Quality & Quantity, 2005, № 39, cc.45–69
88. Goodman, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models. // Biometrika, No 61, 1974, 167- 171 pp.
89. Gregory R.J. Psychological testing: History, principles, and applications (5th edition). - New York: Pearson, 2007.
90. Guilford, J.P. Psychometric Methods. New York, NY: McGraw-Hill 1936
91. Gulliksen H., Theory of mental tests. New York: Wiley, 1950
92. Guttman, L. The structure of interrelations among intelligence tests. In Invitational Conference on Testing Problems. Princeton NJ: Educational Testing Service, 1964, pp. 25-36
93. Guttman, L. Measurement as structural theory. Psychometrika 36, pp. 329–347, 1971
94. Jensen A. R. The g Factor: The Science of Mental Ability, Westport, CT: Greenwood, 1998, p.700.
95. Kelley, T. L. Interpretation of Educational Measurements. Yonkers, N.Y.: World Book, 1927. pp. 363
96. Kingsbury, G.G., Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Ed.), New horizons in testing: Latent trait theory and computerized adaptive testing (pp. 237-254). New York: Academic Press.

97. Kuder, G.F. Use of the International scoring machine for the rapid computation of tables of intercorrelations. Journal of Applied Psychology, 22, 587-596, 1938
98. Kuravsky L. S., Malykh S. B. Application of Markov models for analysis of development of psychological characteristics. - Australian Journal of Educational & Developmental Psychology, 2004, Vol 2, pp 29-40.
99. Kuravsky L. S., P. A. Marmalyuk, G. A. Yuryev, P. N. Dumin. A numerical technique for the identification of discrete-state continuous-time Markov models. Applied Mathematical Sciences, Vol. 9, 2015, no. 8, 379-391.
100. Kuravsky L.S. and Baranov S.N. Condition monitoring of the structures suffered acoustic fatigue failure and forecasting their service life. - Proc. Condition Monitoring 2003, Oxford, United Kingdom, pp. 256-279, July 2003.
101. Kuravsky L.S. and Baranov S.N. Neural networks in fatigue damage recognition: diagnostics and statistical analysis. - Proc. 11th International Congress on Sound and Vibration, St.-Petersburg, Russia, pp. 2929-2944, July 2004.
102. Kuravsky L.S. and Baranov S.N. The concept of multifactor Markov networks and its application to forecasting and diagnostics of technical systems. – In: Proc. Condition Monitoring 2005, Cambridge, United Kingdom, pp. 111-117, July 2005.
103. Kuravsky L.S., Baranov S.N. and Yuryev G.A. Synthesis and identification of hidden Markov models based on a novel statistical technique in condition monitoring. – In: Proc. 7th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Stratford-upon-Avon, England, June 2010. – 23 pp.
104. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N. A Probabilistic Model of Adaptive Training. // Applied Mathematical Sciences. 2016. Vol. 10, 2016, no. 48, 2369 – 2380.
105. Kuravsky L.S., Marmalyuk P.A., Baranov S.N., Alkhimov V.I., Yuryev G.A. and Artyukhina S.V. A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications. – Applied Mathematical Sciences. Vol. 9. 2015. No 21. Pp. 1003–1026.

106. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Dumin P.N., Panfilova A.S. Probabilistic modeling of a testing procedure. // Applied Mathematical Sciences. 2015. T. 9. No 81-84. C. 4053-4066.

107. Lawley D.N. On Problems Connected with Item Selection and Test Construction // Proceedings of the Royal Society of Edinburgh. Section A Mathematical and Physical Sciences. 43 v. LXI, part III, p. 273 - 287, 1943.

108. Lazarsfeld P.F. The interpretation and computation of some latent structures. // Measurement and Prediction, N-Y, John Wiley and Sons, 1950.- p.415-472, 1950.

109. Linacre J. M., Wright B. D. BIGSTEPS: Rasch-model computer programs. Chicago, IL: Mesa Press. 1998.

110. Lord F.M. A Theory of Test Scores. Psychometric Monographs, 7. Richmond. Publ. by Psychometric Society. 1952, 84 pp.

111. Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading MA: Addison-Welsley Publishing Company. 1968.

112. M.D. Zeigenfuse, William H. Batchelder, Mark Steyvers, An item response theory model of matching test performance, Journal of Mathematical Psychology, Volume 95, 2020

113. Matloff N. Art of R Programming - A Tour of Statistical Software Design, 2011. Mills C. N., Stocking M. L. Practical issues in large-scale computerized adaptive testing. Applied Measurement in Education, 9(4), 1996.

114. Mills C. N., Stocking M. L. Practical issues in large-scale computerized adaptive testing. Applied Measurement in Education, 9(4), 1996.

115. Milovanovic I. Python Data Visualization Cookbook. – Packt Publishing, 2015, 302 p.

116. Mkhitaryan G.A., Kibzun A.I., Martyushova Ya.G., Zharkov E.A. Interdisciplinary aspects of development and software implementation of electronic textbooks for students of technical universities Modern Information Technologies and IT-Education. SITITO'2018. Communications in Computer and Information Science, vol. 1201. Springer, Cham (2020), pp. 110-120 (Scopus, WoS)

117. Naumov A.V., Mkhitaryan G.A., Rybalko A.A., Software set of intellectual support and security of LMS MAI CLASS.NET // Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование. 2016. Т. 9. № 4. С. 129-140.

118. Novick, M.R. The axioms and principal results of classical test theory Journal of Mathematical Psychology Volume 3, Issue 1, February 1966, pp.1-18, 1966.

119. Pearl J. Probabilistic Reasoning in Expert Systems: Networks of Plausible Inference. San Francisco: Morgan Kaufmann, 1988.

120. Piton-Gonçalves, J. & Aluísio, S. M. An architecture for multidimensional computer adaptive test with educational purposes. ACM, New York, NY, USA, pp.17-24, 2012

121. Piton-Gonçalves, J.; Aluísio, S. M. Teste Adaptativo Computadorizado Multidimensional com propósitos educacionais: princípios e métodos. Ensaio: Avaliação e Políticas Públicas Em Educação. 23 (87): pp.389–414, 2015.

122. Podvezko V., Podviezko A. Dependence of multi-criteria evaluation result on choice of preference functions and their parameters. // Baltic Journal on Sustainability, No. 16(1), 2011, 143–158 pp.

123. Rasch, G. Probabilistic models for some intelligence and attainment tests. // Copenhagen, Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press. 1960/1980.

124. Richardson M. W. The Relation Between the Difficulty and the Difference Validity of a Test / Psychometrika, 1936, 1: 2, pp.33-49.

125. Richardson M.W. Multimensional psychophysics. Psychological Bulletin. Vol.35: pp.659–660. 1938

126. Richardson M.W. Notes on the Rationale of Item Analysis./Psychometrika, №1: pp.169-76, 1936

127. Rob R. Meijer, Leonardo S. Sotaridona, Detection of Advance Item Knowledge Using Response Times in Computer Adaptive Testing // Law School Admission Council Computerized Testing Report, 2006

128. Rossant C. IPython Interactive Computing and Visualization Cookbook. — Packt Publishing, 2014, 520 p.

129. Snijders Tom A.V. Statistical Models for Social Networks // Annual Review of Sociology Vol. 37, 131-153, 2011.

130. Spearman, C. The proof and measurement of association between two things. By C. Spearman, 1904. The American Journal of Psychology. 100 (3–4): 441–471, 1987.

131. Thurstone L. L. Ability, motivation, and speed. // Psychometrika, 1937, № 2, cc.249–254

132. Tomić V., Marinković Z., Janošević D. PROMETHEE method implementation with multi-criteria decisions. // Mechanical Engineering, Vol. 9, No 2, 2011, pp. 193 – 202 pp.

133. van der Linden, W. J, Conceptual Issues in Response-Time Modeling

134. van der Linden, W. J, Predictive Control of Speededness in Adaptive Testing// Law School Admission Council Computerized Testing Report, 2007

135. van der Linden, W. J, Some New Developments in Adaptive Testing Technology // Journal of Psychology, 2008, Vol. 216(1), pp. 3-11

136. van der Linden, W. J, Using Response Times for Item Selection in Adaptive Testing // Journal of Educational and Behavioral Statistics, 2008, Vol. 33. No.7, pp. 5-20

137. van der Linden, W. J.; Veldkamp, B. P. Constraining item exposure in computerized adaptive testing with shadow tests. Journal of Educational and Behavioral Statistics. 29 (3): 273–291, 2004.

138. Vernon P., Beales, A. The Measurement of Abilities. British Journal of Educational Studies. 6(1): 92, 1957,

139. Wainer, H. (Ed.). Computerized adaptive testing: A Primer (2nd Edition). Mahwah, NJ: ELawrence Erlbaum Associates, 2000

140. Wainer, H.; Mislevy, R.J. Wainer, H. (ed.). Item response theory, calibration, and estimation. Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

141. Wang T., Hanson B.A., Development and Calibration of an Item Response Model that Incorporates Response Time// American Educational Research Association 2001

142. Wechsler D., The Measurement and Appraisal of Adult Intelligence. Fourth Edition. Baltimore: William & Wilkins, 1958. pp. 297

143. Weiss, D. J.; Kingsbury, G. G. Application of computerized adaptive testing to educational problems. Journal of Educational Measurement. 1984, №21 (4), pp.361–375

144. Weiss, D.J. Adaptive Measurement of Individual Change. Zeitschrift für Psychologie, 2008, №216 (1), pp.49-58

145. Weiss, D.J. New horizons in testing: Latent trait theory and computerized adaptive testing. New York: Academic Press, pp. 237–254, 1983

146. Wickham H. Ggplot2 - Elegant Graphics for Data Analysis. 2009.

147. Wise S. L., Lingling Ma Setting Response Time Thresholds for a CAT Item Pool: The Normative Threshold Method // National Council on Measurement in Education, Vancouver, Canada, 2012

148. Wise, S. L., Kong, X., Response time effort: A new measure of examine emotivation in computer-based tests // Applied Measurement in Education, 2005, № 18, cc. 163-183

149. Woodbury M. A., On the standard length of a test // Psychometrika, 1951, № 16, cc. 103–106

150. Woodbury M. A., The stochastic model of mental test theory and an application, 1963, № 28, pp. 391–393

151. Wright B.D., Masters G.N. Rating scale analysis. Rasch measurements. – Chicago: MESA Press, 1982.

152. Wright B.D., Stone M.N. Best Test Design. - Chicago: MESA Press, 1979.

153. Yule G.U. Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series (with discussion). In: Journal of the Royal Statistical Society, Nr. 89, 1926, pp. 1-69