

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)
Институт прикладной математики и компьютерных наук

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК
Руководитель ООП
д-р техн. наук, профессор

_____ С.П. Сущенко

«_____» _____ 2023 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА
РАЗРАБОТКА ВЫЧИСЛИТЕЛЬНО-ИМИТАЦИОННОГО КОМПЛЕКСА
МОДЕЛИ УЗЛА ОБРАБОТКИ ЗАПРОСОВ РАЗНОГО ТИПА

Благинин Алексей Леонидович

Направление подготовки 01.04.02 Прикладная математика и информатика
Направленность (профиль) «Интеллектуальный анализ больших данных»

Руководитель ВКР
канд. физ.-мат. наук

_____ И.Л. Лапатын

«_____» _____ 2023 г.

Автор работы
студент группы № 932128

_____ А.Л. Благинин

«_____» _____ 2023 г.

ОГЛАВЛЕНИЕ

| | |
|--|----|
| Введение | 3 |
| 1 Применение имитационного моделирования в теории массового обслуживания | 5 |
| 2 Разработка программного комплекса | 10 |
| 2.1 Предметная область | 12 |
| 2.2 Объектная модель предметной области | 15 |
| 2.3 Реализации интерфейсов | 20 |
| 2.4 Особенности реализации | 22 |
| 2.5 Реализация программы в качестве программного пакета для Python..... | 23 |
| 2.6 Интеграция в процесс исследования | 27 |
| 3 Вычисление асимптотических результатов | 30 |
| 3.1 Связь характеристических функций с сигналами и их спектрами | 31 |
| 3.2 Дискретизация характеристической функции | 32 |
| 3.3 Иллюстрация работы подхода на биномиальном распределении | 33 |
| 3.4 Иллюстрация работы подхода на задаче теории массового обслуживания | 34 |
| 3.5 Вычисление коэффициента вариации ММРР | 35 |
| 4 Работа с имитационной моделью | 37 |
| 4.1 Определение области применимости асимптотических результатов исследо- вания модели RQ–системы с повторными вызовами и обратной связью | 37 |
| Заключение | 43 |
| Список использованной литературы | 45 |

ВВЕДЕНИЕ

Данная работа посвящена оптимизации процесса экспериментального исследования моделей систем массового обслуживания.

[1] и ММРР [2, 3, 4].

[5]

Тезисы введения:

1. Используя наработки по ранее разработанной имитационной модели, оптимизировать ее для масштабного проведения численных экспериментов.
2. Реализовать метод вычисления асимптотического распределения вероятностей, используя дискретное преобразование Фурье.
3. Реализовать алгоритм вычисления вариации для интервалов между моментами поступления заявок ММРР.
4. Разработать и реализовать вспомогательные программные инструменты для проведения численных экспериментов.
5. Провести апробацию машинного обучения для вычисления вариации длин интервалов между моментами поступления заявок из выходящего процесса рассматриваемой модели системы.

—OLD—

Данная работа посвящена оптимизации процесса экспериментального исследования моделей систем массового обслуживания различного рода [1] и ММРР [2, 3, 4]. Численные методы позволяют в полной мере исследовать рассматриваемую модель, а именно ответить на вопросы, касающиеся определения границ области применимости и наблюдения функционирования системы при нестандартных параметрах, требуется проведение большого числа численных экспериментов и наличие специализированных инструментов для их анализа. По этой причине было принято решение реализовать программный комплекс с набором инструментов для анализа систем теории массового обслуживания, который будет включать в себя алгоритмы вычисления характеристик, имитационную модель и средства для её эффективного использования.

Решение разработать собственное ПО для моделирования было принято ввиду ряда аспектов исследовательской работы, которые требуют создания и

анализа большего объема данных. Для этого требуется автоматизировать существенную часть процесса исследования ввиду ограниченного времени и ресурсов. Помимо этого, трудоемкость на последующих этапах работы значительно снижается благодаря структурированности получаемой информации.

В рамках цели данной работы поставлены следующие задачи:

1. Выработать функциональные требования к программному комплексу;
2. Разработать предметную область имитационной модели и принцип ее функционирования;
3. Реализовать модель с учетом разработанной предметной области и требований.

Работа содержит 48 страниц, 17 рисунков, 2 таблицы, 60 источников.

В первом разделе описывается модель рассматриваемой системы массового обслуживания. Во втором разделе рассматривается разработка и применение имитационной модели. В третьем разделе описаны новые методы вычисления характеристик работы системы. В четвертом разделе рассматривается применение методов машинного обучения для анализа работы рассматриваемой модели системы.

—OLD—

1 Применение имитационного моделирования в теории массового обслуживания

Имитационное моделирование [5] является мощным инструментом для экспериментального изучения систем массового обслуживания. Его актуальность как метода исследования обусловлена следующими факторами:

1. принцип, лежащий в основе моделирования, довольно прост для понимания, что делает этот инструмент доступным для широкого круга исследователей, в том числе для тех, кто не имеет глубоких познаний в теории случайных процессов;
2. при использовании имитационного моделирования в связке со сложными средствами визуализации можно детально наблюдать процесс работы системы в различных ее аспектах, что дает данные, которые невозможно получить при помощи исключительно численного подхода к анализу, более сфокусированного на различных показателях функционирования системы;
3. численные методы в большинстве случаев опираются на решение специально составленных систем уравнений (система уравнений Колмогорова). В свою очередь, имитационное моделирование предлагает методологию, которая помогает рассчитать те же показатели без изменения основного подхода [6];
4. имитационное моделирование зачастую является способом подтвердить аналитические результаты, при исследовании системы массового обслуживания методом асимптотического анализа и определить их область применимости. Поскольку в качестве аналитического решения, как правило, выступает аппроксимация некой функции, то для установки ее качества требуется применение численных методов исследования, среди которых имитационное моделирование предоставляет широкий набор возможностей.

Система массового обслуживания описывает при помощи специального математического аппарата [7] порядок обработки заявок — абстрактных объектов, которые в реальном мире могут выступать в качестве телефонных звон-

ков, сетевых запросов или клиентов, требующих обслуживания. Для описания работы системы используется ряд основных элементов, которые проводят операции над заявками:

1. входящий поток — является случайным потоком событий (заявок), возникающих в системе и требующих обработки;
2. обслуживающий прибор — второй важный элемент системы массового обслуживания, представляющий собой случайных процесс обработки заявок. В системе может быть несколько приборов, при том их блок может также иметь различную топологию: они могут обслуживать заявки последовательно, либо параллельно. Также на конфигурацию системы влияет и выбор самого количества обслуживающих приборов;
3. Для описания класса систем массового обслуживания, в которых заявок может не суметь получить доступ к прибору, используется источник повторных вызовов, иначе орбита, которая агрегирует заявки, которые не смогли захватить прибор и через некоторое случайное время попытаются повторить попытку захвата
4. Для регулирования поступления заявок на прибор используется очередь, иначе буфер, которая накапливает входящие заявки прежде чем они попадут на прибор.

Все перечисленные элементы являются случайными процессами и могут иметь различные характеристики и распределение случайных величин. Задачей теории массового обслуживания является выбор наиболее эффективной структуры системы для ее наилучшей производительности. Данная задача решается путем нахождения и анализа характеристик ее функционирования. Среди аналитических методов решения этой задачи можно выделить метод асимптотического анализа [8] и метод диффузионного анализа [9]. Имитационное моделирование, в свою очередь, является как методом апробации полученных аналитических результатов при помощи численных экспериментов, так и подходом к исследованию систем массового обслуживания, заключающимся в имитации работы исследуемой модели системы массового обслуживания посредством компьютерных вычислений.

Как упомянуто выше, ключевое преимущество имитационного моделирования — прямолинейная реализация алгоритма, что экономит время и снижает трудоемкость исследования. Из относительно простой реализации алгоритмов вытекает и легкое конфигурирование и изменение структуры системы с возможностью быстрого получения результатов в последствии.

При моделировании систем массового обслуживания используется дискретно-событийный подход. Он заключается в генерации случайных величин, имеющих требуемое распределение, которые являются моментами наступления событий в системе. События могут быть различного рода: окончание обслуживания заявки, перемещение заявки на прибор, поломка прибора и т. д. В момент наступления события состояние системы меняется, что и отслеживается в процессе моделирования.

Рассмотрим подход подробнее на примере следующей системы:

на вход системы поступает пуассоновский поток заявок. Заявка занимает прибор в случае, если он свободен, тот, в свою очередь, начинает ее обслуживание в течение экспоненциально распределенного времени с параметром μ . В ситуации, когда прибор уже занят обслуживанием, входящая заявка, после неудавшейся попытки захватить прибор, мгновенно уходит на орбиту и осуществляет там случайную задержку в течение экспоненциально распределенного времени с параметром σ .

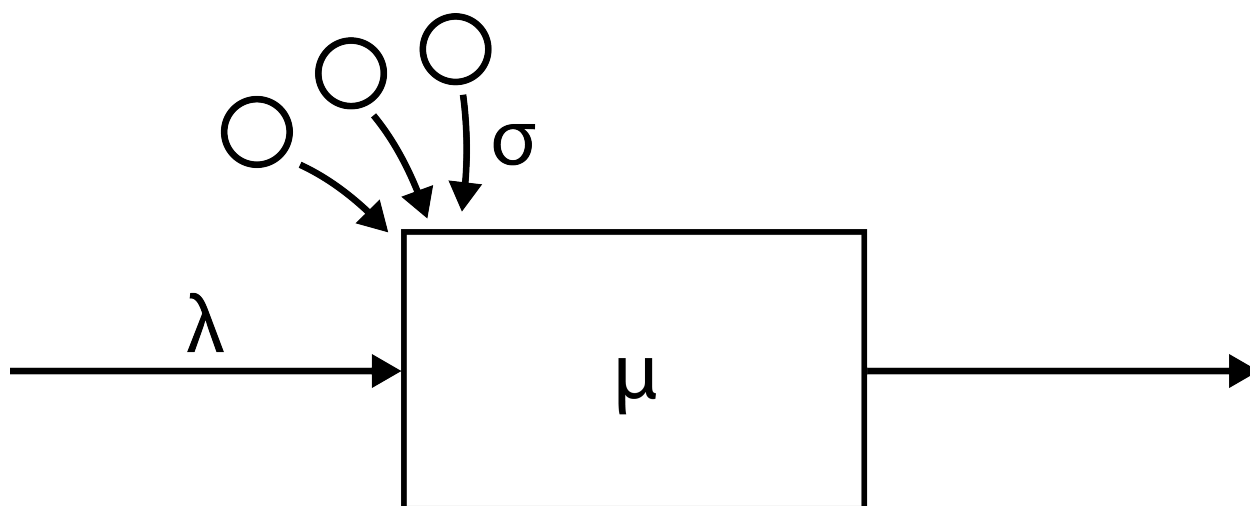


Рисунок 1 — Пример системы ТМО

Алгоритм моделирования функционирования представленной системы массового обслуживания при помощи очереди событий представлен ниже. Вве-

дем следующие обозначения:

1. Q — очередь, хранящая моменты наступления предстоящих событий в системе и функционирующая по принципу кучи, когда извлекаемый элемент является минимальным среди содержащихся в ней;
2. T_{curr} — текущее время моделирования;
3. T_{end} — момент окончания моделирования;
4. t_i — момент наступления события i в системе.



Рисунок 2 — Алгоритм моделирования

Так, при каждой итерации моделирования из Q достается момент наступления следующего события t_i , в свою очередь текущее время моделирования устанавливается как $T_{curr} = t_i$. Наступление события i позволит получить вре-

мя наступления последующего события t_j (например, заявка успешно поступила на прибор из входящего потока в момент t_i , и t_j будет являться момент окончания ее обслуживания), которое поместить в Q и итерация алгоритма завершится, и процесс будет повторяться, пока Q опустеет, либо $T_{curr} = T_{end}$.

Однако для получения достоверных результатов при помощи имитационного моделирования требуется достаточный объем выборки данных, получаемый при работе алгоритма [10, 11]. Определение требуемого объема выборки является отдельной подготовительной задачей и может решаться эмпирически путем подбора такого количества итераций алгоритма, при котором погрешность моделирования, выражаемая, к примеру, среднеквадратическим отклонением [12], становится несущественной.

Так как моделирование подразумевает сравнение получаемой выборки, в частности законов ее распределения, для проверки некой гипотезы о соответствии исследуемой математической модели, также требуется набор критериев и метрик для этого. В общем случае такой метрикой может выступать расстояние Колмогорова, являющимся максимальной по модулю разницей между распределениями вероятностей:

$$\Delta = \max_{0 \leq i < \infty} \left| \sum_{v=0}^i (P_0(v) - P_1(v)) \right|. \quad (1)$$

Данная формула также применима для случаев с многомерными распределениями вероятностей [13].

2 Разработка программного комплекса

В данном разделе описывается процесс проектирования и разработки комплекса алгоритмов и программ для численного исследования процессов, протекающих в системах массового обслуживания, а также расчетов характеристик их функционирования. Ключевой составляющей данного комплекса является имитационная модель. Также программный комплекс содержит утилиты для вычисления дискретного преобразования Фурье, расчета расстояния Колмогорова, а также параметризованное решение системы уравнений Колмогорова для вычисления асимптотического приближения характеристической функции числа обслуженных заявок в системе массового обслуживания с повторными вызовами и обратной связью [14].

Проектирование программы начинается с разработки предметной области по предварительно составленным требованиям. В данном случае, делается упор на универсальность модели и возможность встраивания в непосредственный процесс анализа. Также особое внимание уделяется процедуре проведения численных экспериментов. Численные результаты являются наиболее репрезентативными в случае, если было проведено достаточное количество экспериментов при наборах параметров, которые описывают функционирование модели в различных ситуациях. Такой подход позволяет удостовериться в верности теоретической стороны исследования посредством анализа объемной выборки данных о работе модели.

Предметная область программы была разработана с учетом следующих требований к ее работе и процессу взаимодействия с пользователем. Требования были составлены исходя из имеющегося опыта анализа систем массового обслуживания при помощи имитационного моделирования и опираясь на реализации имеющихся инструментов [15]:

1. Производительность. Инструмент должен обладать достаточным быстродействием, чтобы в течение адекватного времени проводить имитационное моделирование. Данное требование также обосновано необходимостью получать объемные выборки для получения достоверных результатов моделирования;
2. Возможность моделирования широкого спектра моделей теории массового обслуживания, иначе — универсальность;
3. Инструмент должен иметь возможность проведения масштабных экспериментов. Другими словами, программа должна быть оптимизирована для запуска большого количества моделей одновременно и параллельно;
4. Способ задания параметров должен позволять выполнять эксперименты повторно;
5. Возможность конфигурирования модели. Это означает, что инструмент должен позволять настраивать различные аспекты имитационного моделирования для каждого случая в отдельности;

6. Возможность производить сбор данных о работе модели в требуемом контексте. Иначе говоря, инструмент должен позволять собрать конкретные статистические данные имитационного моделирования в зависимости от поставленной задачи;
7. Интеграция в процесс анализа. Другими словами, имитационное моделирование должно быть одним из инструментов среды, в которой проводится анализ. Из данного требования вытекает и выбор целевой программной среды для разработки;
8. Легкая расширяемость. Исходный код программы должен предоставлять возможность добавления новой логики.

2.1 Предметная область

Исходя из выше представленных требований, инструмент должен иметь возможность производить моделирование различных систем массового обслуживания. Исходя из этого, наиболее гибким вариантом реализации данного требования будет включить построение модели в процесс работы с инструментом. Так, у пользователя появляется возможность конструировать требуемую под задачу систему массового обслуживания.

Для достижения предложенного подхода предлагается использовать конструирование блоками [16] — инструмент будет включать в себя набор атомарных элементов систем массового обслуживания (обслуживающий прибор, входящий поток и т.д.), которые будут работать по принципу черного ящика, однако связь между элементами и промежуточные этапы будут доступны для конфигурирования. Такой подход позволяет создать абсолютно любую вариацию системы массового обслуживания и используется не только в имитационном моделировании. Зачастую подобное архитектурное решение можно наблюдать в программных комплексах, предназначенных для программирования сложных взаимодействий между объектами. В 3D-графике зачастую используется схожий процесс работы, основанный на графах, где вершины представляют собой алгоритмы и объекты, а ребрами являются взаимодействия между ними. Например, в ПО для создания эффектов и симуляции взаимодействия физических тел Houdini используется именно такой подход [17].

Принцип работы имитационной модели базируется на обработке событий, происходящих в системе во время ее функционирования. Под событиями понимаются перемещения заявок по элементам системы. Обработка включает в себя регистрацию времени наступления события. Заявками же являются объекты, которые обрабатываются в ходе работы системы. Цикл обработки заявки, как правило, начинается с поступления ее в систему из входящего потока, далее поступление на прибор, обслуживание и последующее покидание системы. Именно их особенностей прохождения заявок через систему складываются характеристики ее функционирования

Исходя из вышеперечисленного, для разработки инструмента выбран объектно–ориентированный подход. Его преимуществом является возможность построения абстракций над логикой работы конкретных элементов системы, что позволяет удовлетворить требованию легкой расширяемости и позволить конструирование модели блоками в процессе анализа.

Объектно–ориентированный подход [18] к реализации предметной области подразумевает выделение из нее отдельных сущностей, которые могут быть представлены одним самостоятельным объектом, выполняющим конкретную задачу.

Процесс моделирования заключается в перемещении заявок от одного элемента к другому. В очередной условный момент времени T_{cur} местонахождение каждой заявки и ее характеристики могут меняться. По этой причине для каждой заявки вводится условный момент времени моделирования, когда она изменит своё состояние — T_{shift} . В момент времени, когда $T_{shift} = T_{curr}$, заявка перемещается в очередной элемент модели, и T_{shift} обновляется им в зависимости от заданных параметров. Иначе процесс моделирования можно представить как последовательное изменение состояния системы, складывающегося из текущих состояний ее элементов. В свою очередь события, происходящее в системе могут быть поделены на несколько основных категорий:

1. прибытие заявки в систему;
2. начало обслуживания заявки на приборе;
3. окончание обслуживания заявки на приборе;
4. перемещение заявки в источник повторных вызовов (орбиту, при ее нали-

чий в системе);

5. попытка повторного обращения из источника повторных вызовов.

В первую очередь, как ранее было указано, выделим элемент модели — объект, который содержит некую вычислительную логику и обрабатывает заявки в соответствии с этой логикой. Пример элемента модели является обслуживающий прибор.

Для того, чтобы соединить элементы в единую модель, потребуется определенным образом указать, как заявки будут перемещаться между элементами. Для этой цели выделим сущность — маршрутизатор. Однако, в зависимости от реализации элемента, он может иметь разное количество связей, с помощью которых будут перемещаться заявки. Таким образом, требуется ввести сущность, которая регулирует количество и качество возможных соединений элемента — слот.

Заявка также представлена сущностью, которая хранит в себе время смены ее состояния на новое, текущее состояние и другие опциональные атрибуты.

Поскольку элемент модели работает по принципу черного ящика, то не следует отягощать его логику сбором данных о моделировании. Для этой цели можно использовать маршрутизатор, отслеживая характеристики перемещения заявок в нужных участках модели. Исходя из этого, логику сбора статистики можно вынести в отдельную сущность, что позволит добавить в данный процесс разнообразную логику.

Итак, в предметной области программы были выделены следующие сущности:

- Элемент системы — объект, являющийся частью модели и выполняющий обработку заявок на протяжении ее работы.
- Заявка — объект, перемещающийся по системе и подлежащий обработке.
- Маршрутизатор — объект, образующий односторонний туннель для перемещения заявок от элемента системы к другому.
- Слот — объект, служащий для соединения двух элементов при помощи маршрутизатора.

- Сборщик статистики — объект, собирающий данные о перемещении заявок по маршрутизатору;
- Генератор задержки — служит для генерации случайных величин с требуемым распределением.

2.2 Объектная модель предметной области

На основе выделенных сущностей и описанного принципа работы программы была построена ее объектная модель, в последствии использованная для реализации. Для графического представления предметной области используется язык моделирования UML [19], в частности диаграмма классов [20], позволяющая графически изобразить выделенные в предметной области в качестве классов и взаимосвязь между ними.

Для обеспечения универсальности и возможности добавления новых разновидностей элементов модели был введен общий для них интерфейс *Producer*. Такое решение, в качестве побочного эффекта, также может позволять использовать различные элементы системы в несвойственных для них позициях в модели.

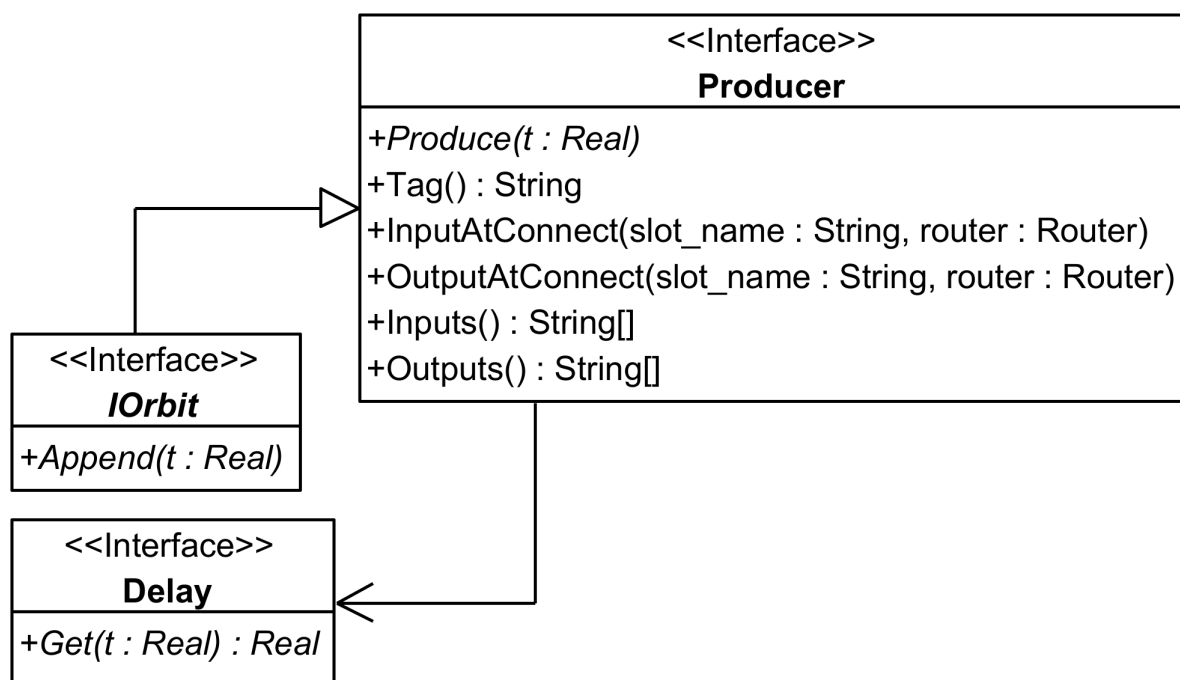


Рисунок 3 — Интерфейс *Producer*

В абстрактном методе *Produce* должна находиться логика работы элемен-

та системы, именно этот метод будет вызываться в каждой итерации работы модели, возвращая вектор моментов времени наступления событий в элементе. Также в интерфейсе определены вспомогательные методы Tag, InputAtConnect, OutputAtConnect, Inputs и Outputs, служащие для соединения элементов между собой. Метод Tag возвращает константную строку, определенную для каждой реализации интерфейса Producer. Слоты элемента разделены на два класса — входящие, которые позволяют принимать элементу заявки, и выходящие, позволяющие ему их отправлять. Методы InputAtConnect и OutputAtConnect позволяют присоединять к маршрутизатору входящие и выходящие слоты соответственно, указывая название слота для соединения.

Другим важным элементом предметной области является объект, генерирующий время задержки заявок в процессе их перемещения и позволяющий проводить эксперименты, когда элементы системы могут задавать задержку для заявки разного рода. Для этой цели служит интерфейс Delay. Он имеет единственный метод Get, возвращающий сгенерированную случайную величину.

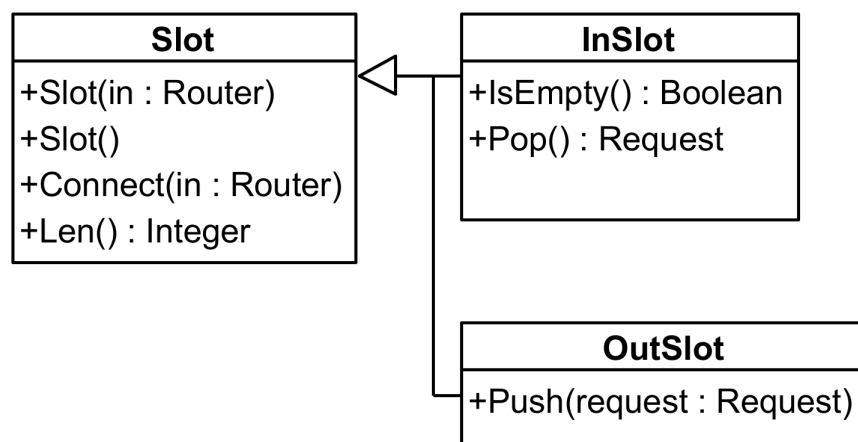


Рисунок 4 — Класс Slot и его наследники

Слоты используются в элементах системы для задания ограничения на возможные связи между другими элементами. К примеру, входящий поток может иметь только один выход для заявок, которые им генерируются, и не имеет входов, поскольку порождает заявки самостоятельно. В свою очередь, орбита должна иметь один вход для поступающих на нее заявок и один выход для отправления их обратно. Для разграничения логики работы слота на принимающий и отправляющий введено два наследника — InSlot, способный только при-

нимать заявки при помощи метода Pop и OutSlot, предназначенный лишь для их отправления при помощи метода Push. Среда для переправки заявок между слотами обеспечивается за счет класса Router.

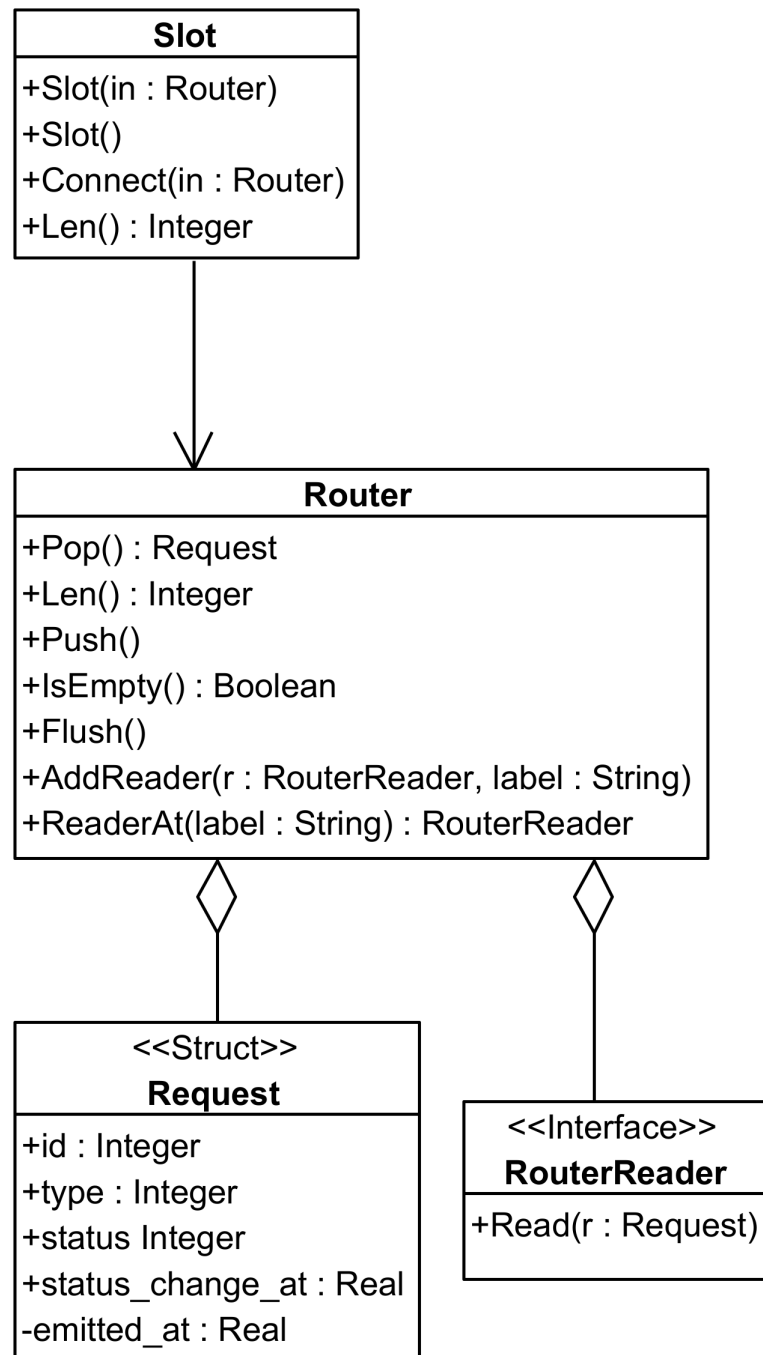


Рисунок 5 — Класс Router

Маршрутизатор представляет собой очередь, работающую по принципу FIFO. Таким образом, он позволяет элементам системы взаимодействовать с заявками между итерациями, когда очередная заявка еще ожидает обработки, но

текущая итерация подошла к концу. Управление очередью заявок осуществляется при помощи методов Push (поместить заявку в очередь), Pop (достать заявок из очереди), Len (получить текущую длину очереди), IsEmpty (пуста ли очередь). Методы AddReader и ReaderAt предназначены для управления наборов сборщиков статистики. AddReader позволяет добавить новый сборщик статистики под указанным ярлыком, который позволит обращаться к нему в последствии при помощи метода ReaderAt для доступа к собранным данным.

Заявка представлена структурой Request, содержащей момент изменения ее состояния $status_change_at$ (T_{shift}), тип, служащий для разделения способа происхождения заявки, статус, отражающий этап ее обработки (обслужена, в процессе обслуживания, в пути, покинула систему) и время, когда заявка была создана.

Ниже представлена полная объектная модель предметной области, объединяющая ранее представленные классы и интерфейсы.

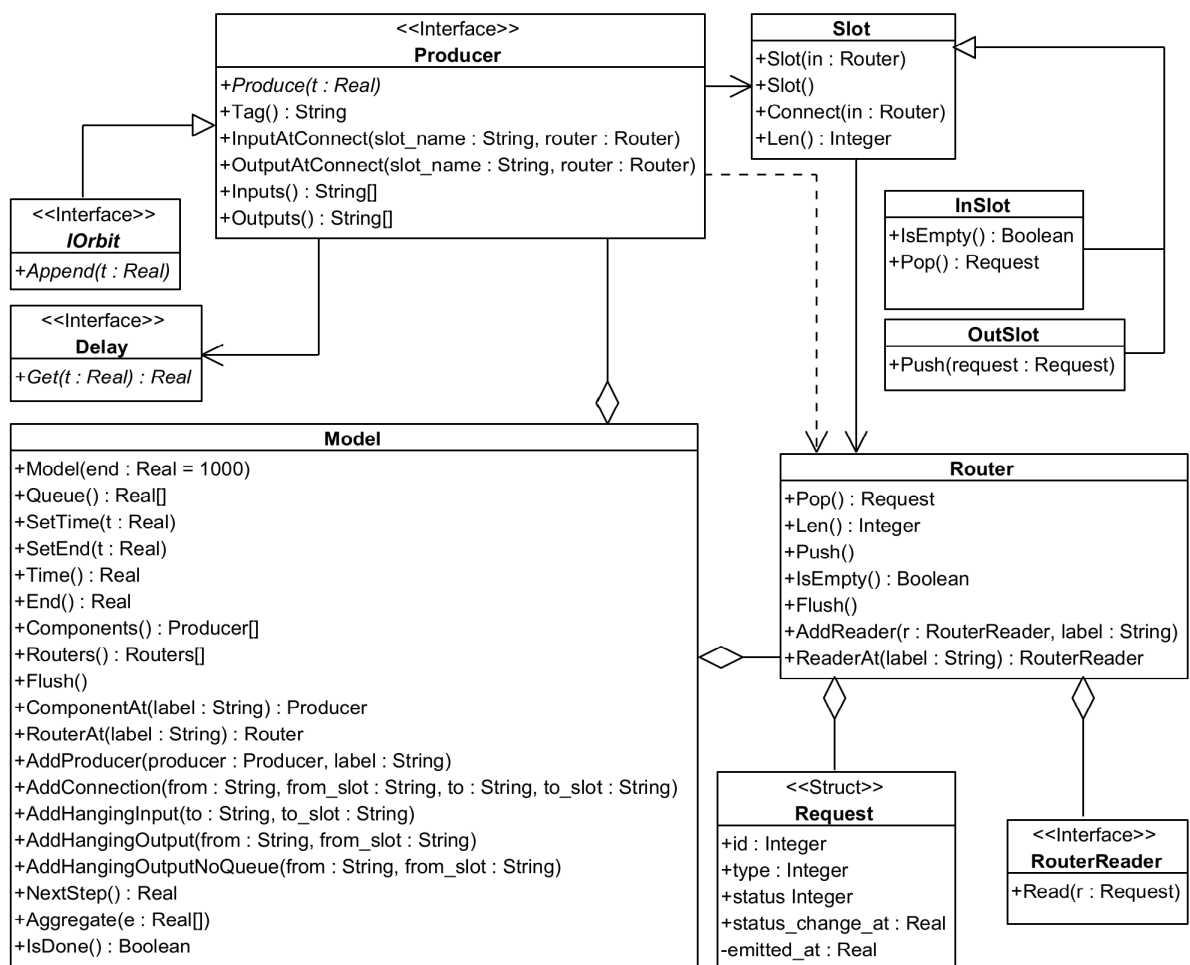


Рисунок 6 — Объектная модель предметной области

Как видно на рисунке 6, предметная область содержит специальный агрегирующий класс `Model`, предназначенный для управления процессом моделирования в рамках одной системы. Он позволяет построить модель при помощи методов `AddProducer` (добавление нового элемента в модель) и `AddConnection` (добавление маршрутизатора между существующими элементами). `AddConnection` имеет три частных случая: `AddHangingInput` — добавляет маршрутизатор на входящий слот одного элемента, `AddHangingOutput` — добавляет маршрутизатор на выходящий слот элемента, `AddHangingOutputNoQueue` — аналогичен `AddHangingOutput`, но заявки не сохраняются в маршрутизаторе. Три перечисленных метода позволяют вручную отправлять заявки в модель, а также устанавливать поток заявок между отдельными моделями.

Метод `Queue` возвращает текущий вектор моментов наступления событий в системе. `SetTime`, `Time`, `SetEnd`, `End` устанавливают либо возвращают текущее условное время и условное время окончания моделирования соответственно. Методы `Components` и `Routers` возвращают список имеющихся в модели элементов и маршрутизаторов соответственно. Методы `ComponentAt` и `RouterAt` предоставляют доступ к конкретному элементу или маршрутизатору. Метод `Aggregate` добавляет в вектор предстоящих событий новые, а метод `NextStep` сдвигает время моделирования на момент наступления ближайшего события и возвращает его. Метод `IsDone` позволяет узнать, закончилось ли моделирование (текущее время моделирования стало равно времени окончания моделирования).

Иначе содержимое модели можно представить графически следующим образом (рисунок 7):

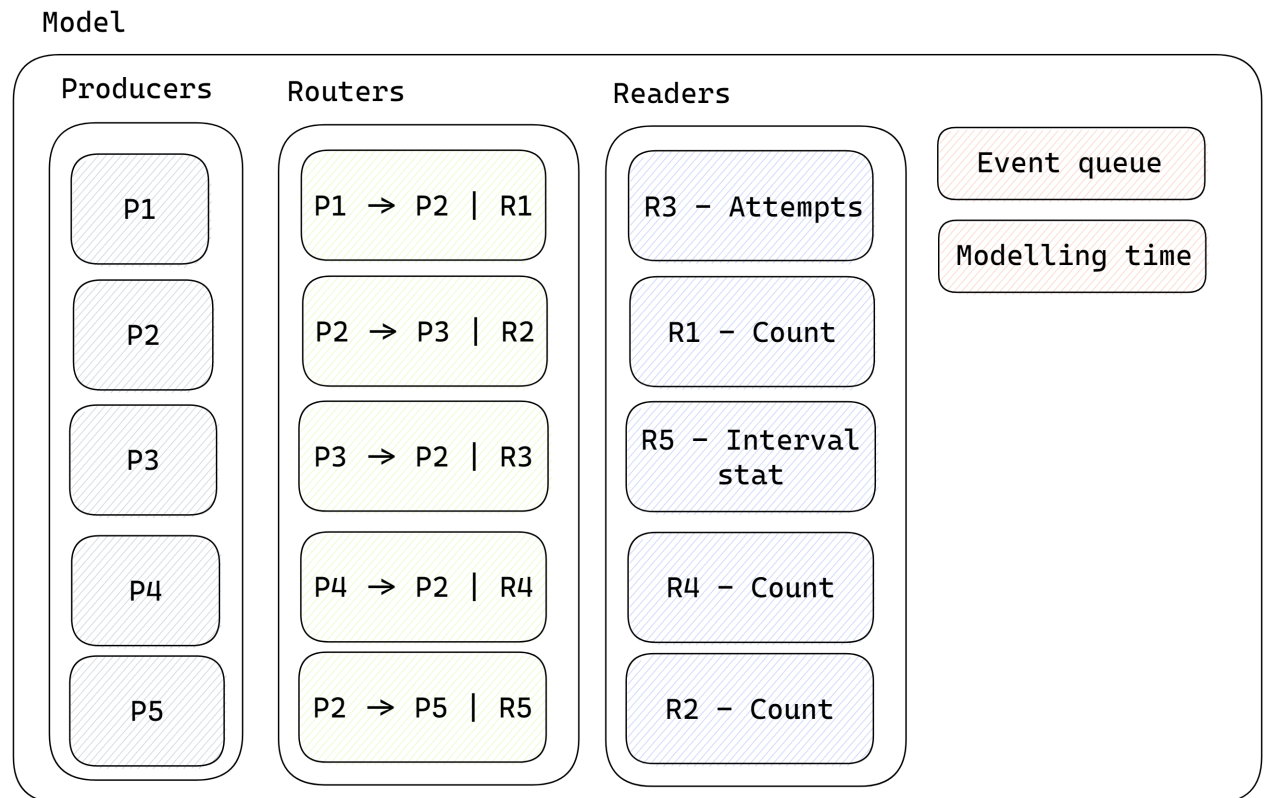


Рисунок 7 — Объектная модель предметной области

2.3 Реализации интерфейсов

В данном разделе представлены имеющиеся реализации интерфейсов и родительских классов, описанных ранее в предметной области, но не представленных на рисунке 6.

Для интерфейса `Producer` существует три ряда реализаций, разделенных по предназначению.

Первым классом являются обслуживающие приборы. Среди них:

1. `SimpleNode` — реализация простейшего обслуживающего прибора, который обслуживает входящие заявки и передает их далее. В случае, если заявка не может получить доступ к прибору, она теряется;
2. `RQNode` — реализация обслуживающего прибора с орбитой. Функционирует аналогично простейшему прибору, однако заявки, не сумевшие попасть на прибор перемещаются на орбиту, откуда снова пытаются захватить его;
3. `RQTNode` — аналогичен прибору с орбитой, но обслуживает два типа за-

явок: входящие и приходящие из источника повторных вызовов.

Далее описаны источники заявок — объекты, имеющие только выход для порожденных заявок:

1. SimpleInput — реализация простейшего потока;
2. MMPPInput — реализация MAP–потока, имеющего управляющую цепь, которая определяет состояние процесса, в соответствии с которым меняется и интенсивность порождения заявок.

Для орбиты имеется лишь одна реализация.

Отметим, что каждая из реализаций имеет в качестве параметра генераторы случайных задержек, что позволяет существенно менять характеристики функционирования элемента. Для этого интерфейса (Delay) имеются следующие реализации:

1. ExponentialDelay — генератор экспоненциально распределенного времени задержки;
2. UniformDelay — генератор равномерно распределенного времени задержки;
3. GammaDelay — генератор задержки, имеющей гамма распределение;
4. WeibullDelay — генератор задержки, имеющей распределение Вейбулла [21];
5. LognormalDelay — генератор задержки, имеющей логнормальное распределение.

Для маршрутизатора, помимо основной реализации, имеются два особых случая:

1. NoneRouter — очередь всегда пуста, а заявки, попадающие в нее, уничтожаются. Данный класс предназначен для тех случаев, когда в заявки надо направить в тупик после обслуживания и их учет не требуется;
2. OutputRouter — отличается от NoneRouter тем, что учет проходящих по очереди заявок ведется, однако они по-прежнему не хранятся.

Сборщики статистики:

1. `IntervalRouterReader` — сборщик статистики, предназначенный для получения распределения вероятностей числа прошедших через маршрутизатор заявок разного типа. Класс имеет методы для получения суммарного распределения, среднего, среднеквадратического отклонения и другие;
2. `AttemptCounter` — сборщик, учитывающий количество раз, которое каждая заявка прошла через маршрутизатор, а также суммарное время, потраченное на перемещение до цели;
3. `TimeCounter` — сборщик, учитывающий моменты времени, когда заявки проходили через маршрутизатор.

2.4 Особенности реализации

В данном разделе описаны подробности реализации имитационной модели.

Исходный код, реализующий предметную область, выполнен на языке программирования C++ стандарта 2020 года (C++2a). Выбор в сторону C++ был сделан исходя из требований к инструменту, которые включают производительность и возможность интеграции непосредственно в процесс исследования, о чем подробнее изложено ниже. C++ отлично подходит для написания производительного программного обеспечения, а также обеспечивает гибкость в плане архитектурных решений и управления памятью. Помимо этого C++ является объектно-ориентированным языком программирования, что важно для реализации предметной области. Однако с некоторыми различиями, такими, как отсутствие интерфейсов в явном виде. Вместо них C++ поддерживает абстрактные классы с чистыми виртуальными функциями [22]. Помимо этого важными аспектами являются:

1. ручное управление памятью. С одной стороны, необходимость разработчику постоянно следить за выделяемой памятью и разрушением экземпляров объектов является недостатком, так как осложняет процесс разработки и способствует возникновению ошибок в управлении памятью, с другой — дает возможность организовать подходящий под задачу процесс выделения памяти;

2. компиляция. Компилируемые языки всегда превосходят в производительности интерпретируемые, поскольку проверка типов, предупреждение возникших исключений и других ошибок происходят на этапе компиляции, а не в процессе работы программы;
3. статическая типизация. Заранее известные типы объектов ускоряют как компиляцию, так и работу программы ввиду отсутствия проверки типов;
4. объявление без инициализации. C++ позволяет объявить переменную, не инициализируя ее некоторым значением, что экономит время работы.

Для сборки программы использовался компилятор GCC [23], поскольку он поддерживается всеми платформами, а также имеет широкий спектр параметров для оптимизации кода на этапе компиляции [24].

Также стоит отметить подход к генерации случайных величин во время моделирования. Поскольку получаемые величины по своей природе являются псевдослучайными, то генерируются они с использованием в качестве параметров такта центрального процессора или системного времени. При использовании такого генератора во время имитационного моделирования получаемые данные будут недостоверны, поскольку генерация будет производиться крайне часто — для каждого элемента модели в каждой итерации, что вызовет появление повторяющихся значений в генерируемой последовательности чисел. Для решения данной проблемы используется вихрь Мерсенна [25] — алгоритм генерации псевдослучайных чисел, основывающийся на свойствах простых чисел Мерсенна. Алгоритм гораздо менее предсказуем, имеет большой период повторения и практически полное отсутствие статистической закономерности.

В стандартной библиотеке C++ имеется реализация вихря Мерсенна под названием MT19937 [26], где число в названии связано с величиной его периода — 2^{19937} , которой хватит на генерацию псевдослучайных чисел без повторения в течение продолжительного времени моделирования.

2.5 Реализация программы в качестве программного пакета для Python

Ключевой особенностью реализуемого ПО является возможности работы с ним в среде, где проводится остальные операции по исследованию и анализу системы массового обслуживания. По этой причине инструмент является

программным пакетом (библиотекой) для языка Python. Данный язык крайне удобен и популярен для проведения статистического анализа, машинного обучения и других научных исследований, а имеющая оболочка для интерактивной работы IPython позволяет объединить все аспекты работы в общей среде исполнения.

Python реализован на C++, что позволяет разрабатывать к нему расширительные пакеты не только на самом языке, но и на языке реализации, что дает преимущество в производительности и вариативность. Примером подобного пакета, является фундаментальная библиотека для вычислений линейной алгебры и работы с матрицами NumPy, которая лежит в основе многих других пакетов и написана на C++ [27].

Исходный код, относящийся к предметной области строго изолирован, чтобы предоставить максимальную расширяемость имеющихся интерфейсов. За привязку исходного кода с библиотеке Python отвечает библиотека Pybind11 [28]. Она позволяет скомпилировать код, написанный на языке C++ в пакет для Python, при этом код изначально может быть изначально не предназначен для этого. Таким образом, становится возможным строго разделить интерфейс, с которым взаимодействует пользователь и внутреннюю логику работы программы.

Ниже приведен пример привязки исходного кода класса Router. Тело методов опущено для лучшей читаемости.

Код на C++:

```
class Router
{
public:
    std::vector<Request> q;
    std::unordered_map<std::string, RouterReader &> readers;
    friend class InSlot;
    friend class OutSlot;
    friend class Slot;

    Router()
    void AddReader(RouterReader &r, std::string label)
    std::unordered_map<std::string, RouterReader &> &Readers()
    RouterReader &ReaderAt(std::string label)
    virtual Request Pop()
    virtual int Len()
    virtual void Push(Request request)
    virtual bool IsEmpty()
    virtual void Flush()
```



```
};
```

Листинг 1 — Исходный код класса Router

Как видно из листинга 1, маршрутизатор содержит очередь `q`, словарь ссылок на сборщиков статистики `readers`, а также объявления о дружественных классах для семейства классов `Slot`, поскольку они тесно связаны между собой. Ниже представлена привязка данного класса при помощи `Pybind11`, позволяющая инстанцировать объекты в интерпретаторе Python и работать с ними, но управление памятью и вызов функций будет происходить на стороне C++. Другими словами, с помощью такого подхода создается оболочка для реализации предметной области программы на C++:

```
py::class_<Router>(m, "Router", "Request queue")
.def(py::init())
.def("len", &Router::Len, "Returns number of requests contained")
.def("push", &Router::Push, "request"_a, "Push request in queue")
.def("pop", &Router::Pop, "Pop request from queue")
.def("is_empty", &Router::IsEmpty, "Check if queue is empty")
.def("flush", &Router::Flush, "Clears router queue")
.def("add_reader", &Router::AddReader, "reader"_a, "label"_a,
    py::keep_alive<1, 2>(), "Add request reader")
.def("readers", &Router::Readers, py::return_value_policy::
    reference, "dictionary of readers")
.def("reader_at", &Router::ReaderAt, py::keep_alive<1, 0>(), py
    ::return_value_policy::reference, "get reader")
.def_readonly("pushed_count", &Router::pushed_count, "number of
    pushed requests")
.def_readonly("popped_count", &Router::popped_count, "number of
    popped requests")
.def_readwrite("__readers__", &Router::readers)
.def_readonly("__q__", &Router::q);
```

Листинг 2 — Привязка класса Router при помощи Pybind11

Так, `Pybind11` позволяет переименовать сигнатуры функций, в частности, указать новые названия методов, аргументов, а также указать пояснение для документации к каждому методу и классу. То же самое возможно сделать для всего модуля в целом, что позволит совместить его в одном пакете с имеющимся кодом на Python, либо создать вокруг нее еще один слой интерфейса:

```
PYBIND11_MODULE(simulation, m)
{
    m.attr("__name__") = "q_analysis.simulation";
    m.doc() = "Python library for retrial queuing system modeling";
    ...
}
```

Помимо этого, важно отметить, что при привязке требуется указать, как именно будет происходить распоряжение памятью при создании и удалении объектов. В Python все переменные являются указателями, которые содержат в себе счетчик ссылок, означающий количество ссылающихся указателей на область в памяти. Как только счетчик становится равным нулю, объект уничтожается интерпретатором, что может повлечь за собой различного рода ошибки доступа к памяти и ее утечку. Во избежание перечисленных ошибок Pybind11 позволяет указать время жизни объекта, являющегося аргументом и возвращаемым значением по принципу Nurse–Patient:

```
py::class_<Router>(m, "Router", "Request queue")
    .def("add_reader", &Router::AddReader, "reader"_a, "label"_a,
        py::keep_alive<1, 2>(), "Add request reader")
    .def("reader_at", &Router::ReaderAt, py::keep_alive<1, 0>(), py
        ::return_value_policy::reference, "get reader")
```

Листинг 4 — Указание времени жизни объекта и тип возвращаемого значения в Pybind11

В листинге 4 для метода `add_reader` указано выражение `py::keep_alive<1, 2>()`, что означает следующее: аргумент метода `reader` (индекс 2) не должен удаляться, пока не удален вызывающий метод объект (индекс 1). Данное выражение предотвратит ситуацию, когда аргумент `reader` может быть удален из памяти еще до того, как попадет в функцию `add_reader`. Иными словами, выражение `keep_alive<1, 2>` указывает, что объект `j` должен прожить хотя бы столько же, сколько проживет объект `i`. Индексом, равным единице, всегда является вызывающий объект; последующие индексы относятся к аргументам; индекс, равный нулю, означает возвращаемое значение функции. Пример использования указания времени жизни возвращаемого значения можно наблюдать в привязке метода `reader_at`: поскольку данный метод является аналогом оператора индексации, крайне важно, чтобы ссылка на объект, возвращаемая методом не была удалена при обращении. Также для этого же метода указана политика возвращаемого значения — `reference`. Это означает, что при возвращении значения из функции управление данным объектом не переходит к Python, и ответственным за своевременное удаление объекта из памяти является код на стороне C++.

Для корректной строковой репрезентации объектов пакета используется атрибут Python–объектов под названием `__repr__`, так как по умолчанию обь-

екты классов, принадлежащие стороне C++ будут отображаться в формате `<module.class_name object at <адрес в памяти> >`.

2.6 Интеграция в процесс исследования

После успешной компиляции проекта, становится возможным использовать программу как нативный пакет Python, что и позволяет встроить программу непосредственно в среду, где проводится исследование. Данное решение также избавляет от необходимости реализации специальных классов для инстанцирования объектов с применением порождающих шаблонов проектирования (таких, как абстрактная фабрика [18]. В свою очередь процесс работы был вдохновлен известной библиотекой для машинного обучения PyTorch [29], где модель, в данном случае нейронная сеть, предварительно составляется из набора слоев требуемой конфигурации и уже в последствии обучается. Попытка использовать подобный принцип была сделана и в случае имитационной модели:

```
In [1]: from q_analysis import simulation as q
        model = rq.Model()
        model.set_time(0)
        model.set_end(1000000)

In [2]: model.add_producer(rq.SimpleInput(rq.ExponentialDelay(1.1),0,0),"input")
        model.add_producer(rq.SimpleInput(rq.ExponentialDelay(0.6),1,0),"call")
        model.add_producer(rq.Orbit(rq.ExponentialDelay(0.81)),"orbit")
        model.add_producer(
            rq.RqtNode(rq.ExponentialDelay(1.3),
                rq.ExponentialDelay(1)),"node")

In [3]: nodein = model.add_connection("input","out_slot","node","in_slot")
        model.add_connection("call","out_slot","node","call_slot")
        model.add_connection("orbit","out_slot","node","orbit_slot")
        output = model.add_hanging_output_noqueue("node","out_slot")
        model.add_connection("node","orbit_append_slot","orbit","in_slot")

In [4]: model.router_at(output).add_reader(rq.TimeCounter(),"count")
```

В ячейке [1] производится создание модели и задание времени моделирования. Далее, в ячейке [2], производится добавление элементов в модель:

- простейший поток с экспоненциально распределенными моментами порождения заявок, интенсивностью 1.1 и меткой «input»;

- простейший поток с экспоненциально распределенными моментами порождения заявок, интенсивностью 0.6 и меткой «call» в качестве источника вызываемых заявок;
- орбита с интенсивностью повторного обращения заявок 0.81 и меткой «orbit»;
- обслуживающий прибор с повторными обращениями и обратной связью, обрабатывающий входящие заявки в течение экспоненциально распределенной задержки с интенсивностью 1.3 и вызываемые заявки в течение экспоненциально распределенной задержки с интенсивностью 1. Имеет метку «node».

В ячейке [3] производится маршрутизация заявок между элементами модели:

- заявки из входящего потока «input» попадают на прибор «node»;
- заявки из источника вызываемых заявок «call» попадают на прибор «node»;
- заявки, находящиеся на орбите «orbit» пытаются попасть обратно на прибор «node»;
- заявки, не сумевшие захватить прибор «node» отправляются на орбиту «orbit»;
- заявки, успешно обработанные прибором «node» отправляются в тупик.

В ячейке [4] к маршрутизатору, который представлен как тупик для обслуженных заявок, добавляется сборщик моментов прохождения заявок через маршрутизатор «count». После этого модель готова к запуску и графически может быть представлена как на изображено на рисунке 8.

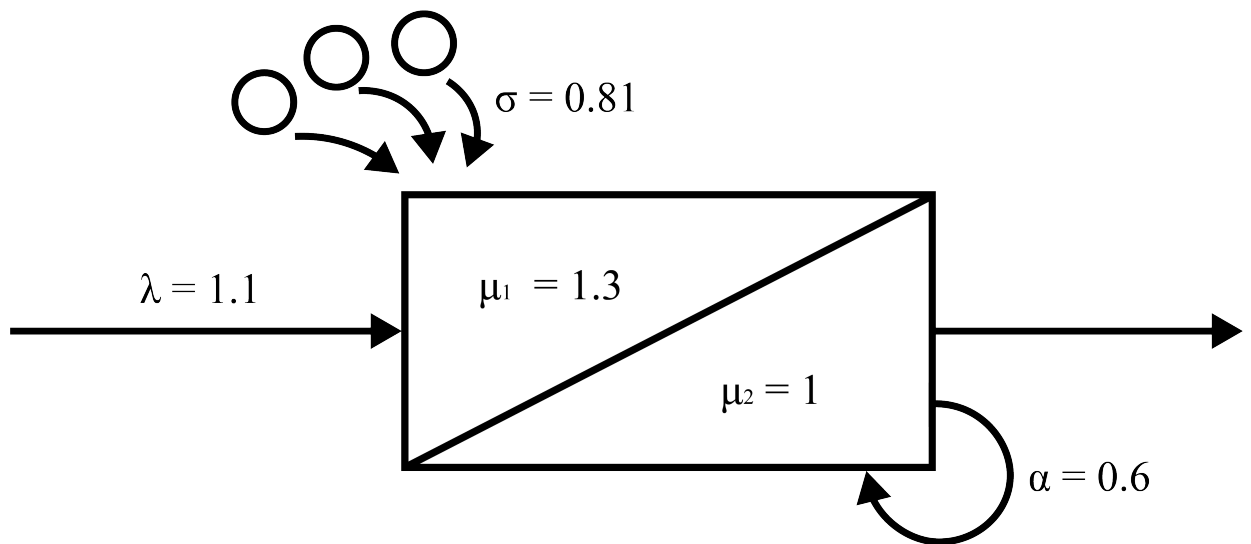


Рисунок 8 — Пример составленной модели

Таким образом, реализация программного пакета в качестве библиотеки для Python позволяет проводить численные эксперименты непосредственно в среде, где проводится статистический анализ, апробация результатов и остальные аспекты исследования.

3 Вычисление асимптотических результатов

Как известно, любая случайная величина может быть однозначно определена функцией распределения вероятностей или характеристической функцией [30]. При этом характеристическая функция может быть определена для любой случайной величины (дискретной или непрерывной).

Объектом исследования теории массового обслуживания являются случайные величины и случайные процессы, которые могут быть как непрерывными (например, время ожидания заявки до начала обслуживания, период занятости системы, объем занятого ресурса), так и дискретными (например, число заявок в очереди, число событий, наступивших в потоке за некоторый промежуток времени, число занятых каналов обслуживания). При этом эти величины по своей природе принимают в основном неотрицательные значения.

Для неотрицательных случайных величин характеристическая функция является частным случаем преобразования Лапласа-Стилтьеса от функции распределения вероятностей с мнимым аргументом. Соответственно, характеристическая функция обладает свойствами преобразования Лапласа-Стилтьеса. С другой стороны, в общем случае взаимное соответствие характеристической функции и функции распределения вероятностей соответствует теории Фурье-анализа, безусловным преимуществом которого является свойство двойственности, то есть схожесть переходов от одной функции к другой. Это позволяет, решая задачу для характеристических функций, в последствии по явным интегральным формулам переходить к функциям распределения.

Для дискретных распределений принято использовать не характеристическую функцию, а производящую [31], но во многих работах для дискретных распределений используется тоже характеристическая функция. Это позволяет получать формулы перехода к распределению в терминах рядов Фурье через интегрирование характеристической функции по периоду 2π . С точки зрения записи формул никаких проблем не возникает, но при проведении численных экспериментов, в зависимости от сложности вида характеристической функции, численное интегрирование является задачей либо очень трудоемкой, либо вообще невозможной.

В данной работе предлагается использовать теорию Фурье-анализа сигналов для обращения характеристических функций дискретных распределе-

ний. Идея заключается в применении принципов теоремы Котельникова [32, 33] для дискретизации характеристических функций и применении дискретного преобразования Фурье, которое вместо интегрирования использует суммирование и умножение.

3.1 Связь характеристических функций с сигналами и их спектрами

Характеристической функцией $h(u)$ случайной величины ξ называется

$$h(u) = M\{e^{ju\xi}\}. \quad (2)$$

Характеристической функцией $h(u)$ дискретной случайной величины ξ с распределением вероятностей p_i называется

$$h(u) = M\{e^{ju\xi}\} = \sum_{i=-\infty}^{\infty} e^{ju\xi_i} p_i. \quad (3)$$

Характеристической функцией $h(u)$ дискретной неотрицательной целочисленной случайной величины ξ ($\xi = 0, 1, 2, \dots$) с распределением вероятностей p_i ($i = 0, 1, 2, \dots$) называется

$$h(u) = M\{e^{ju\xi}\} = \sum_{i=0}^{\infty} e^{ju\xi_i} p_i. \quad (4)$$

При этом для нахождения распределения вероятностей p_i ($i = 0, 1, 2, \dots$) по характеристической функции $h(u)$ используется формула обращения

$$p_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ju\xi_i} h(u) du. \quad (5)$$

Можно заметить, что формулы (3) и (4) являются формулами разложения сигнала $h(u)$ в ряд Фурье в комплексной форме [34], а формула (5) нахождения вероятностей p_i совпадает с формулой вычисления коэффициентов ряда Фурье [34]. Таким образом, в терминах теории сигналов характеристическая функция (3) является непрерывным комплекснозначным периодическим сигналом, а распределение вероятностей p_i ($i = 0, 1, 2, \dots$) является спектром этого сигнала.

3.2 Дискретизация характеристической функции

Основные вычислительные сложности в задачах теории массового обслуживания возникают как раз при интегрировании по формуле (5) для нахождения распределения вероятностей p_i . При усложнении вида полученной характеристической функции $h(u)$ в зависимости от вычислительных ресурсов компьютера вероятности p_i либо считаются долго, либо их вовсе не удается вычислить.

Для уменьшения вычислительной сложности нахождения вероятностей предлагается воспользоваться аппаратом Дискретного преобразования Фурье [35], которое в соответствие дискретному сигналу h_k (последовательность N значений) ставит в соответствие его спектральные отсчеты p_i^* (последовательность N значений)

$$p_i^* = \left| \frac{1}{N} \sum_{k=0}^{N-1} h_k \cdot e^{-j \cdot i \cdot \frac{2\pi}{N} k} \right|, i = 0, 1, \dots, (N - 1). \quad (6)$$

Таким образом, необходимо дискретизировать характеристическую функцию $h(u)$ на периоде 2π таким образом, чтобы дискретное преобразование Фурье p_i^* от него было максимально близко к p_i

$$\Delta u = \frac{2\pi}{N}, h_k = h(k * \Delta u), k = 0, 1, \dots, N - 1. \quad (7)$$

При дискретизации сигнала (характеристической функции $h(u)$) с шагом Δu его спектр (распределение вероятностей p_i^*) начинает дублироваться с периодом Ω

$$\Omega = \frac{2\pi}{\Delta u} = N. \quad (8)$$

В этом случае, если распределение вероятностей p_i на интервале от 0 до N равно или близко к 1, то периодизация спектра не внесет в него значимых изменений на периоде N . Если же интервал от 0 до N не содержит всей или почти всей информации о распределении, то при дискретизации характеристической функции и периодизации спектра будет происходить наложение копий с периодом N , что приведет к расхождению вычисленного p_i^* и истинного p_i . Факт дублирования спектра при периодизации сигнала и возможного его наложения лежит в основе теоремы Котельникова, которая определяет выбор частоты дискретизации, позволяющее избежать такого наложения спектров.

3.3 Иллюстрация работы подхода на биномиальном распределении

Рассмотрим пример дискретизации характеристической функции на примере биномиального распределения с параметрами n (количество испытаний) и p (вероятность наступления события).

Характеристическая функция будет иметь вид

$$h(u) = (q + pe^{iu})^n. \quad (9)$$

Продемонстрируем, что обращение характеристической функции можно провести как при помощи интегрального обратного преобразования Фурье для дискретных случайных величин (5), так и при помощи дискретного преобразования Фурье (6). Результат вычислений представлен на графике

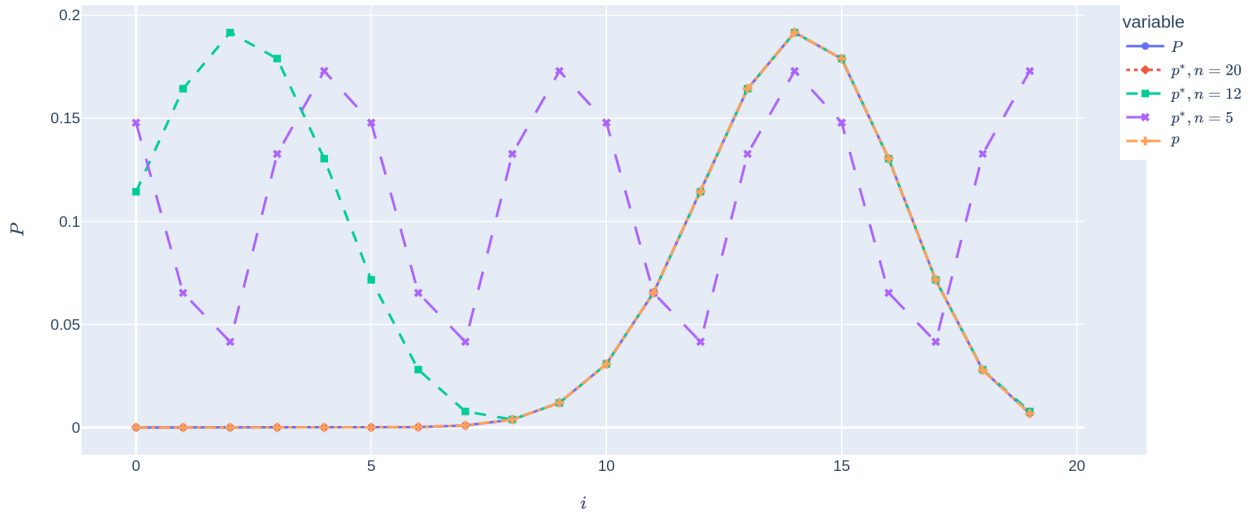


Рисунок 9 — Обращение характеристической функции

Вычисление было произведено на языке Python с заданными параметрами

$$n = 20, p = 0.7, N = \{20, 12\}.$$

Из рисунка 9 видно, что теоретическое биномиальное распределение P , результат обратного интегрального p (5) и дискретного p^* (6) преобразований совпадают при $N = 20$. Однако, как показывает график, при неверном выборе шага дискретизации (в данном случае $N < n$, $N = 12$) распределение начинает искажаться, так как условие теоремы Котельникова [32] не выполняется.

3.4 Иллюстрация работы подхода на задаче теории массового обслуживания

В данном разделе продемонстрируем эффективность предлагаемого подхода при реализации численных расчетов в задаче из [14]. В данной работе получено асимптотическое приближение характеристической функции $h(u, t)$ числа обслуженных заявок в системе с повторными обращениями и вызываемыми заявками

$$h(u, t) = \mathbf{R}e^{\mathbf{G}(u)t}\mathbf{E}, \quad (10)$$

которая при фиксированном t является функцией только от u .

Здесь матрица $\mathbf{G}(u)$ содержит коэффициенты системы дифференциальных уравнений Колмогорова. Ее элементы выражаются через параметры модели. \mathbf{R} — вектор-строка, \mathbf{E} — единичный вектор-столбец.

Для вычисления характеристической функции необходимо вычислять матричную экспоненту $e^{\mathbf{G}(u)}$, что безусловно делает задачу трудоемкой. Здесь матричную экспоненту вычисляем при помощи преобразования подобия матриц [36]:

$$e^{\mathbf{G}(u)} = \mathbf{T}(u) \cdot \mathbf{GJ}(u) \cdot \mathbf{T}(u)^{-1},$$

где $\mathbf{T}(u)$ — матрица собственных векторов $\mathbf{G}(u)$, $\mathbf{GJ}(u)$ — диагональная матрица собственных чисел Λ_n матрицы $\mathbf{G}(u)$.

Вычисление распределения вероятностей числа обслуженных заявок в системе за некоторое фиксированное время t через интегрирование с помощью формулы (5) является, как уже было упомянуто, достаточно трудоемкой вычислительной задачей. Для решения этой проблемы предлагается воспользоваться формулой (6) ДПФ.

На рисунке 10 видно, что результат дискретного преобразования Фурье (p^*) полностью совпадает с результатом вычисления при помощи интегрирования (p)

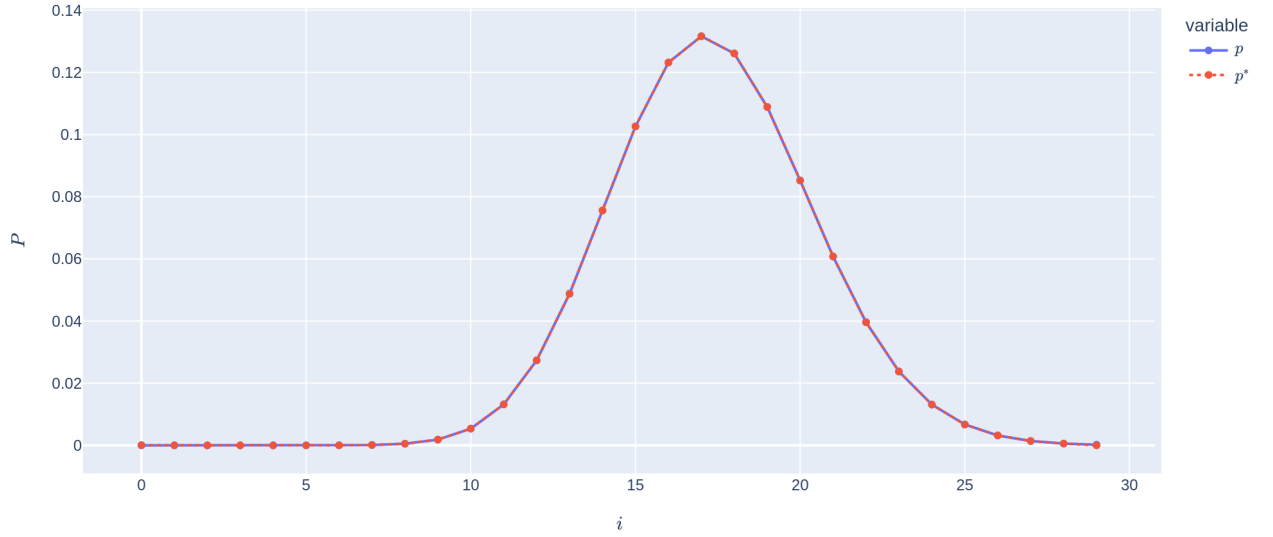


Рисунок 10 — Сравнение распределений вероятностей, полученных с помощью интегрирования и ДПФ

Для проверки скорости работы данного подхода к вычислению был проведен ряд тестов (400 запусков) со сравнением скорости работы алгоритмов и точности получаемого распределения при помощи расстояния Колмогорова

$$\Delta = \max_{0 < i < \infty} \left| \sum_{v=0}^i (P_0(v) - P_1(v)) \right|.$$

В среднем ДПФ быстрее интегрирования более, чем в 100 раз, а среднее расстояние Колмогорова — $4.16 \cdot 10^{-6}$.

3.5 Вычисление коэффициента вариации ММРР

В данной работе одним из объектов изучения является вариация длин интервалов между моментами поступления заявок ММРР.

Согласно [37], вариация длин интервалов между моментами поступления заявок ММРР вычисляется как

$$Var = \frac{\sqrt{v}}{Lg^{-1}}, \quad (11)$$

где v — дисперсия длин интервалов между моментами поступления групп запросов и рассчитывается как

$$v = \frac{2Lg \cdot r \cdot (-D0)^{-1} \cdot E - 1}{Lg^2}, \quad (12)$$

где вектор–строка r — стационарное распределение вероятностей процесса $\{k(t), n(t)\}$, E — единичный вектор–столбец размерности N , Lg — интенсивность входящего потока

$$Lg = r \cdot \Lambda \cdot E, \quad (13)$$

$$D0 = Q - \Lambda - Q \cdot D,$$

где D — матрица, содержащая вероятности наступления события в потоке при смене его состояния.

4 Работа с имитационной моделью

В данном разделе изложены подходы к работе с реализованным пакетом программ, показывающие течение процесса исследования при помощи имитационного моделирования и дополнительных утилит для расчета характеристик.

4.1 Определение области применимости асимптотических результатов исследования модели RQ-системы с повторными вызовами и обратной связью

В главе 3.4 описаны асимптотические результаты, в частности приближение характеристической функции числа обслуженных заявок разного типа за время t . Последующей задачей ставится определение области применимости полученного приближения.

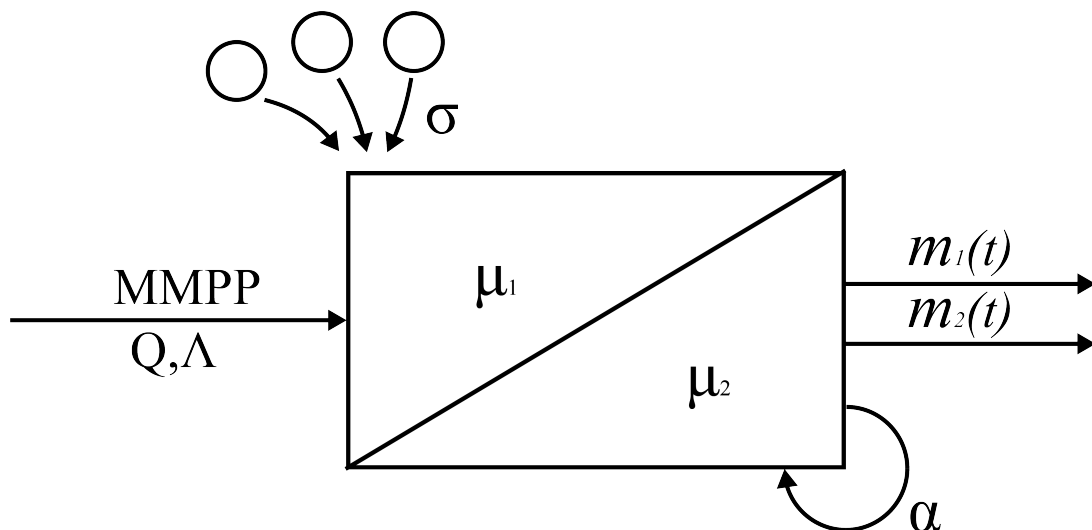


Рисунок 11 — Модель RQ-системы с повторными вызовами и обратной связью

Поскольку решение было получено при асимптотическом условии бесконечной задержки заявок в источнике повторных вызовов, требуется определить, при каких параметрах модели асимптотические результаты становятся недостоверными. Для этой цели как раз отлично подходит имитационное моделирование.

В первую очередь, импортируем модуль наш `q_analysis`. Он, в свою очередь, содержит подмодуль `simulation` для имитационного моделирования. Именно содержит реализацию ранее описанной предметной области.

```
In [5]: import q_analysis.simulation as q
```

Создадим модель и установим время моделирования, равное 1000000 условных единиц.

```
In [6]: model = rq.Model()
model.set_time(0)
model.set_end(1000000)
model
```

```
Out[6]: Model{ time: 0, end : 1000000, event_queue_len : 0, num_components : 0,
          num_connections : 0}
```

Далее добавим элементы модели (метод `add_producer`).

В модель добавятся следующие элементы:

- ММРР поток с матрицей Λ следующего вида

$$\Lambda = \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 1.12 & 0 \\ 0 & 0 & 0.45 \end{bmatrix}$$

и матрицей Q

$$\Lambda = \begin{bmatrix} -0.4 & 0.3 & 0.1 \\ 0.5 & -0.6 & 0.1 \\ 0.3 & 0.6 & -0.9 \end{bmatrix}$$

;

- Орбита с экспоненциальной задержкой с интенсивностью 1;
- Простейший поток с экспоненциальной задержкой с интенсивностью 1 (будет использоваться как источник вызываемых заявок прибора);
- Обслуживающий прибор с экспоненциальным временем обслуживания входящих и вызываемых заявок с интенсивностью 1.15.

Чтобы обращаться к элементам модели, им даются ярлыки - `input`, `call`, `orbit`, `node`.

In [7]:

```
model.add_producer(rq.MMPPInput(  
    [1,1.12,0.45],  
    [[ -0.4, 0.3, 0.1],  
     [0.5,-0.6,0.1],  
     [0.3,0.6,-0.9]],0,0),"input")  
model.add_producer(rq.SimpleInput(rq.ExponentialDelay(1),1,0),"call")  
model.add_producer(rq.Orbit(rq.ExponentialDelay(1)),"orbit")  
model.add_producer(rq.RqtNode(rq.ExponentialDelay(1.15),rq.ExponentialDelay(  
    1.15)),"node")  
print("Components:",model.components())
```

Out [7]:

```
Components: {'node': RQTNode, 'orbit': Orbit, 'input': MMPP, 'call':  
    SimpleInput}
```

Далее требуется соединить элементы модели при помощи маршрутизаторов. Для этого используется метод `add_connection`. В параметрах указывается ранее заданный ярлык элемента модели, потом его вход или выход, функция возвращает ярлык нового соединения.

In [8]:

```
model.add_connection("input","out_slot","node","in_slot")
```

В данном случае мы указали входящий поток `input` и его выход `out_slot` как источник заявок, и прибор `node` и его вход `in_slot` как вход для приходящих заявок. Получим следующий ключ, описывающий соединение:

Out [8]:

```
'input:out_slot:node:in_slot'
```

С помощью полученной строки в последствии можно обращаться к соответствующими маршрутизатору при помощи метода `reader_at`.

Далее добавляются остальные соединения. Для обслуживающего прибора маршрутизатор для обслуженных заявок не будет их сохранять в памяти. Это достигается использованием метода `add_hanging_output_noqueue`

In [9]:

```
model.add_connection("call","out_slot","node","call_slot")  
model.add_connection("node","orbit_append_slot","orbit","in_slot")  
orb = model.add_connection("orbit","out_slot","node","orbit_slot")  
output = model.add_hanging_output_noqueue("node","out_slot")  
print("Connections",model.routers())
```

Out [9]:

```
Connections {
  'onq:node:out_slot': Router{ queue_len: 0},
  'orbit:out_slot:node:orbit_slot': Router{ queue_len: 0},
  'node:orbit_append_slot:orbit:in_slot': Router{ queue_len: 0},
  'input:out_slot:node:in_slot': Router{ queue_len: 0},
  'call:out_slot:node:call_slot': Router{ queue_len: 0}
}
```

Ярлыки маршрутизаторов называются по следующей схеме:

- *ярлык выходного элемента : имя выхода : ярлык входного элемента : имя входа ;*
- для аккумулирующего входного маршрутизатора (`model.add_hanging_input`)
— *i : ярлык входного элемента : имя входа*
- для аккумулирующего выходного маршрутизатора (`model.add_hanging_output`)
— *o : ярлык выходного элемента : имя выхода*
- для выходного маршрутизатора (`model.add_hanging_output_noqueue`) — *onq : ярлык выходного элемента : имя выхода*

Добавим сбор статистики с интервалом в 20 условных единиц, что будет означать, что мы измеряем, сколько заявок было обслужено в системе за $t = 20$:

In

[10]:

```
model.router_at(output).add_reader(rq.IntervalRouterReader(20), "stat")
model.router_at(output).readers()
```

Out [10]:

```
'stat': IntervalRouterReader
```

Теперь когда модель создана, нам требуется описать, как именно будет проходить ее запуск. Маршрутизация определяет топологию передачи заявок по системе, однако порядок взаимодействия элементов модели указывается в качестве отдельного алгоритма, так как, в зависимости от специфики задачи, он может варьироваться. Процесс моделирования сводится к итеративному вызову метода `produce` у элементов модели с указанием в качестве параметра текущего времени моделирования. Результат работы `produce` (список моментов наступления предстоящих событий) передается в функцию модели `aggregate` для того, чтобы далее смещать время на нужные моменты. Именно порядок вызовов `produce` составляет алгоритм моделирования. В данном случае он будет следующим:

1. генерация очередной заявки входящих потоком;
2. возвращение заявок с орбиты;
3. генерация очередной заявки источником вызываемых заявок;
4. обслуживание заявки;
5. сбор заявок, не сумевших захватить прибор.

Построим цикл:

```
In [11]: from tqdm import tqdm
e = model.end()
with tqdm(total=int(e)) as pbar:
    t = c = 0
    while True:
        c+=1
        told = t
        t = model.next_step()
        pbar.update(t-told)
        model.aggregate(model.component_at("input").produce(t))
        model.aggregate(model.component_at("orbit").produce(t))
        model.aggregate(model.component_at("call").produce(t))
        model.aggregate(model.component_at("node").produce(t))
        model.aggregate(model.component_at("orbit").append(t))
        if model.is_done():
            break
    print("Time: ",model.time())
    print("Iters: ",c)
```

```
Out[11]: Time: 1000000.0437646629
Iters: 21710021
```

По окончании моделирования мы можем проверить, какое распределение числа обслуженных заявок за интервал 20:

In
[12]:

```
distr = model.router_at(output).reader_at('stat').get_distribution_2d()
import plotly.graph_objects as go
fig = go.Figure(data=[go.Surface(z=distr) ])

fig.update_traces(contours_z=dict(show=True, usecolormap=False,
highlightcolor="limegreen", project_z=True))

fig.update_layout(title='Model 2d distribution', autosize=False,
width=1000, height=1000,
margin=dict(l=65, r=50, b=65, t=90))

fig.show()
```

Model 2d distribution

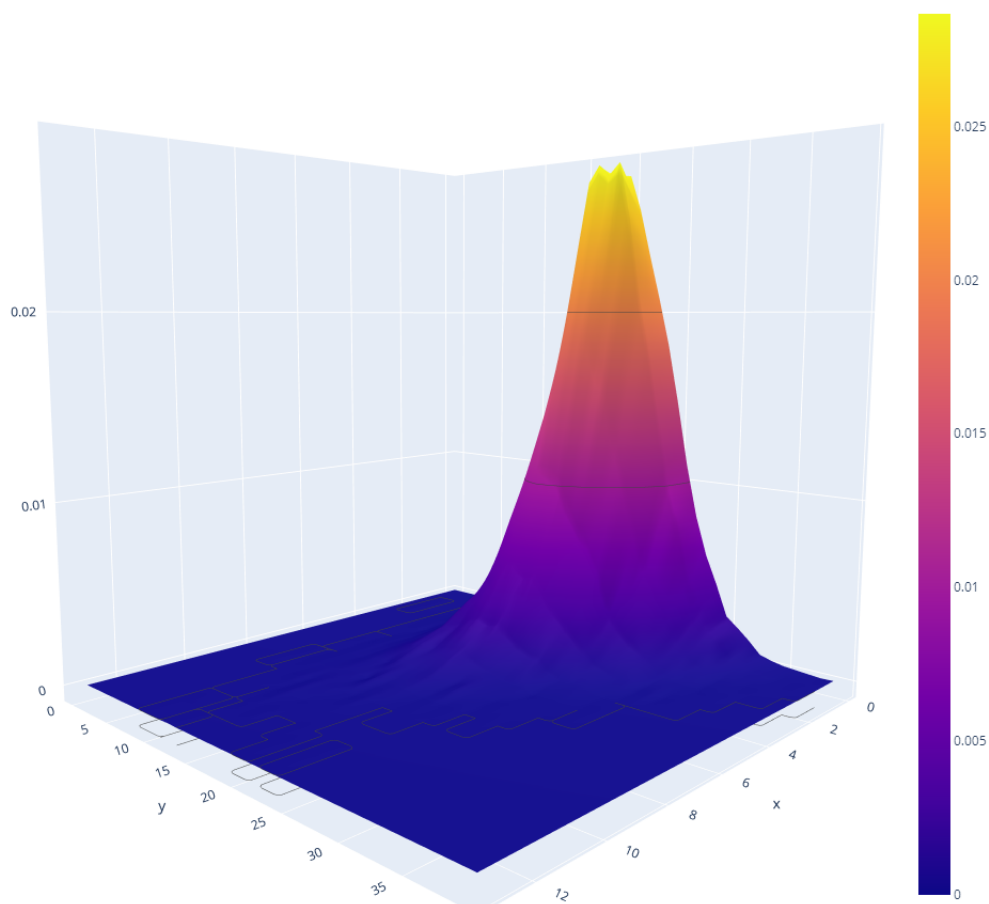


Рисунок 12 — Распределение числа обслуженных заявок по окончании имитационного моделирования

Теперь необходимо сравнить данное распределение вероятностей с аналогичным, но полученным при помощи асимптотического приближения харак-

теристической функции числа обслуженных заявок. Для этого нам потребуется численно решить систему дифференциальных уравнений Колмогорова при помощи имеющегося в пакете класса RQSystem:

ЗАКЛЮЧЕНИЕ

В рамках данной работы был рассмотрен ряд методов оптимизации экспериментального исследования модели системы массового обслуживания с повторными вызовами и вызываемыми заявками. Согласно указанной цели исследования был выполнен ряд задач.

Была переработана имитационная модель с ориентацией на масштабные запуски и параллелизм. Была улучшена ее производительность и отказоустойчивость при задании параметров, затрудняющих процесс моделирования. Также была реализована возможность задания параметров для всех элементов системы из файла, который заранее содержит информацию о требуемых результатах.

Для работы с результатами моделирования были разработаны программные инструменты на языке Python, позволяющие генерировать необходимое количество наборов параметров, проводить множественные запуски имитационной модели и агрегировать полученные данные.

Помимо этого, в рамках программного комплекса были реализованы алгоритмы вычисления характеристик рассматриваемой модели системы массового обслуживания, включающие также вспомогательные функции для вычисления точности распределения вероятностей при помощи расстояния Колмогорова. Ключевой особенностью данного инструмента является наличие функций для подсчета распределения вероятностей при помощи дискретного преобразования Фурье. Их реализация позволяет существенно оптимизировать процесс проведения экспериментов, так как время работы алгоритмов во много раз превосходит ранее использованный подход.

В конечном итоге, разработанные инструменты позволили провести апробацию методов машинного обучения для расчета одной из характеристик работы системы — вариации длин интервалов между моментами покидания заявками прибора. На основе выборки, составленной из 286 тысяч запусков имитационной модели, были обучены алгоритмы LinearRegression, RandomForest,

GradientBoost и CatBoost для решения задачи регрессии. Были выявлены зависимости целевой переменной от ряда признаков, в частности, задержки заявок на орбите и коэффициента вариации входящего потока. Обученные модели позволяют получать результаты с достоверностью около 80%. В дальнейшем исследовании стоит цель увеличить точность предсказаний за счет изменения процедуры генерации наборов параметров таким образом, чтобы распределение каждого признака было более равномерным и охватывало как можно большее количество возможных значений во избежание искажения результатов.

Часть результатов проведенного исследования были представлены в качестве докладов на двух международных конференциях:

- XXIV International Conference on Distributed Computer and Communication Networks (DCCN) (сентябрь 2021, Томск) [14];
- XX Международная конференции по информационным технологиям и математическому моделированию имени А.Ф. Терпугова (ITMM) (декабрь 2021, Томск);

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- [1] Phung-Duc T. Retrial queueing models: A survey on theory and applications // arXiv preprint arXiv:1906.09560. — 2019.
- [2] Baiocchi A., Blefari-Melazzi N. Steady-state analysis of the MMPP/G/1/K queue // IEEE transactions on communications. — 1993. — Vol. 41, no. 4. — P. 531–534.
- [3] Lapatin I. L., Nazarov A. A. Asymptotic Analysis of the Output Process in Retrial Queue with Markov-Modulated Poisson Input Under Low Rate of Retrials Condition // International Conference on Distributed Computer and Communication Networks / Springer. — 2019. — P. 315–324.
- [4] Meier-Hellstern K. A fitting algorithm for Markov-modulated Poisson processes having two arrival rates // European Journal of Operational Research. — 1987. — Vol. 29, no. 3. — P. 370–377.
- [5] Задорожный Владимир Николаевич. Методы аналитико-имитационного моделирования систем с очередями и стохастических сетей : Ph. D. thesis / Владимир Николаевич Задорожный ; СПб., 2011. — 2011.
- [6] Glynn Peter W, Iglehart Donald L. Simulation methods for queues: An overview // Queueing systems. — 1988. — Vol. 3, no. 3. — P. 221–255.
- [7] Назаров А. А., Терпугов А. Ф. Теория массового обслуживания: [учебное пособие по специальностям 010200 (010501) «Прикладная математика и информатика» и 061800 (080116) «Математические методы в экономике». — Томск: Изд-во НТЛ, 2010.
- [8] Назаров Анатолий Андреевич [et al.]. Асимптотический анализ марковизуемых систем. — 1991.
- [9] Исследование RQ-системы $M|M|1$ с разнотипными вызываемыми заявками и ненадежным прибором методом асимптотически-диффузионного анализа / Анатолий Андреевич Назаров [et al.] // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. — 2021. — no. 57. — P. 74–83.

- [10] Имитационное и статистическое моделирование / Виктор Иванович Лобач [et al.] // Практикум для студентов математических и экономических специальностей. Минск: БГУ. — 2004.
- [11] Моисеев Александр Николаевич [et al.]. Исследование математических моделей систем и сетей массового обслуживания с высокоинтенсивными непуассоновскими входящими потоками. — 2016.
- [12] Алиев Тауфик Измайлович. Погрешности моделирования высоконагруженных систем в GPSS World // Научно-технический вестник информационных технологий, механики и оптики. — 2013. — no. 1 (83). — P. 70–75.
- [13] Fasano Giovanni, Franceschini Alberto. A multidimensional version of the Kolmogorov–Smirnov test // Monthly Notices of the Royal Astronomical Society. — 1987. — Vol. 225, no. 1. — P. 155–170.
- [14] Blagin Alexey, Lapatin Ivan. Approximation of the Two-Dimensional Output Process of a Retrial Queue with MMPP Input // International Conference on Distributed Computer and Communication Networks / Springer. — 2021. — P. 333–345.
- [15] The comparison of structured modeling and simulation modeling of queueing systems / Igor Yakimov [et al.] // Information Technologies and Mathematical Modelling. Queueing Theory and Applications: 16th International Conference, ITMM 2017, Named After AF Terpugov, Kazan, Russia, September 29–October 3, 2017, Proceedings 16 / Springer. — 2017. — P. 256–267.
- [16] Simulation using building blocks / Edwin Valentin [et al.] // Proceedings conference on AI, Simulation and Planning / Citeseer. — 2002. — P. 65–71.
- [17] Claes Peter. Controlling Fluid Simulations with Custom Fields in Houdini Master Thesis. — 2009.
- [18] Fowler Martin. Analysis patterns: reusable object models. — Addison-Wesley Professional, 1997.
- [19] Pilone Dan, Pitman Neil. UML 2.0 in a Nutshell. — "O'Reilly Media, Inc. 2005.

- [20] Herchi Hatem, Abdessalem Wahiba Ben. From user requirements to UML class diagram // arXiv preprint arXiv:1211.0713. — 2012.
- [21] Hallinan Jr Arthur J. A review of the Weibull distribution // Journal of Quality Technology. — 1993. — Vol. 25, no. 2. — P. 85–93.
- [22] Schmid Hubert. C++ 11: pure virtual functions with override and final. — 2012.
- [23] PyTorch. — <https://gcc.gnu.org/>. — Accessed: 2023-04-29.
- [24] Branco David, Henriques Pedro Rangel. Impact of GCC optimization levels in energy consumption during C/C++ program execution // 2015 IEEE 13th International Scientific Conference on Informatics / IEEE. — 2015. — P. 52–56.
- [25] Matsumoto Makoto, Nishimura Takuji. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator // ACM Transactions on Modeling and Computer Simulation (TOMACS). — 1998. — Vol. 8, no. 1. — P. 3–30.
- [26] Mersenne Twister Engine. — https://en.cppreference.com/w/cpp/numeric/random/mersenne_twister_engine. — Accessed: 2023-05-02.
- [27] The fundamental package for scientific computing with Python. — <https://github.com/numpy/numpy>. — Accessed: 2023-04-29.
- [28] Seamless operability between C++11 and Python. — <https://github.com/pybind/pybind11>. — Accessed: 2023-04-29.
- [29] PyTorch. — <https://pytorch.org/>. — Accessed: 2023-04-29.
- [30] Гнеденко Борис Владимирович. Курс теории вероятностей. — URSS, 2010.
- [31] Калинина Е.С., Кудрявцева М.В. О применении аппарата производящих функций в теории вероятностей // Современные тенденции развития науки и технологий. — 2016. — ¹ 3-1. — □. 18–22.
- [32] Ястребов И.П. Дискретизация непрерывных сигналов во времени. Теорема Котельникова // Нижний Новгород: Нижегородский государственный университет им. НИ Лобачевского. — 2012.

- [33] Кузнецов Н.А., Козякин В.С. Теорема Котельникова—основа цифрового оценивания и моделирования непрерывных процессов // Радиотехника. — 2008. — № 8.
- [34] Ряды Фурье / В.П. Долгополов [и др.]. — 2011.
- [35] Тимошенко Л.И. Дискретное преобразование Фурье и его быстрые алгоритмы // Современные наукоемкие технологии. — 2014. — № 12-2. — С. 188–193.
- [36] Bronson R. Matrix methods: An introduction. — Gulf Professional Publishing, 1991.
- [37] Вишневский ВМ, Дудин АН, Клименок ВИ. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях // М.: Техносфера. — 2018.
- [38] Artalejo Jesus R., Gomez-Corral A. Retrial Queueing Systems: A Computational Approach. — Springer-Verlag Berlin Heidelberg, 2008.
- [39] Falin G., Templeton J. Retrial queues. — CRC Press, 1997. — Vol. 75.
- [40] Егоров И. М. Программирование: Учебное методическое пособие. — Томск: Томский межвузовский центр дистанционного образования, 2006. — Р. 79.
- [41] Artalejo Jesus R. Accessible bibliography on retrial queues: progress in 2000–2009 // Mathematical and computer modelling. — 2010. — Vol. 51, no. 9-10. — P. 1071–1081.
- [42] Falin G. Model of coupled switching in presence of recurrent calls // Engineering Cybernetics. — 1979. — Vol. 17, no. 1. — P. 53–59.
- [43] Artalejo Jesus R., Resing JAC. Mean value analysis of single server retrial queues // Asia-Pacific Journal of Operational Research. — 2010. — Vol. 27, no. 03. — P. 335–345.
- [44] Erlang A. K. The theory of probabilities and telephone conversations // Nyt. Tidsskr. Mat. Ser. B. — 1909. — Vol. 20. — P. 33–39.

- [45] Bjornstad J. H. Traffic characteristics and queueing theory: implications and applications to web server systems : Master's thesis / J. H. Bjornstad. — 2006.
- [46] Fischer W., Meier-Hellstern K. The Markov-modulated Poisson process (MMPP) cookbook // Performance evaluation. — 1993. — Vol. 18, no. 2. — P. 149–171.
- [47] Meiert K. S. A Statistical procedure for fitting Markov-modulated Poisson process (applied probability). — 1986.
- [48] IEEE 802.3 az: the road to energy efficient ethernet / K. Christensen [et al.] // IEEE Communications Magazine. — 2010. — Vol. 48, no. 11. — P. 50–56.
- [49] Blagin A. L., Lapatin I. L. The Two-Dimensional Output Process of Retrial Queue with Two-Way Communication // Information Technologies and Mathematical Modelling. Queueing Theory and Applications: 19th International Conference, ITMM 2020, Named after AF Terpugov, Tomsk, Russia, December 2–5, 2020, Revised Selected Papers / Springer Nature. — P. 279–290.
- [50] Lapatin I. L., Nazarov A. A., Blagin A. L. Asymptotic analysis of the output process in retrial queue with two-way communication and MAP input. — 2020.
- [51] Kendall D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain // The Annals of Mathematical Statistics. — 1953. — P. 338–354.
- [52] Kulkarni V. G. On queueing systems with retrials // Journal of Applied Probability. — 1983. — P. 380–389.
- [53] Gharbi N., Dutheil C. An algorithmic approach for analysis of finite-source retrial systems with unreliable servers // Computers & Mathematics with Applications. — 2011. — Vol. 62, no. 6. — P. 2535–2546.
- [54] Nazarov A. A., Paul S., Gudkova I. Asymptotic analysis of Markovian retrial queue with two-way communication under low rate of retrials condition. — 2017.
- [55] Burke P. J. The output process of a stationary M/M/s queueing system // The Annals of Mathematical Statistics. — 1968. — Vol. 39, no. 4. — P. 1144–1152.

- [56] Paul S., Phung-Duc T. Retrial Queueing Model with Two-Way Communication, Unreliable Server and Resume of Interrupted Call for Cognitive Radio Networks // Information Technologies and Mathematical Modelling. Queueing Theory and Applications. — Springer, 2018. — P. 213–224.
- [57] Mirasol N. M. The output of an M/G/ ∞ queueing system is poisson // Operations Research. — 1963. — Vol. 11, no. 2. — P. 282–284.
- [58] Daley D. J. Queueing output processes // Advances in Applied Probability. — 1976. — Vol. 8, no. 2. — P. 395–415.
- [59] Wieland J. R., Pasupathy R., Schmeiser B. W. Queueing network simulation analysis: queueing-network stability: simulation-based checking // Proceedings of the 35th conference on Winter simulation: driving innovation. — 2003. — P. 520–527.
- [60] ПрЭВМ №2020664030. — https://new.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2020664929&TypeFile=html. — Accessed: 2021-06-02.
- [61] Благинин А., Лапатин И., Назаров А. Исследование двумерного выходящего потока марковской модели узла обработки запросов с повторными обращениями и вызываемыми заявками // Информационные технологии и математическое моделирование (ИТММ-2020) : материалы XIX Международной конференции имени А. Ф. Терпугова. — Томск: Изд-во НТЛ, 2021. — P. 159–165.
- [62] Lapatin I. L., Nazarov A. A., Blaginin A. L. Asymptotic Analysis of the Output Process in Retrial Queue with Two-Way Communication and MAP Input // Пятая Международная конференция по стохастическим методам (МКСМ-5) : материалы Международной научной конференции. — Изд-во РУДН, 2020. — P. 340–344.
- [63] Лапатин И. Л., Назаров А. А., Благинин А. Л. Исследование выходящего потока марковской модели узла обработки запросов с повторными обращениями и вызываемыми заявками // Математика, ее приложения и математическое образование (МПМО'20) : материалы VII Международной конференции. — Улан-Удэ: Изд-во ВСГУТУ, 2020. — P. 157–159.

- [64] Lauder Anthony, Kent Stuart. Precise visual specification of design patterns // European Conference on Object-Oriented Programming / Springer. — 1998. — P. 114–134.
- [65] Yang Gang, Yao Luo-gen, Ouyang Zi-sheng. The MAP/PH/N retrieval queue in a random environment // Acta Mathematicae Applicatae Sinica, English Series. — 2013. — Vol. 29, no. 4. — P. 725–738.
- [66] O’dwyer Arthur. Mastering the C++ 17 STL: Make full use of the standard library components in C++ 17. — Packt Publishing Ltd, 2017.
- [67] JSON: data model, query languages and schema specification / Pierre Bourhis [et al.] // Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems. — 2017. — P. 123–135.
- [68] Taylor Dave. Wicked Cool Shell Scripts: 101 Scripts for Linux, Mac OS X, and Unix Systems. — No Starch Press, 2004.
- [69] Hunt John. Multiprocessing // Advanced Guide to Python 3 Programming. — Springer, 2019. — P. 363–376.
- [70] Nussbaumer Henri J. The fast Fourier transform // Fast Fourier Transform and Convolution Algorithms. — Springer, 1981. — P. 80–111.
- [71] Bergland Glenn D. A guided tour of the fast Fourier transform // IEEE spectrum. — 1969. — Vol. 6, no. 7. — P. 41–52.
- [72] Array programming with NumPy / Charles R Harris [et al.] // Nature. — 2020. — Vol. 585, no. 7825. — P. 357–362.
- [73] Libbrecht Maxwell W, Noble William Stafford. Machine learning applications in genetics and genomics // Nature Reviews Genetics. — 2015. — Vol. 16, no. 6. — P. 321–332.
- [74] Shinde Pramila P, Shah Seema. A review of machine learning and deep learning applications // 2018 Fourth international conference on computing communication control and automation (ICCUBEA) / IEEE. — 2018. — P. 1–6.

- [75] Soofi Aized Amin, Awan Arshad. Classification techniques in machine learning: applications and issues // Journal of Basic and Applied Sciences. — 2017. — Vol. 13. — P. 459–465.
- [76] Learning Deep Generative Models for Queuing Systems / Cesar Ojeda [et al.] // Proceedings of the AAAI Conference on Artificial Intelligence. — Vol. 35. — 2021. — P. 9214–9222.
- [77] Scheduling data analytics work with performance guarantees: queuing and machine learning models in synergy / Ji Xue [et al.] // Cluster Computing. — 2016. — Vol. 19, no. 2. — P. 849–864.
- [78] Balla Husamelddin AMN, Sheng Chen Guang, Weipeng Jing. Reliability enhancement in cloud computing via optimized job scheduling implementing reinforcement learning algorithm and queuing theory // 2018 1st international conference on data intelligence and security (ICDIS) / IEEE. — 2018. — P. 127–130.
- [79] Cox Nicholas. PAIRPLOT: Stata module for plots of paired observations. — 2007.