

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

О.К. АЛЬСОВА

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СИСТЕМ

Учебное пособие

по дисциплине «Имитационное моделирование информационно-вычислительных
систем»

НОВОСИБИРСК, 2015

Работа подготовлена на кафедре вычислительной техники НГТУ

В пособии рассмотрены базовые определения и понятия теории моделирования, основные вопросы организации имитационного моделирования систем на компьютере, основные этапы технологии разработки имитационной модели системы, специализированные инструментальные средства моделирования, основы планирования и проведения машинных экспериментов с моделями систем, методы статистической обработки результатов моделирования на примере информационно-вычислительных систем. Также рассмотрены основные разделы теории Марковских случайных процессов, теории массового обслуживания в приложении к решению задач моделирования систем.

Основное внимание уделено темам, изучение которых необходимо для успешного выполнения лабораторных работ.

Пособие предназначено для бакалавров и магистрантов 1 курса АВТФ, обучающихся по направлениям «Информатика и вычислительная техника», «Программная инженерия».

Новосибирский государственный
технический университет, 2015

СОДЕРЖАНИЕ

1. Введение в теорию моделирования.....	5
1.1. Основные определения и понятия.....	6
1.2. Основные методы моделирования.....	8
1.3. Классификация видов моделирования.....	9
1.4. Математическая модель системы.....	12
1.5. Основные подходы к построению моделей системы.....	13
1.6. Контрольные вопросы и упражнения.....	17
2. Имитационное моделирование систем.....	19
2.1. Определение имитационного моделирования.....	19
2.2. Условия и проблемы применения имитационного моделирования.....	19
2.3. Технология разработки имитационной модели.....	21
2.4. Программное обеспечение имитационного моделирования систем.....	26
2.4.1. Классификация инструментальных средств моделирования.....	26
2.4.2. Возможности специализированных пакетов имитационного моделирования.....	28
2.4.2.1. Основные характеристики.....	28
2.4.2.2. Оборудование и программное обеспечение.....	31
2.4.2.3. Анимация и динамическая графика.....	32
2.4.2.4. Статистические возможности.....	34
2.4.2.5. Отчеты с выходными данными и графика.....	37
2.4.3. Специализированные средства имитационного моделирования систем ..	38
2.4.3.1. Язык моделирования GPSS.....	38
2.4.3.2. Расширенный редактор GPSS.....	42
2.4.3.3. Пакет имитационного моделирования ExtendSim.....	42
2.4.3.4. Другие специализированные средства имитационного моделирования	44
2.5. Принципы построения моделирующих алгоритмов.....	47
2.6. Методы генерации случайных последовательностей.....	56
2.6.1. Генерация случайных чисел.....	56
2.6.2. Генерация случайных событий.....	68
2.6.3. Имитация случайных величин.....	71
2.6.4. Имитация случайных векторов.....	81
2.6.5. Имитация случайных процессов.....	84
2.7. Метод статистического моделирования.....	87
2.8. Статистическая обработка результатов моделирования.....	89
2.9. Сравнение альтернативных конфигураций системы.....	98
2.10. Контрольные вопросы и упражнения.....	102
3. Отдельные разделы теории моделирования систем.....	103
3.1. Элементы теории Марковских случайных процессов, применяемые при моделировании систем.....	103
3.1.1. Потоки событий.....	103
3.1.2. Случайный процесс.....	109
3.1.3. Непрерывные цепи Маркова. Расчет вероятностей состояний.....	109
3.1.4. Марковский случайный процесс «гибели и размножения».	114

Расчет вероятностей состояний	114
3.2. Элементы теории массового обслуживания,.....	116
применяемые при моделировании систем	116
3.2.1. Основные определения и понятия	116
3.2.2. Типы систем массового обслуживания, описываемые	120
Марковским случайным процессом «гибели и размножения»	120
3.2.3. Расчет характеристик типовых систем массового обслуживания	121
3.2.4. Системы массового обслуживания	125
с «взаимопомощью» между каналами	125
3.2.5. Системы массового обслуживания с ошибками	127
3.3. Контрольные вопросы и упражнения	128
4. Планирование машинных экспериментов с моделями систем	129
4.1. Основные понятия теории планирования экспериментов	131
4.2. Этапы планирования и проведения эксперимента	132
4.3. Контрольные вопросы и упражнения	147
5. Список литературы	148

1. Введение в теорию моделирования

Моделирование (в широком смысле) – это основной метод исследования во всех областях знаний. Методы моделирования используются для оценок характеристик сложных систем и принятия научно обоснованных решений в разных сферах человеческой деятельности. Существующую или проектируемую систему можно эффективно исследовать с помощью математических моделей (аналитических и имитационных) с целью оптимизации процесса функционирования системы. Модель системы реализуется на современных компьютерах, которые в этом случае выступают в качестве инструмента исследователя с моделью системы.

В настоящее время моделирование широко используется в сфере управления различными системами, где основными являются процессы принятия решений на основе получаемой информации. Также методы моделирования находят применение при исследовании, проектировании, внедрении информационно-вычислительных систем (ИВС) и автоматизированных систем управления (АСУ).

Выбор метода моделирования и необходимая детализация проекта зависят от этапа разработки системы. На этапах обследования объекта управления (например, промышленного предприятия) и разработки технического задания на проектирование ИВС, АСУ модели носят описательный характер и преследуют цель наиболее полно представить в компактной форме информацию об объекте, необходимую разработчику системы.

На этапе разработки технического проекта ИВС, АСУ моделирование служит для решения задачи проектирования, т.е. выбора оптимального варианта по определенному критерию при заданных ограничениях из множества допустимых (построение оптимизационных моделей).

На этапе внедрения и эксплуатации ИВС, АСУ строятся модели для проигрывания возможных ситуаций для принятия обоснованных и перспективных решений по управлению объектом. Моделирование (имитацию) также широко применяют при обучении и тренировке персонала (деловые игры).

1.1. Основные определения и понятия

Приведем основные определения и понятия, используемые в теории моделирования.

Теория моделирования – теория замещения объекта-оригинала его моделью и исследования свойств объекта на его модели [1].

Моделирование – метод исследования, основанный на замене исследуемого объекта-оригинала его моделью и на работе с ней (вместо объекта) [1, 18].

Модель (объекта - оригинала) (от лат. *modus* – «мера», «объем», «образ») – вспомогательный объект, отражающий наиболее существенные для исследования закономерности, суть, свойства, особенности строения и функционирования объекта-оригинала [18].

Когда говорят о моделировании, обычно имеют в виду моделирование некоторой системы.

Система – совокупность взаимосвязанных элементов, объединенных для реализации общей цели, обособленная от окружающей среды и взаимодействующая с ней как целостное целое и проявляющая при этом основные системные свойства. В [18] выделено 15 основных системных свойств, к которым относятся: эмергентность (эмерджентность); цельность; структурированность; целостность; подчиненность цели; иерархичность; бесконечность; эргатичность; открытость; необратимость; единство структурной устойчивости и неустойчивости; нелинейность; потенциальная многовариантность актуальных структур; критичность; непредсказуемость в критической области.

В курсе рассматриваются вопросы, связанные с моделированием систем массового обслуживания (СМО), на примере информационно-вычислительных систем.

Понятие СМО охватывает широкий круг систем. *СМО* – любая система по обслуживанию заявок в условиях ограничения на ресурсы системы. Примеры: библиотека, вычислительный центр, система передачи данных, информационная система, автоматизированная система управления, аэропорт, производственный

участок, АЗС и т.д. Любая информационно-вычислительная система может быть представлена, формализовано описана и исследована, как СМО.

Цели моделирования СМО:

1. Оценить эффективность функционирования системы, т.е. ответить на вопросы: как долго ждут заявки в очереди на обслуживания, справляются ли каналы обслуживания с потоком заявок, время пребывания заявки в системе и т.п.
2. Оптимизировать выходные характеристики системы, т.е. подобрать такие параметры системы, при которых система функционирует оптимально.

Исследоваться могут как существующие системы, так и системы, которые только планируется реализовать.

Методы теории моделирования инвариантны к предметной области, т.е. могут применяться для моделирования систем любого типа и широко используются на практике. Например, на этапе проектирования информационно-поисковой системы хорошо бы сначала подумать какой поток заявок, и какого типа будет система обслуживать, какова будет скорость выполнения запросов, скорость передачи по каналам связи и т.п.

При моделировании систем используют два подхода: классический (индуктивный), сложившийся исторически первым, и системный [1, 18].

При *классическом подходе* реальный объект, подлежащий моделированию, разбивается на подсистемы, выбираются исходные данные для моделирования и ставятся цели, отражающие отдельные стороны процесса моделирования. По отдельной совокупности исходных данных ставится цель моделирования отдельной стороны функционирования системы, на базе этой цели формируется некоторая компонента будущей модели. Совокупность компонент объединяется в модель. Таким образом, происходит суммирование компонент, каждая компонента решает свои собственные задачи и изолирована от других частей модели. Применим подход только для простых систем, где можно не учитывать взаимосвязи между элементами (компонентами) системы.

Системный подход – методологическая концепция, основанная на стремлении построить целостную картину изучаемого объекта с учетом важных для решаемой задачи элементов объекта, связей между ними и внешних связей с другими объектами и окружающей средой. В основе системного подхода лежит рассмотрение системы как интегрированного целого, причем это рассмотрение при разработке начинается с главного – формулировки цели функционирования. Важным для системного подхода является определение структуры системы – совокупности связей между элементами системы, отражающих их взаимодействие.

Существуют структурные и функциональные подходы к исследованию структуры системы и ее свойств. При структурном подходе выявляются состав выделенных элементов системы и связи между ними. При функциональном подходе рассматриваются алгоритмы поведения системы (функции – свойства, приводящие к достижению цели).

1.2. Основные методы моделирования

Методы моделирования можно условно классифицировать на три основные группы: аналитические, численные и имитационные [1, 19].

1. Аналитические методы моделирования. Аналитические методы позволяют получить характеристики системы как некоторые функции параметров ее функционирования. Таким образом, аналитическая модель представляет собой систему уравнений, при решении которой получают параметры, необходимые для расчета выходных характеристик системы (среднее время обработки задания, пропускную способность и т.д.). Аналитические методы дают точные значения характеристик системы, но применяются для решения только узкого класса задач. Причины этого заключается в следующем. Во-первых, вследствие сложности большинства реальных систем их законченное математическое описание (модель) либо не существует, либо еще не разработаны аналитические методы решения созданной математической модели. Во-вторых, при выводе формул, на которых основываются аналитические методы, принимаются определенные допущения, ко-

торые не всегда соответствуют реальной системе. В этом случае от применения аналитических методов приходится отказываться. Подробно аналитические методы моделирования СМО рассмотрены в разделе 3 учебного пособия.

2. Численные методы моделирования. Численные методы предполагают преобразование модели к уравнениям, решение которых возможно методами вычислительной математики. Класс задач, решаемых этими методами, значительно шире. В результате применения численных методов получают приближенные значения (оценки) выходных характеристик системы с заданной точностью.

3. Имитационные методы моделирования. С развитием вычислительной техники широкое применение получили имитационные методы моделирования для анализа систем, преобладающими в которых являются стохастические воздействия.

Суть имитационного моделирования (ИМ) заключается в имитации процесса функционирования системы во времени, с соблюдением таких же соотношений длительности операций как в системе оригинале. При этом имитируются элементарные явления, составляющие процесс, сохраняется их логическая структура, последовательность протекания во времени. В результате применения ИМ получают оценки выходных характеристик системы, которые необходимы при решении задач анализа, управления и проектирования систем.

В основе имитационного моделирования лежит разработка и выполнение на компьютере программы, отражающей структуру и функционирование (поведение) моделируемого объекта или процесса во времени [21]. Программа представляет имитационную модель этого объекта (системы, процесса), выполнение программы можно считать имитацией поведения исходной системы во времени. Имитационная модель – это упрощенное подобие реальной системы, либо существующей, либо той, которую предполагается создать в будущем. Вопросы, связанные с применением на практике имитационных методов моделированием систем, рассмотрены в разделе 2 учебного пособия.

1.3. Классификация видов моделирования

В основу классификации видов моделирования можно положить различные признаки [1, 18, 19]. В зависимости от характера изучаемых процессов в системе моделирование может быть разделено на детерминированное и стохастическое; статическое и динамическое; дискретное и непрерывное.

Детерминированное моделирование применяется для исследования систем, поведение которых можно абсолютно точно предвидеть. Например, путь, пройденный автомобилем, при равноускоренном движении в идеальных условиях; устройство, возводящее в квадрат число и т.п. Соответственно в этих системах протекает детерминированный процесс, который адекватно описывается детерминированной моделью.

Стохастическое (теоретико-вероятностное) моделирование применяется для исследования системы, состояние которой зависит не только от контролируемых, но и от неконтролируемых воздействий или в ней самой есть источник случайности. К стохастическим системам относятся все системы, которые включают человека, например, заводы, аэропорты, информационно-вычислительные системы и сети, магазины, предприятия бытового обслуживания и т.п. Стохастические модели учитывают вероятностный характер параметров моделируемого объекта. Например, в модели работы вычислительного кластера не могут быть точно определены моменты поступления заявок на обработку. Эти моменты являются случайными величинами, поэтому модель является стохастической. Анализ подобных моделей, как правило, выполняется на компьютере на основе статистики, набираемой в ходе имитационных экспериментов при многократном прогоне моделей для различных значений исходных случайных величин, выбранных в соответствии с их статистическими характеристиками.

Статистическое моделирование служит для описания систем в какой-либо момент времени. *Динамическое моделирование* отражает изменение системы во времени (выходные характеристики системы в данный момент времени определяются характером входных воздействий в прошлом и настоящем). Примером ди-

намических систем являются биологические, экономические, социальные системы; такие искусственные системы как завод, предприятие, поточная линия и т.п.

Дискретное моделирование применяют для исследования систем, в которых входные и выходные характеристики измеряется или изменяется во времени дискретно, в противном случае применяют *непрерывное моделирование*. Например, электронные часы, электросчетчик – дискретные системы; солнечные часы, нагревательные приборы – непрерывные системы.

В зависимости от формы представления объекта (системы) можно выделить *мысленное и реальное моделирование*.

При *реальном (натурном) моделировании* исследование характеристик системы проводится на реальном объекте, либо на его части. Реальное моделирование – наиболее адекватно, но его возможности, с учетом особенностей реальных объектов, ограничены. Например, проведение реального моделирования с АСУ предприятия требует, во-первых, создания АСУ; во-вторых, проведения экспериментов с предприятием, что невозможно. К реальному моделированию относят производственный эксперимент и комплексные испытания, которые обладают высокой степенью достоверности. Другой вид реального моделирования – *физическое*. При физическом моделировании исследование проводится на установках, которые сохраняют природу явления и обладают физическим подобием.

Мысленное моделирование применяется для моделирования систем, которые практически не реализуемы на заданном интервале времени. В основе мысленного моделирования лежит создание идеальной модели, основанной на идеальной, мыслительной аналогии. Различают два вида мысленного моделирования: *образное (наглядное) и знаковое*.

При *образном моделировании* на базе представлений человека о реальных объектах создаются различные наглядные модели, отображающие явления и процессы, протекающие в объекте. Например, модели частиц газов в кинетической теории газов в виде упругих шаров, воздействующих друг на друга во время столкновения.

При *знаковом моделировании* описывают моделируемую систему с помощью условных знаков, символов, в частности, в виде математических, физических и химических формул. Наиболее мощный и развитый класс знаковых моделей представляют математические модели.

Математическая модель – это искусственно созданный объект в виде математических, знаковых формул, который отображает и воспроизводит структуру, свойства, взаимосвязи и отношения между элементами исследуемого объекта [1, 18]. Далее рассматриваются только математические модели и соответственно математическое моделирование.

Математическое моделирование – метод исследования, основанный на замене исследуемого объекта-оригинала его математической моделью и на работе с ней (вместо объекта) [18]. Математическое моделирование можно разделить на аналитическое (АМ), имитационное (ИМ), комбинированное (КМ).

При АМ создается аналитическая модель объекта в виде алгебраических, дифференциальных, конечно-разностных уравнений. Аналитическая модель исследуется либо аналитическими методами, либо численными методами.

При ИМ создается имитационная модель, используется метод статистического моделирования для реализации имитационной модели на компьютере.

При КМ проводится декомпозиция процесса функционирования системы на подпроцессы. Для тех из них, где это возможно, используют аналитические методы, в противном случае – имитационные.

1.4. Математическая модель системы

Важный этап моделирования – создание математической модели исследуемой системы. На базе математической модели проводится анализ характеристик системы. При компьютерном моделировании с помощью математической модели создается алгоритм программы имитации. Формальное описание объекта исследования необходимо также для взаимопонимания между специалистами разных областей, объединенных для решения какой-либо глобальной задачи.

В общем случае математическую модель любой динамической системы можно представить в следующем виде:

$$\vec{y}(t) = F(\vec{x}, \vec{h}, t), \quad (1.1)$$

где $\vec{x}$ – совокупность входных воздействий на систему;

$\vec{h}$ – совокупность внутренних параметров системы;

$\vec{y}$ – совокупность выходных характеристик системы;

F – закон функционирования системы.

Процесс функционирования системы можно рассматривать также как последовательную смену состояний: $z_1(t), z_2(t), \dots, z_k(t)$:

$$\vec{z}(t) = G(\vec{z}^0, \vec{x}, \vec{h}, t), \quad (1.2)$$

где $\vec{z}^0$ – совокупность начальных состояний.

Таким образом, общую математическую модель системы можно также представить следующим образом:

$$\vec{y}(t) = F(\vec{z}(t)). \quad (1.3)$$

1.5. Основные подходы к построению моделей системы

При построении математических моделей процессов функционирования систем можно выделить следующие основные подходы: непрерывно-детерминированный, дискретно-детерминированный, дискретно-стохастический, непрерывно-стохастический, сетевой и обобщенный [1, 19]. Соответственно этим подходам были разработаны типовые математические схемы создания моделей.

1. Непрерывно-детерминированный подход (D – схемы) основан на использовании систем дифференциальных уравнений в качестве математических моделей. Созданные на основе этого подхода модели исследуются, как правило, аналитическим способом. Одно из приложений непрерывно-детерминированного подхода – анализ систем автоматического управления непрерывными процессами, например, системы управления температурой печи. Система моделируется в течение непрерывного интервала времени t .

2. Дискретно-детерминированный подход (F – схемы) реализуется с помощью математического аппарата теории автоматов. Система представляется в виде автомата, перерабатывающего дискретную информацию и меняющего свои внутренние состояния лишь в допустимые моменты времени. Математическая модель при этом подходе – конечный автомат, характеризующийся конечным множеством X входных сигналов, конечным множеством Y выходных сигналов, конечным множеством Z внутренних состояний, начальным состоянием $Z_0 \in Z$; функцией переходов $g(z,x)$; функцией выходов $v(z,x)$. Автомат функционирует в дискретном автоматном времени, моменты которого – такты (примыкающие друг к другу равные интервалы времени, каждому из которых соответствуют постоянные значения входного и выходного сигналов и внутренние состояния). Задается конечный F – автомат таблицей переходов и выходов, либо с помощью графа.

Дискретно-детерминированный подход применяется для описания широкого класса процессов функционирования реальных объектов в АСУ. Примеры таких объектов – узлы компьютера, устройства контроля, регулирования и управления, системы временной и пространственной коммутации в технике связи и т.д.

3. Дискретно-стохастический подход (P – схемы) использует в качестве математического аппарата вероятностные автоматы, которые можно определить, как дискретные потактные преобразователи информации с памятью, функционирование которых в каждом такте зависит только от состояния памяти в них и может быть описано статистически. Для такого автомата характерно задание таблицы вероятностей перехода автомата в некоторое состояние и появления некоторого выходного сигнала в зависимости от текущего состояния и входного сигнала. Исследование автомата может проводиться как аналитическими, так и имитационными методами. Этот подход применим для изучения эксплуатационных характеристик производственных объектов (например, надежности, ремонтпригодности, отказоустойчивости и т.п.).

4. Непрерывно-стохастический подход (Q – схема) применяется для формализации процессов обслуживания [1-2]. Этот подход наиболее известен ввиду того, что большинство производственных (экономических, технических и т.д.) си-

систем по своей сути – это системы массового обслуживания. Типовая математическая схема моделирования таких систем – Q – схема. В любой системе массового обслуживания можно выделить элементарный прибор. Соответственно в этом приборе выделяют некоторой емкости накопитель заявок, ожидающих обслуживания; канал обслуживания; потоки событий: поток заявок на обслуживание, характеризующийся моментами времени поступления и атрибутами (признаками) заявок (например, приоритетами), и поток обслуживания, характеризующийся моментами начала и окончания обслуживания заявок (рисунок 1.1, 1.2). Под непрерывностью в данном случае имеется в виду непрерывность по времени: система моделируется в течение непрерывного интервала времени t , моменты поступления заявки в систему и окончания обслуживания заявки – случайны.

Для исследования систем массового обслуживания применяются имитационные методы (раздел 2) и аналитические методы (раздел 3).

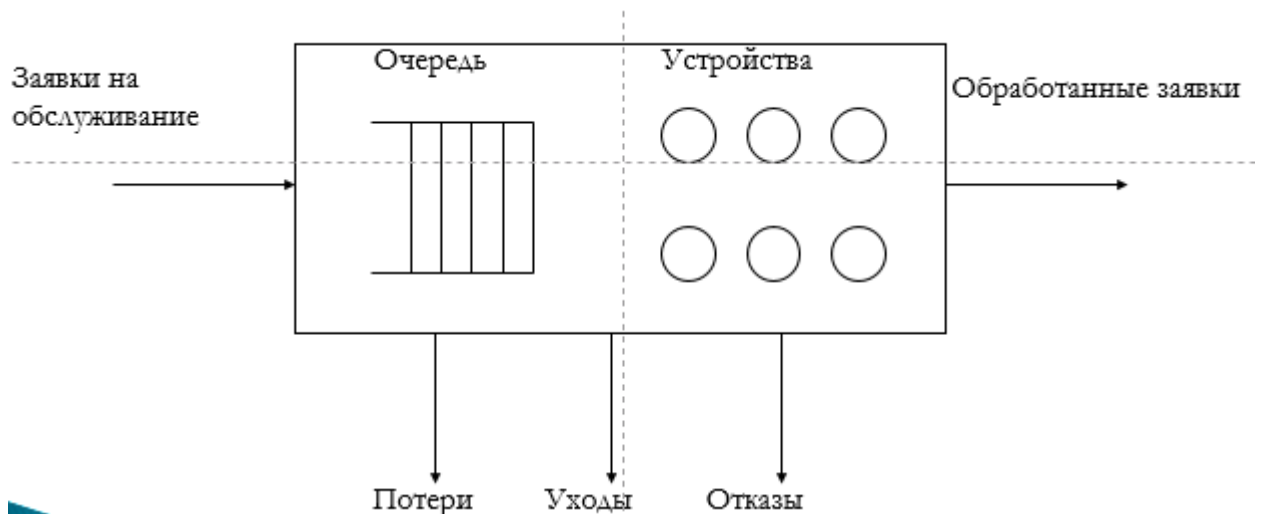


Рисунок 1.1 – Структурная схема СМО

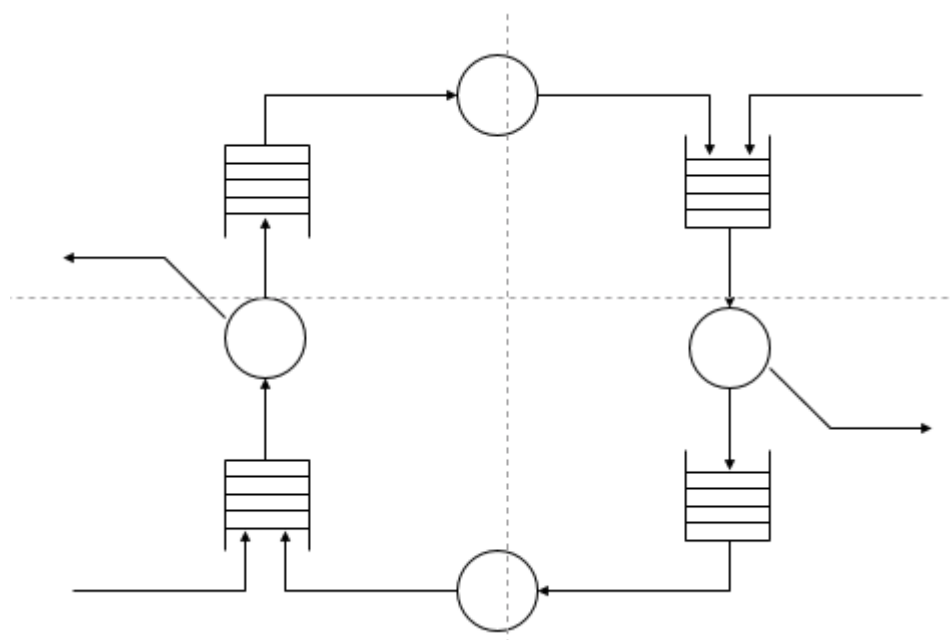


Рисунок 1.2. – Структура Q-схемы

За имитационным моделированием систем массового обслуживания исторически закрепился термин «дискретно-событийное моделирование» [21]. Система называется дискретно-событийной если изменение состояний в ней происходит под влиянием явно определенных (дискретных) событий. Находясь в некотором состоянии, система сохраняет его (не изменяет своих характеристик) до наступления очередного события, под воздействием которого переменные системы (и, следовательно, ее состояние) изменяются скачком. Например, при построении модели вычислительной системы состояние системы может быть представлено количеством заявок в системе, числом занятых компьютеров. Состояние системы изменяется, если новая заявка поступает в систему или, когда освобождается компьютер. Приход заявки или окончание обслуживания заявки – событие, которое изменяет состояние системы.

5. Сетевой подход (N – схема) используется для формализованного описания и анализа причинно-следственных связей в сложных системах, где одновременно протекает несколько процессов. Самым распространенным в настоящее время формализмом, описывающим структуру и взаимодействие параллельных систем и процессов, являются сети Петри.

6. Обобщенный подход (A – схема) применяется для описания любых видов систем и базируется на понятии агрегативной системы, представляющей собой формальную схему общего вида (A – схему). При агрегативном описании сложная система разбивается на конечное число частей (подсистем), с сохранением связей, обеспечивающих их взаимодействие. В свою очередь подсистемы также разбиваются на части. Процесс разбиения продолжается до тех пор, пока не образуются подсистемы, которые в условиях рассматриваемой задачи моделирования удобны для математического описания. В результате такой декомпозиции сложная система представляется в виде многоуровневой конструкции из взаимосвязанных элементов, объединенных в подсистемы различных уровней.

В рамках курса изучаются методы и инструментальные средства моделирования систем на основе непрерывно-стохастического подхода.

1.6. Контрольные вопросы и упражнения

1. Как определяются понятия «теория моделирования»; «система»; «модель системы»; «математическое моделирование»?

2. В каких областях знаний используется математическое моделирование, какие задачи решаются на его основе?

3. Какие методы применяются при математическом моделировании системы? Приведите примеры систем, для моделирования которых используются эти методы.

4. Какие существуют классификационные признаки видов моделирования систем? Приведите примеры.

5. Какие основные подходы используются при построении математических моделей системы? Каковы условия использования и особенности каждого подхода? Приведите примеры.

6. При проектировании информационной системы возможны два подхода.

При первом подходе проектирование реализуется в виде цепочки принятия проектных решений (рисунок 1.3): разработка технического задания (ТЗ), систе-

мотехническое проектирование (СП), схмотехническое проектирование (СХП), конструкторское проектирование (КП), технологическая проработка (ТП).

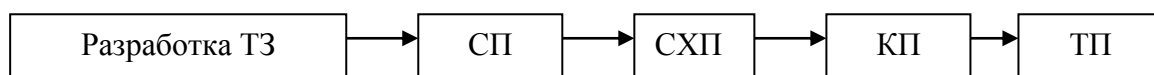


Рисунок 1.3 – Первый подход к проектированию информационной системы

При втором подходе проектирование опирается на критерий (или их совокупность), отражающий цель проектирования. Процесс проектирования показан на рисунке 1.4.

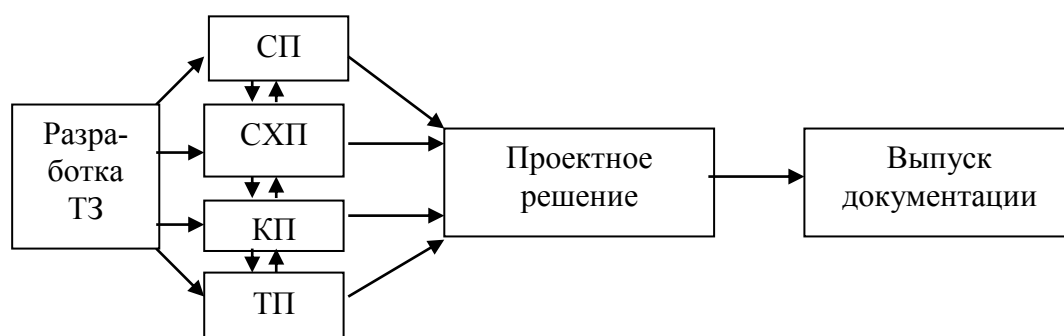


Рисунок 1.4 – Второй подход к проектированию информационной системы

Стрелками показаны возможные пути передачи информации. Какая из двух технологий проектирования более современна? В какой из технологий (или в обеих) осуществлен системный подход? Почему?

7. Опишите наиболее эффективный, с вашей точки зрения, вид моделирования каждой из перечисленных систем и объясните свой ответ: небольшой производственный участок существующей фабрики; устройство, возводящее входной сигнал в квадрат; пункт скорой помощи в существующей больнице; проектируемая АСУ предприятия; функционирующая локальная сеть в вычислительном центре. Какие цели могут быть поставлены при моделировании этих систем?

8. Предположим, что решено изучить все системы п. 7 посредством математической модели. Каким должно быть моделирование – статическим или динамическим, детерминированным или стохастическим, непрерывным или дискретным? Какой вид математического моделирования должен быть использован?

2. Имитационное моделирование систем

2.1. Определение имитационного моделирования

Известный американский ученый Роберт Шеннон дает следующее определение [29]: «Имитационное моделирование – процесс конструирования модели реальной системы и постановки экспериментов на этой модели с целью либо понять поведение системы, либо оценить (в рамках ограничений, накладываемых некоторым критерием или совокупностью критериев) различные стратегии, обеспечивающие функционирование данной системы». Все имитационные модели используют принцип черного ящика. Это означает, что они выдают выходной сигнал системы при поступлении в нее некоторого входного сигнала. Поэтому в отличие от аналитических моделей для получения результатов необходимо осуществлять «прогон» имитационных моделей, т.е. подачу некоторой последовательности сигналов, объектов или данных на вход модели и фиксацию выходной информации, а не «решать» их. Происходит своего рода «выборка» состояний объекта моделирования (состояния – свойства системы в конкретные моменты времени) из пространства (множества) состояний (совокупность всех возможных значений состояний). Насколько репрезентативной окажется эта выборка, настолько результаты моделирования будут соответствовать действительности. Этот вывод показывает важность статистических методов оценки результатов имитации.

2.2. Условия и проблемы применения имитационного моделирования

Применение имитационного моделирования целесообразно при выполнении определенных условий [29]:

1. Не существует законченной математической постановки данной задачи, либо еще не разработаны аналитические методы решения сформулированной ма-

тематической модели. К этой категории относятся многие модели массового обслуживания, связанные с рассмотрением очередей.

2. Аналитические методы имеются, но математические процедуры столь сложны и трудоемки, что имитационное моделирование дает более простой способ решения задачи.

3. Кроме оценки определенных характеристик системы, необходимо осуществить на имитационной модели наблюдение за ходом процесса в течение определенного периода.

4. Имитационное моделирование может быть единственно возможным вследствие трудности постановки эксперимента и наблюдения явления в реальных условиях (например, наблюдение движения космического корабля).

5. Может понадобиться сжатие шкалы времени (как замедление, так и ускорение; например, исследование проблем развития и упадка городов).

Дополнительным преимуществом ИМ можно считать широкие возможности его применения в сфере образования и профессиональной подготовки. Использование имитационной модели позволяет экспериментатору видеть и «разыгрывать» на модели реальные процессы и ситуации.

Во многих случаях имитационное моделирование – это единственный способ получить представление о поведении сложной системы и провести ее анализ [21].

Необходимо обозначить ряд проблем, возникающих в процессе имитационного моделирования системы.

Первая проблема, которая касается и аналитических методов моделирования, состоит в нахождении «золотой середины» между простотой и точностью модели. Если модель слишком упрощена и в ней не учтены некоторые существенные факторы, то высока вероятность получения недостоверных результатов. С другой стороны, если модель сложная и в нее включены факторы, имеющие незначительное влияние на изучаемую систему, то резко повышаются затраты на создание такой модели и возрастает риск ошибки в логической структуре модели. Поэтому перед созданием модели необходимо проделать большой объем работы

по анализу структуры системы и взаимосвязей между ее элементами, изучению совокупности входных воздействий, тщательной обработке имеющихся статистических данных об исследуемой системе.

Вторая проблема заключается в искусственном воспроизведении случайных воздействий окружающей среды. Этот вопрос очень важен, так как большинство реальных систем – стохастические, и при их моделировании необходимы несмещенные случайные последовательности, в противном случае, результаты, полученные на модели, могут не соответствовать действительности. Существует два основных направления разрешения этой проблемы: аппаратная и программная (псевдослучайная) генерация случайных последовательностей (п. 2.6).

Третья наиболее сложная проблема – оценка качества модели и полученных с ее помощью результатов (эта проблема актуальна и для аналитических методов). Адекватность модели может быть оценена по полученным результатам методом экспертных оценок, сравнением с другими моделями (уже подтвердившими свою достоверность). В свою очередь, для проверки полученных результатов проводится их сравнение с уже имеющимися данными.

2.3. Технология разработки имитационной модели

Процесс моделирования системы можно укрупнено представить в виде последовательного выполнения трех основных этапов: построение концептуальной модели системы и ее формализация; алгоритмизация и машинная реализация модели системы; получение и интерпретация результатов моделирования системы. На рисунке 2.1 приведена схема технологического процесса имитационного моделирования объекта (системы), составленная на основе [1, 2, 18].

На первом этапе «Построение концептуальной модели системы и ее формализация» анализируется и описывается объект моделирования, строится концептуальная модель системы, определяющая цели, задачи, исходные данные, стратегии, условия, показатели качества моделирования на предметно – содержательном уровне.

Концептуальная модель (КМ) – это абстрактная модель, определяющая состав и структуру системы, свойства элементов и причинно-следственные связи, присущие анализируемой системе и существенные для достижения целей моделирования. В таких моделях обычно в словесной форме приводятся сведения о природе и параметрах (характеристиках) элементарных явлений исследуемой системы, о виде и степени взаимодействия между ними, о месте и значении каждого элементарного явления в общем процессе функционирования системы. При создании КМ практически параллельно формируется область исходных данных (информационное пространство системы) – этап подготовки исходных данных. На данном этапе выявляются количественные характеристики (параметры) функционирования системы и ее элементов, численные значения которых составят исходные данные для моделирования. Очевидно, что значительная часть параметров системы – это случайные величины. Поэтому особое значение при формировании исходных данных имеют выбор законов распределения случайных величин, аппроксимация функций и т.д.

Затем формализуется концептуальная модель, т.е. реализуется переход от содержательного описания объекта к его математической модели, и составляется техническая документация по первому этапу моделирования.

На втором этапе «Алгоритмизация и машинная реализация модели системы» математическая модель, разработанная на первом этапе, воплощается в конкретную машинную (программную модель). На данном этапе выбираются методы генераций случайных воздействий на объект моделирования и принцип построения моделирующего алгоритма (δz или Δt), который далее представляется в виде логической схемы (блок-схемы) процесса функционирования исследуемой системы. Логическая схема в свою очередь служит основой для разработки и проверки достоверности программной модели системы. Одновременно с построением логической схемы модели получают, если это возможно, математические соотношения в виде явных функций, т.е. строят аналитические модели. Таким образом, в общем случае построенная машинная модель системы будет иметь комбиниро-

ванный характер, т.е. отражать аналитико-имитационный подход к моделированию системы.

На данном этапе также происходит уточнение или выбор программного пакета моделирования с учетом ряда критериев. Непременное условие при этом – достаточность и полнота средств для реализации концептуальной модели. Среди других критериев можно назвать доступность, простоту и легкость освоения, скорость и корректность создания программной модели. После выбора среды моделирования концептуальная модель, сформулированная на предыдущем этапе, воплощается в компьютерную модель, т.е. решается задача алгоритмизации и детализации модели. Завершается этап составлением технической документации.

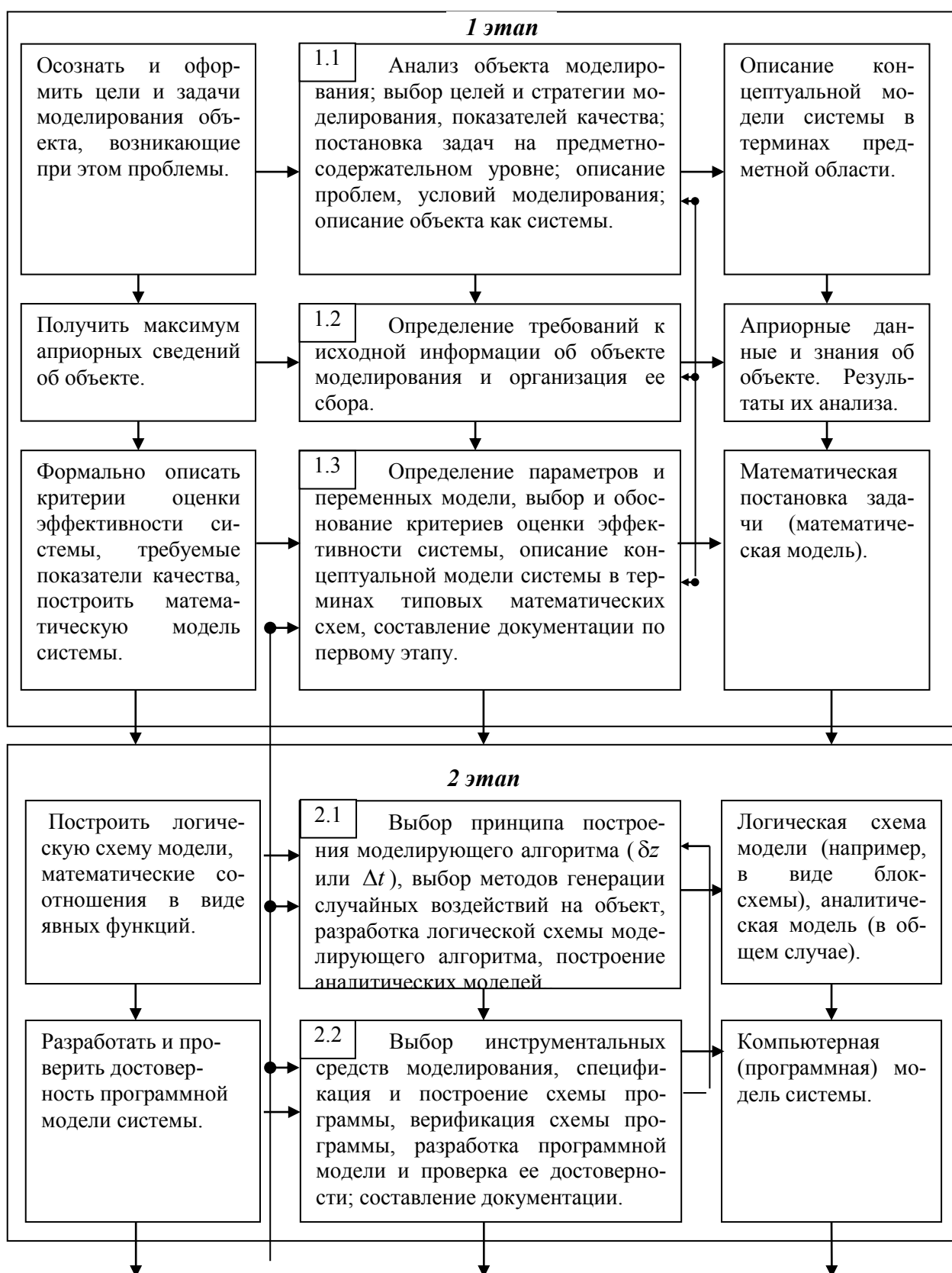


Рисунок 2.1 – Укрупненная схема технологического процесса имитационного моделирования объекта (системы)

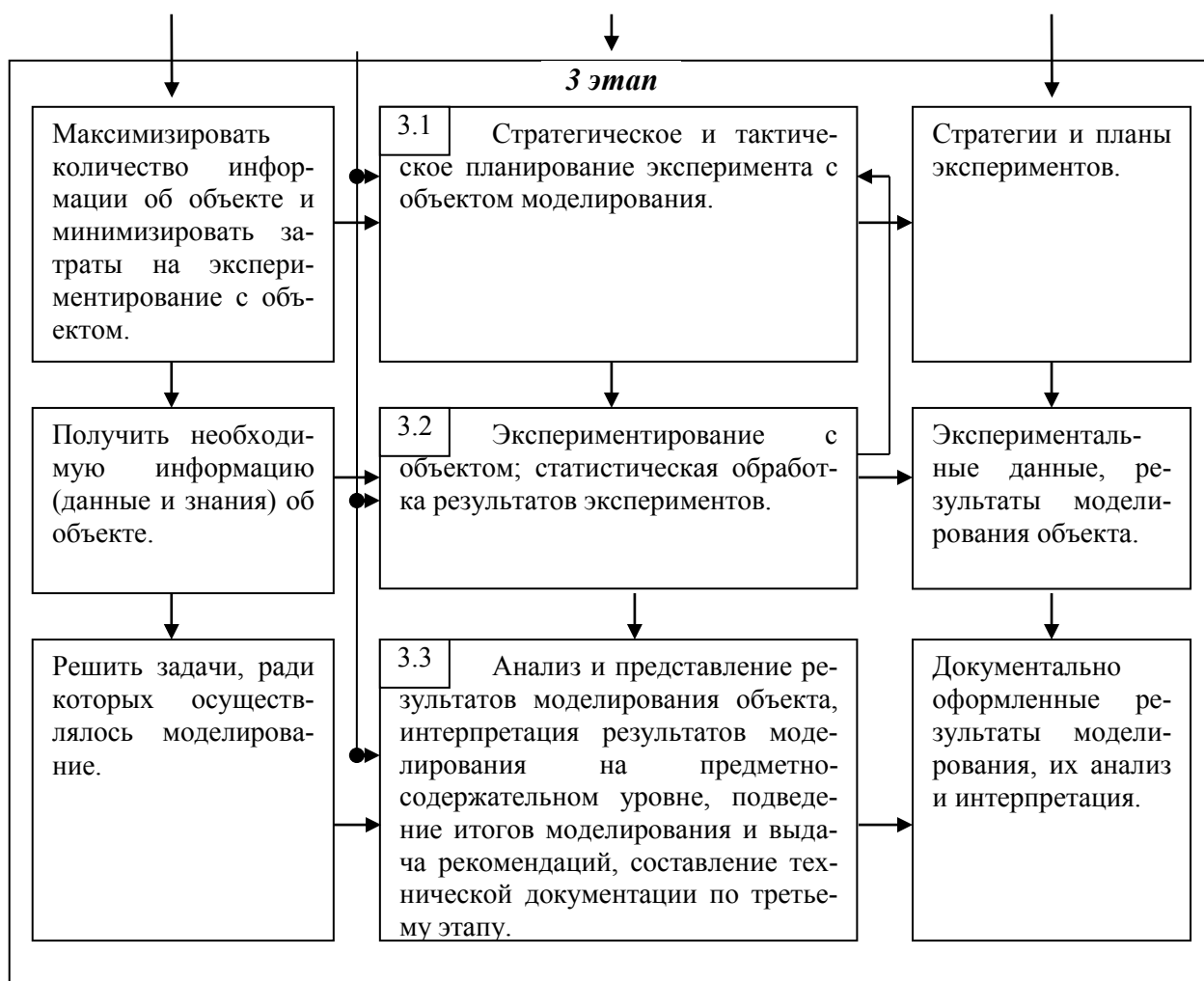


Рисунок 2.1 (продолжение) – Укрупненная схема технологического процесса имитационного моделирования объекта (системы)

На третьем этапе «Получение и интерпретация результатов моделирования системы» планируются и проводятся машинные эксперименты с моделью системы, статистически обрабатываются полученные экспериментальные данные. Далее анализируются и интерпретируются полученные результаты, подводятся итоги моделирования, выдаются рекомендации по поводу использования разработанной модели и решения поставленных задач моделирования, составляется техническая документация.

Далее подробно рассматривается содержание каждого из этапов.

2.4. Программное обеспечение имитационного моделирования систем

2.4.1. Классификация инструментальных средств моделирования

В основе имитационного моделирования – разработка компьютерной программы, реализующей процесс функционирования исследуемой системы или объекта во времени. Одно из наиболее важных решений, которые приходится принимать разработчику имитационной модели, касается выбора *программного* обеспечения. Если программное обеспечение недостаточно гибко или с ним сложно работать, то имитационное моделирование может дать неправильные результаты или оказаться вообще невыполнимым. Компьютерная программа имитации может быть разработана с помощью различных инструментальных средств. Приведем их основные типы [3]:

- универсальные языки программирования (C, C++, Java, Паскаль и т.п.);
- специализированные языки и среды имитационного моделирования (например, GPSS);
- специализированные визуальные среды имитационного моделирования (Arena, ExtendSim, AnyLogic, ...);
- специализированные объектно-ориентированные среды имитационного моделирования.

Для эффективной разработки имитационной модели системы важно, чтобы инструментальное средство обладало рядом функциональных возможностей. К таким возможностям относятся:

- генерирование случайных чисел, то есть последовательности значений, равномерно распределенных в интервале $[0,1]$;
- генерирование последовательности значений случайных величин с заданным законом распределения вероятностей (например, экспоненциальное распределение);
- продвижение модельного времени;
- определение следующего события по списку событий и передача управления соответствующему блоку кода;

- добавление записей в список или удаление записей из списка;
- сбор выходных статистических данных и создание отчета с результатами;
- определение сбойных ситуаций.

В сущности, именно наличие этих и некоторых других общих функциональных возможностей в моделирующих программах привело к разработке специальных программных пакетов имитационного моделирования.

Использование специализированного пакета имитационного моделирования в сравнении с применением универсального языка программирования дает несколько преимуществ.

1. Пакеты имитационного моделирования автоматически предоставляют большинство функциональных возможностей, требующихся для создания имитационной модели, что позволяет существенно сократить время, необходимое для программирования, и общую стоимость проекта.
2. Пакеты имитационного моделирования обеспечивают естественную среду для создания имитационных моделей. Их основные моделирующие конструкции больше подходят для имитационного моделирования, чем соответствующие конструкции в универсальных языках программирования, таких как С.
3. Имитационные модели, которые созданы с помощью пакетов моделирования, как правило, проще модифицировать и использовать.
4. Пакеты имитационного моделирования обеспечивают более совершенные механизмы обнаружения ошибок, поскольку они выполняют автоматический поиск ошибок многих типов. И так как модель не требует большого числа структурных компонентов, уменьшаются шансы совершить какую-либо ошибку.

Тем не менее, часто для разработки имитационных моделей по-прежнему используют универсальные языки программирования, которые тоже обладают некоторыми преимуществами.

1. Языки программирования знает большинство разработчиков, чего нельзя сказать о пакетах имитационного моделирования.

2. Скорость выполнения прогона имитационных моделей, написанных на языках С или С++, обычно выше, чем моделей, созданных с помощью пакетов имитационного моделирования (это связано с тем, что такие пакеты часто разрабатываются для самых разнообразных систем посредством одного набора моделирующих конструкций, тогда как программа на языке С может быть более удачно написана под конкретную систему).
3. При программировании универсальные языки обеспечивают большую гибкость, чем пакеты имитационного моделирования.
4. Стоимость применяемого программного обеспечения обычно ниже (хотя общая стоимость проекта не всегда).

Таким образом, использование разных классов инструментальных средств имеет свои преимущества, поэтому разработчику моделей следует очень обдуманно подходить к выбору каждого из них.

2.4.2. Возможности специализированных пакетов имитационного моделирования

При выборе специализированного пакета имитационного моделирования следует учитывать все предоставляемые возможности, которые можно объединить в следующие группы:

- основные характеристики;
- совместимое оборудование и программное обеспечение;
- анимация;
- средства получения и обработки статистических данных;
- услуги, предоставляемые заказчикам и документация;
- отчеты с выходными данными и графиками.

2.4.2.1. Основные характеристики

Одно из основных свойств программного продукта имитационного моделирования, по мнению многих разработчиков, – *гибкость при моделировании*, то есть возможность моделировать системы с различным уровнем сложности технологических операций. Учитывая то, что не существует двух полностью идентичных систем, пакет имитационного моделирования, где применяется *фиксированное* число моделирующих конструкций, и нет возможностей программирования, обязательно окажется неподходящим для некоторых систем, встречающихся на практике. В идеале должна существовать возможность моделировать любую систему, используя исключительно конструкции программного пакета и не прибегая к программам, написанным на каком-либо языке, например на С. Далее перечислены некоторые возможности, придающие гибкость программному продукту имитационного моделирования.

1. Возможность определять и изменять атрибуты объектов и глобальных переменных, а также применять как те, так и другие в логике решений (например, конструкции If...then, else).
2. Возможность использовать математические выражения и функции (логарифмы, возведение в степень и т. п.).
3. Возможность создавать новые моделирующие конструкции и изменять уже существующие, а также применять новые и измененные моделирующие конструкции в данной и будущих моделях.

Следующим важным свойством средств имитационного моделирования является *простота в применении* (и изучении), поэтому многие современные пакеты моделирования снабжены графическим интерфейсом пользователя. В такой программе должны быть моделирующие конструктивные элементы: (скажем» пиктограммы или блоки), не слишком «примитивные», но и не слишком «изощренные». В первом случае понадобится очень много конструктивных элементов для моделирования даже достаточно простой ситуации; во втором - диалоговое окно каждого конструктивного элемента будет содержать чересчур большое число параметров, необходимых для обеспечения соответствующей гибкости программы. Управлять такими параметрами можно с помощью вкладок в диалоговых окнах.

Для сложных систем может оказаться полезным иерархическое моделирование. *Иерархия* позволяет сгруппировать несколько конструктивных основных элементов моделирования в новые структурные компоненты более высокого уровня. Эти новые структурные компоненты затем можно объединить в структурные компоненты еще более высокого уровня и т.д. Последние структурные компоненты помещают в библиотеку доступных структурных компонентов, и их можно повторно использовать в данной или будущих моделях. Повторное применение частей модели с расширением логических возможностей повышает эффективность моделирования. Иерархия является важной концепцией многих пакетов моделирования. Она также помогает избежать «неразберихи» на экране в графически ориентированных моделях, которые состоят из множества пиктограмм и блоков.

Программное обеспечение должно быть снабжено хорошими *средствами отладки*, такими как интерактивный отладчик. Мощный отладчик позволяет:

- отслеживать отдельные объекты по всей модели, чтобы убедиться в правильности их обработки;
- проверять состояние модели при каждом возникновении определенного события (например, при поломке станка);
- устанавливать значения определенных атрибутов или переменных, для того, чтобы заставить объект продвинуться до конца по логическому пути, который встречается с малой вероятностью.

При моделировании некоторых систем очень важна *высокая скорость работы модели*. Это касается моделей военных систем и моделей, в которых требуется обрабатывать большое количество объектов (например, модель быстродействующей сети связи).

Если имитационная модель будет использоваться кем-нибудь кроме самого разработчика, желательно, чтобы существовала возможность создавать *удобный дружественный интерфейс*, с помощью которого неспециалист будет легко вводить параметры моделирования, например, среднее время обслуживания или продолжительность моделирования.

В настоящее время большой интерес вызывает возможность *импортировать данные из других приложений* или *экспортировать данные в другие приложения* (например, таблицы Excel или базы данных).

Желательно, чтобы пакет имитационного моделирования мог автоматически *моделировать различные сценарии, которые повторяются с изменением какого-либо параметра моделирования* (например, числа станков на производственном рабочем месте). Кроме того, должна существовать возможность построения графика для какого-либо рабочего показателя (например, среднего времени пребывания в производственной системе) в виде функции изменяемого параметра.

Иногда возникает необходимость ввести в имитационную модель сложный блок программной логики, написанный на каком-либо языке программирования, поэтому желательно, чтобы в пакете имитационного моделирования можно было *запускать внешние программы*.

Еще одна полезная возможность заключается в *сохранении состояния модели в конце прогона* и использовании сохраненных данных для того, чтобы позднее возобновить моделирование.

Наконец, при покупке программных средств имитационного моделирования большое значение имеет их *стоимость*. На сегодняшний день стоимость различных программных средств имитационного моделирования колеблется от 500 до 50 000 долларов. Однако есть и другие расходы, на которые следует обратить внимание, а именно: затраты на эксплуатацию и обслуживание, затраты на модернизацию и стоимость любого дополнительного оборудования и программного обеспечения, которое может понадобиться.

2.4.2.2. Оборудование и программное обеспечение

При выборе программного средства имитационного моделирования необходимо обращать внимание на то, для какой *компьютерной платформы* оно предназначено. Почти все программные средства работают на персональных компьютерах с системой Windows, некоторые работают на рабочих станциях UNIX и компьютерах Apple. Если пакет может работать на разных платформах, тогда он должен обеспечивать *совместимость с*

другими платформами. Кроме того, следует также учесть *необходимый объем оперативной памяти.* Нужно также помнить о том, *какие операционные системы поддерживает пакет.*

2.4.2.3. Анимация и динамическая графика

Наличие во многих пакетах встроенных средств анимации — одна из причин роста интереса к имитационному моделированию. При использовании *анимации* все ключевые элементы системы представлены на экране пиктограммами, которые динамически изменяют свое положение, цвет и форму по мере развития имитационной модели во времени. Так, в производственной системе, пиктограмма изображающая автопогрузчик, будет менять свое положение при возникновении соответствующих изменений в модели, а пиктограмма, обозначающая станок, может менять цвет при переходе станка из одного состояния в другое (например, из состояния незанятости в состояние занятости) в модели. Анимацию обычно используют в таких случаях:

- для представления сути имитационной модели (или самого моделирования) руководителю или другим сотрудникам, которые не ознакомлены с техническими подробностями ее работы (или ранее не интересовались ими);
- с целью отладки моделирующей компьютерной программы;
- если надо показать, что имитационная модель *не* правильна;
- чтобы предложить для системы более совершенный технологический процесс (не все можно определить только по численным результатам моделирования);
- в обучении обслуживающего персонала;
- при обмене информацией между участниками проекта.

Существует два основных типа анимации: совместная и раздельная (также именуемая *воспроизведением*). *Совместная анимация* осуществляется во время прогона имитационной модели. Но следует иметь в виду, что обычно при выполнении рабочих прогонов анимация отключается, так как она замедляет их выполнение. При *раздельной анимации* изменения состояния системы сохраняются в файле на диске и применяются

для управления графикой *после* завершения моделирования. В некоторых программных средствах имитационного моделирования реализованы оба типа анимации.

К средствам анимации также предъявляются определенные требования. Прежде всего, необходимо, чтобы программное обеспечение имитационного моделирования использовало *анимацию по умолчанию* как часть процесса создания модели. Поскольку в первую очередь средства анимации необходимы для визуализации работы системы и обмена информацией, должна существовать возможность *создания пиктограмм с высокой разрешающей способностью* и сохранения их для последующего применения. Кроме того, нужно, чтобы поставляемые программные средства были оснащены *библиотекой стандартных пиктограмм* и обеспечивали *плавное движение пиктограмм*. Должен также существовать элемент управления для *ускорения, или замедления эффектов анимации*. Необходима и возможность *увеличения или уменьшения масштаба изображения*, а также *прокрутки изображения* для просмотра различных частей системы, если она полностью не помещается на экране. В некоторых программных средствах требуются *именованные, представления* анимации, соответствующие различным частям моделируемой системы и позволяющие создавать меню из представлений. Желательно, чтобы для анимации применялась *векторная графика, а не растровая*. Первый вид графики обеспечивает вращение объектов (скажем, несущего винта вертолета) и сохранение направления движения (например, транспортного средства, осуществляющего поворот).

Некоторые средства имитационного моделирования с совместной анимацией позволяют остановить моделирование «на ходу» при просмотре анимации, внести изменения в параметры моделирования (например, число станков на рабочем месте), а затем возобновить моделирование. Однако использование такой возможности может оказаться *опасным для статистических данных*, если состояние системы и статистические счетчики не будут переустановлены.

Некоторые из пакетов имитационного моделирования обеспечивают *объемную (трехмерную) анимацию* (точка наблюдения поворачивается вокруг любой из трех осей). Такая анимация имеет большое значение при презентации модели и в ситуациях, когда особенно важно оценить вертикальные габариты. В пакетах с

трехмерной анимацией можно также предоставить зрителю возможность как бы перемещаться по системе на определенном объекте.

Должна существовать и возможность применения в анимации импортируемых *CAD-чертежей и графических объектов Clip Art*.

Часто может быть полезным вывод на экран *динамической графики и статистических данных* по мере выполнения моделирования. В качестве динамической графики обычно используются часы, круговая шкала, индикатор уровня (отображающий, например, очередь) и динамически обновляемые гистограммы и графики временной зависимости. Как пример графика временной зависимости можно привести график числа заявок в некоторой очереди, обновляемый по мере продвижения модельного времени.

2.4.2.4. Статистические возможности

Если имитационная модель не снабжена надежными средствами статистического анализа, невозможно будет получать достоверные данные о работе моделируемой системы. Прежде всего, в программном средстве необходим хороший *генератор случайных чисел*, то есть механизм для генерирования независимых значений, равномерно распределенных в интервале $(0,1)$. Следует иметь в виду, что не все генераторы случайных чисел, которыми снабжены компьютеры или программное обеспечение, обладают удовлетворительными статистическими характеристиками.

Генератор должен поддерживать, по меньшей мере, 100 потоков (а желательно и больше), которые можно назначать для различных случайных факторов, имеющихся в имитационной модели, в частности для интервалов времени между поступлениями или времени обслуживания. Это позволит эффективно сравнивать проекты разных систем с помощью полученной статистики.

В тех случаях, когда для разных потоков заданы начальные значения, которые используются по умолчанию, программное обеспечение имитационного моделирования должно давать одинаковые результаты при многократном моделирова-

нии. Нежелательной является такая ситуация, когда начальные значения зависят от внутренних часов компьютера. У пользователя всегда должна быть возможность, если есть такая необходимость, задавать начальные значения для всех потоков самостоятельно.

В целом каждый фактор случайности системы должен быть представлен в имитационной модели *распределением вероятностей*, а не только средним значением. Если можно подобрать стандартное *теоретическое распределение*, позволяющее оптимально представить некоторый фактор случайности, то такое распределение следует использовать в модели системы. Программное обеспечение должно поддерживать следующие *непрерывные распределения*: экспоненциальное, Вейбулла, логарифмически нормальное, нормальное, равномерное и треугольное, а также гамма- и бета-распределение. Треугольное и бета-распределение, как правило, используются в качестве модели фактора случайности, когда не доступны никакие системные данные. Надо учитывать также, что *очень немногие* входные случайные переменные при моделировании реальных систем имеют нормальное распределение. Кроме непрерывных должны быть доступны еще и *дискретные распределения*: биномиальное, геометрическое и отрицательное биномиальное, а также распределение Пуассона и дискретно-равномерное.

Если нет возможности найти теоретическое распределение, удачно представляющее некоторый фактор случайности, следует воспользоваться *эмпирическим* (или *определенным пользователем*) *распределением*, которое основывается на полученных данных. При этом случайные числа получают путем выборки значений случайной величины с функцией распределения, созданной по наблюдаемым системным данным.

Программа должна содержать команду, позволяющую выполнять *независимые повторения* (репликации) или *прогоны* имитационной модели. Это означает, что

- для каждого из прогонов используются отдельные наборы разных случайных чисел;
- для каждого прогона применяются одни и те же входные условия;
- при каждом прогоне статистические счетчики переводятся в исходное состояние.

Следует учитывать, что результаты, полученные при различных прогонах имитационной модели, независимы и в то же время являются вероятностными копиями друг друга. Это позволяет применять к таким результатам классические статистические процедуры.

Чтобы построить *доверительный интервал* для среднего значения (в частности, среднего времени пребывания в системе для детали на заводе), необходим статистически устойчивый метод. Он должен быть простым для понимания и обеспечивать надежные статистические результаты. В этом смысле *метод повторения, или репликации*, несомненно, является наиболее удачным.

Если трудно определить длительный прогон или «установившийся режим» системы, то желательно указать продолжительность переходного периода работы модели, необходимого для того, чтобы статистические счетчики сбрасывались в 0 (но состояние системы при этом не изменялось). В идеале программные средства имитационного моделирования должны иметь возможность *определять продолжительность переходного процесса работы модели* на основе экспериментальных прогонов.

Программные средства имитационного моделирования должны предоставлять возможность указывать, *для оценивания каких критериев работы системы будут собираться выходные данные*, а не просто выдавать сведения, вычисляемые по умолчанию и не представляющие для пользователя никакого интереса.

В настоящее время разработчиков, желающих приобрести программные средства имитационного моделирования, интересует еще вопрос *оптимизации*. Предположим, что существует некоторое число искомых переменных (входных факторов), каждая из которых имеет свой ряд возможных значений. (Для искомых переменных также могут задаваться линейные ограничения.) Кроме того, существует целевая функция, которая должна быть максимизирована (или минимизирована), то есть функция одной или нескольких выходных случайных переменных имитационного моделирования (например, функция производительности в производственной системе). Тогда цель оптимизации — выполнить прогоны имитационной модели (для каждого из прогонов используются определенные значения искомых переменных) с помощью интеллектуальных модулей оптимизации, и

найти в итоге комбинацию искомых переменных, дающих оптимальное или близкое к нему решение. В таких модулях оптимизации применяются следующие эвристические средства; генетические алгоритмы, моделированный «отжиг», нейронные сети, а также рассеивающий поиск и запрещающий поиск.

2.4.2.5. Отчеты с выходными данными и графика

Для оценки показателей работы модели в программе должна быть заложена возможность создавать *стандартные отчеты*. При этом надо учитывать потребность в *отчетах заданного формата*, например, для представления данных руководству. Если программа имитационного моделирования достаточно гибкая, она обеспечивает вычисление оценок показателей работы, определенных пользователем, и предоставляет возможность записывать их в отчете заданного формата.

Для каждого показателя работы модели (например, времени пребывания детали на заводе) обычно вычисляются среднее, максимальное и минимальное наблюдаемые значения.

Кроме того, есть необходимость получать отчеты, как в промежуточных точках прогона имитационной модели, так и по его завершении.

Программа имитационного моделирования должна обеспечивать разнообразную статическую графику. Прежде всего, необходимо иметь возможность создавать *гистограммы* для некоторых наблюдаемых данных. Для непрерывных данных гистограмма представляет собой оценку лежащей в ее основе функции плотности распределения вероятностей, для дискретных — оценку функции суммарного их распределения.

Также очень важными являются графики временной зависимости. На *графике временной зависимости* значения одной или нескольких ключевых системных переменных (например, число требований в определенной очереди) изображаются в течение всего времени моделирования, обеспечивая, таким образом, представление о динамическом поведении моделируемой системы за весь период моделирования (тогда как анимация обеспечивает представление о динамическом поведении системы в текущий момент времени).

Некоторые программы имитационного моделирования позволяют выводить результаты в виде *столбиковых* или *круговых диаграмм*. Кроме того, желательно, чтобы, *результаты для различных моделируемых сценариев сохранялись в базе данных и при необходимости могли быть изображены на одном графике*.

Наконец, полезным для определения зависимости в выходных данных, получаемых в результате одного прогона имитационной модели, может быть *график корреляционной функции*.

Должна также существовать возможность экспортировать *результаты наблюдений выходов отдельной модели* (например, время пребывания в системе) в такие программные пакеты, как электронные таблицы, базы данных, статистические и графические пакеты, с целью проведения дальнейшего анализа.

2.4.3. Специализированные средства имитационного моделирования систем

В рамках курса предлагается изучить основные возможности и получить опыт разработки простых имитационных моделей в средах *GPSS World* и *ExtendSim*. Кратко опишем основные возможности этих инструментальных средств моделирования.

2.4.3.1. Язык моделирования GPSS

GPSS – один из первых языков моделирования, разработанный еще в начале 1970-х годов фирмой IBM. Система моделирования GPSS World является торговой маркой фирмы MINUTEMAN Software. Благодаря простоте и удобству использования конструкций языка GPSS является одним из самых распространенных специализированных языков моделирования в мире [2, 7, 9, 10, 23, 28].

Модель системы на GPSS представляет собой последовательность блоков, каждый из которых соответствует некоторому оператору (подпрограмме). Каждый блок имеет определенное количество реквизитов, называемых операндами, которые отделяются друг от друга запятой (это аналоги параметров процедур и

функций в языках программирования), но положение операндов строго фиксировано, и отсутствие некоторого операнда отмечается запятой. Блоки записываются в следующем формате: *метка имя_оператора операнды; комментарии*.

Операторы располагаются в программе последовательно. Оператор может иметь метку для ссылки на этот оператор в других операторах.

Имя_оператора – это ключевое слово, указывающее конкретную функцию, выполняемую данным оператором. Операнды задают параметры выполняемой функции. Оператор в зависимости от типа может содержать до девяти операндов, расположенных в определённой последовательности и обозначаемых буквами латинского алфавита от A до I. Некоторые операторы вообще не имеют операндов, а в некоторых операнды могут быть пропущены.

После трансляции исходной программы в памяти ПК создаётся текущая модель, являющаяся совокупностью разного типа объектов. Каждому объекту соответствуют числовые атрибуты, описывающие его состояние в данный момент модельного времени. Значения атрибутов могут быть арифметическими или логическими. Атрибуты, доступные программисту, называются стандартными числовыми атрибутами. Кроме того, имеется ряд так называемых системных атрибутов, относящихся не к отдельным объектам, а к модели в целом. Значение атрибутов всех объектов модели по окончании моделирования выводятся в стандартный отчёт. Каждый объект GPSS имеет имя и номер. Имя объекта представляет собой начинающуюся с буквы последовательность букв латинского алфавита, цифр и символа подчёркивания.

В GPSS реализован сбор статистических данных, описывающих поведение модели и автоматическая печать итоговой статистики по завершении моделирования.

В язык включены и многие другие полезные элементы. Например, GPSS обслуживает таймер модельного времени, планирует события, которые должны произойти позднее в течение времени моделирования, вызывает их своевременное появление и управляет очередностью поступления.

В GPSS реализована концепция "передачи управления от блока к блоку". Последовательность блоков GPSS-модели отражает направления, по которым происходит движение перемещающихся элементов. Каждый такой элемент называется транзактом. Транзакты являются динамическими элементами GPSS-модели. Работа этой модели заключается в перемещении транзактов от блока к блоку. В самом начале моделирования в GPSS-модели нет ни одного транзакта.

В процессе моделирования транзакты входят в модель в определенные моменты времени в соответствии с заданными параметрами. Подобным же образом транзакты покидают модель. В общем случае в модели существует большое число транзактов, но в один момент времени двигается только один.

Центральной задачей, выполняемой симулятором, является определение того, какой транзакт надо выбрать следующим для продвижения в модели, когда его предшественник прекратил продвижение. С этой целью симулятор рассматривает каждый транзакт как элемент некоторого списка. В относительно простых моделях используются список текущих событий (время продвижения меньше или равно модельному) и список будущих событий (время продвижения больше модельного). Симулятор имеет таймер модельного времени.

Моделирование системы на GPSS, по существу, представляет собой имитацию последовательности переходов системы из одного состояния в другое в некоторые моменты времени, поэтому важнейшим является организация моделирования во времени. Время в GPSS-модели отображается в условных единицах (целыми числами), масштаб которых устанавливается исследователем. Приращение модельного времени на каждом шаге моделирования - величина случайная, равная интервалу времени между последовательными событиями.

Для имитации процессов, протекающих в моделируемой системе, в GPSS предусмотрены следующие механизмы:

- все транзакты, порождаемые в процессе моделирования, образуют списки, в которых транзакты отсортированы, во-первых, по времени, во-вторых, при равных временах у транзактов, по приоритетам;

- в процессе моделирования может быть образовано до пяти списков: *список текущих событий* содержит транзакты, которые должны перемещаться в текущий момент модельного времени; все они имеют одинаковое время, равное текущему, и если их приоритеты совпадают, то очередной транзакт для перемещения выбирается по правилу FIFO; *список будущих событий* содержит транзакты, которые будут перемещаться по модели в будущие моменты времени; в этот список попадают транзакты, вошедшие в блоки имитации задержки или заблокированные по каким-либо условиям; *списки прерываний, синхронизации и пользователя* содержат транзакты, обслуживание которых прервано по некоторым условиям, задаваемым блоками прерывания, синхронизации или находящимся под управлением пользователя.

Транзакты представляют собой описание динамических процессов в реальных системах. Они могут описывать реальные физические объекты, например, заявки, поступающие в вычислительную систему. Транзакты можно «генерировать» и «уничтожать» в процессе моделирования системы S. Основным атрибутом любого транзакта являются его параметры, число которых для каждого транзакта может быть от 0 до 1020. Использование параметров определяется пользователем GPSS. Например, при моделировании сети передачи данных программист может резервировать первый параметр для значения длины сообщения, второй — для номера источника (терминала), выдавшего это сообщение, и т. д. Важным атрибутом любого транзакта является уровень приоритета PR, изменяющийся в пределах от 0 до 127. В случае, когда два транзакта соперничают при занятии данного устройства, сначала обрабатывается тот, у которого приоритет выше. Если приоритеты у обоих транзактов одинаковые, сначала обрабатывается тот, у которого время ожидания обработки больше. В одном задании может выполняться как один, так и несколько прогонов одной модели. При этом текущим значением абсолютного времени модели AC1 будет называться суммарное время по всем реализованным прогонам, а текущим значением относительного времени модели C1—системное время в пределах одного прогона.

Формальными основными моделями, для имитации которых используется GPSS, являются системы массового обслуживания (Q-схемы), конечные и вероятностные автоматы (F и P-схемы), сети Петри (N-схемы), агрегаты (A-схемы) [1-2].

2.4.3.2. Расширенный редактор GPSS

Один из существенных недостатков среды GPSS World – отсутствие удобного графического интерфейса, но в ноябре 2014 года фирма «Элина-Компьютер» выпустила расширенный редактор к среде GPSS World (версия 1.7) [9]. Редактор расширяет возможности стандартного GPSS, с его помощью можно:

1. Написать модель в текстовом редакторе, который поддерживает автовыравнивание, подсветку синтаксиса, контекстную подсказку, обнаружение ошибок в операторах, группировку и ряд других средств автоматизации ввода.
2. Сконструировать модель с помощью графического редактора схем.
3. Провести серию экспериментов с моделью системы.
4. Сформировать и распечатать отчет об исследовании в формате Microsoft Word, который будет содержать все данные и описания моделей, схем, экспериментов и результаты моделирования.
5. Создать авто-формы ввода исходных данных и вывода результатов моделирования. Сформировать выполняемый exe-модуль, работающий независимо от расширенного редактора.

Новые возможности никоим образом не изменяют спецификацию модели, расширенный редактор использует для запуска моделирования средства стандартного GPSS World, поэтому его наличие обязательно.

Редактор удобно использовать для создания и отладки модели, к тому же редактор полностью русифицирован.

Видимо, появление расширенного редактора повысит интерес к использованию GPSS.

2.4.3.3. Пакет имитационного моделирования ExtendSim

Универсальная среда имитационного моделирования ExtendSim поставляется компанией Imagine That (Сан-Хосе, Калифорния) [3]. Среда ExtendSim предназначена для разработки имитационных моделей систем разных типов, как дискретных, так и непрерывных. В ExtendSim используется так же, как и в пакете GPSS, транзакт – парадигма, т.е. созданная заявка (транзакт) перемещается от одного блока модели к другому блоку и активирует их работу. Но, в отличие от GPSS, ExtendSim – среда визуального моделирования: каждый блок имеет условное графическое обозначение, блоки связаны между собой линиями-соединителями. Блоки включены в библиотеки, всего в ExtendSim семь специализированных библиотек для моделирования различных типов систем. Процесс разработки модели заключается в выборе блоков из библиотеки, размещении блоков в модельном окне, настройке блоков и связывании блоков с помощью линий-соединителей. С каждым блоком связано диалоговое окно, в котором задаются параметры работы блока (например, закон распределения интервалов между поступлением заявок, время обслуживания заявки, дисциплина обслуживания и т.п.).

Пакет ExtendSim позволяет моделировать самые разнообразные конфигурации систем, поскольку он содержит внутренний язык ModL для настройки существующих блоков и создания новых. Новые блоки помещают в новую библиотеку и применяют повторно как в данной, так и в других моделях. Чтобы наглядно представить связи между блоками, нужно дважды щелкнуть мышью на выбранном блоке, удерживая нажатой клавишу Alt, иерархия модели может включать неограниченное число уровней, здесь также может использоваться механизм наследования. В настоящее время пакет Extend обеспечивает двухмерную и трехмерную анимацию, при этом в качестве вспомогательного средства служит платформа Proof Animation [Wolverine, 1996].

С каждой имитационной моделью в пакете Extend связана программа Notebook, в которой могут содержаться элементы диалоговых окон и результаты моделирования. Таким образом, Notebook может применяться как интерфейс для мо-

дели или как средство для вывода на экран важных результатов моделирования во время прогона имитационной модели.

Число потоков случайных чисел в пакете Extend не ограничено. Более того, имеется доступ к 18 стандартным теоретическим распределениям вероятностей, а также к эмпирическим распределениям.

В пакете Extend существует простой способ выполнения независимых повторных прогонов моделируемой системы, а также построения точечных оценок и доверительных интервалов для показателей работы системы. Пакет позволяет строить графики, в частности гистограммы, а также графики с временной зависимостью. В настоящее время добавлен модуль оптимизации.

Кроме того, в пакете Extend есть возможности для моделирования непрерывных систем.

Данный пакет позволяет выполнять функционально-стоимостной анализ с использованием ABC-метода, благодаря чему по мере продвижения объекта в системе можно связывать с ним фиксированные и переменные затраты. Например, в производственной системе с деталью могут быть связаны фиксированные затраты на сырье, а переменные затраты будут зависеть от времени ожидания этой детали в очереди.

В целом, ExtendSim позволяет создать наглядную, интуитивно-понятную имитационную модель объекта.

2.4.3.4. Другие специализированные средства имитационного моделирования

Среди универсальных пакетов имитационного моделирования также хорошо известны:

1. Система моделирования AnyLogic

Система AnyLogic [21] поддерживает три технологии создания имитационных моделей: процессно-ориентированный (дискретно-событийный), системно-динамический и агентный, а также любую их комбинацию. Графический интер-

фейс AnyLogic, инструменты и библиотеки позволяют быстро создавать модели для широко спектра задач от моделирования производства, логистики, бизнес-процессов до стратегических моделей развития компании и рынков. AnyLogic стал корпоративным стандартом на бизнес-моделирование во многих транснациональных компаниях, широко используется в образовании. Сайт: <http://www.anylogic.ru/>.

2. Система Arena

Arena - система дискретного моделирования. Сфера основных приложений системы – имитационное моделирование производственных технологических процессов и операций, складской учет, банковская деятельность, оптимизация обслуживания клиентов в сфере услуг, транспортные задачи.

Разработчик: Rockwell Automation Inc., Wexford, PA, США.

Сайт: <http://www.arenasimulation.com/>

3. Среда моделирования AutoMod

Система AutoMod предназначена для моделирования систем логистики и производства. Программное обеспечение разработано для детального анализа операций и потоков. Хотя главным образом используется в производстве и материальном анализе систем обработки, гибкая архитектура AutoMod позволяет использовать в широком диапазоне прикладных областей, от аэропортов до промышленности полупроводников.

Разработчик: Brooks Automation, США.

Сайт: <http://www.automod.se/eng/home.html>

4. Среда моделирования AweSim

AweSim – это универсальная система имитационного моделирования для сети с дискретной или непрерывной интерпретацией. Возможные области применения: бизнес, промышленность, здравоохранение, военное дело. Будучи совместимым с языком имитационного моделирования Visual SLAM, продукт включает построение интерактивной модели, одновременную и последующую анимацию, статистическую информацию в текстовом и графическом видах, интерактивное представление и выбор сценариев. Сетевые модели строятся графически и могут

быть иерархическими. Они могут быть расширены по заданным пользователем правилам, написанным на языке C или Visual Basic. Одновременно могут отображаться несколько анимированных изображений. Сценарии сравниваются статистически, после чего из них выбирают набор альтернатив с лучшими показателями.

Разработчик: Symix Systems Inc., США.

5. Система Boson NetSim

Система Boson NetSim – это коммерческий симулятор, по сути своей эмулятор, сетевых устройств компании Cisco на основе Cisco IOS. Данная система позволяет получить практические знания по работе с сетевыми устройствами, начиная от обычных управляемых свичей и заканчивая роутерами 7-го поколения. В поставку включается утилита для моделирования сети. В ней можно смоделировать любой тип сети или взять готовую сеть из примеров. Сертификация специалистов Cisco CCNP (Cisco Certified Network Professional) проходит именно в этой программе.

Сайт: <http://www.boson.com/netsim-cisco-network-simulator>

Сайт: <http://www.boson.com/network-simulator/ccnp-cisco-network-simulator>.

В общем, в развитии специализированных средств моделирования можно выделить два направления:

1. Средства моделирования для анализа широкого класса систем. К ним относятся языки и среды имитационного моделирования (GPSS, SimScript, Simula, Arena, Extend, AveSim, AnyLogic и другие) [2, 3, 4, 7, 8-11, 21, 23, 28], пакеты прикладных программ, использующих для моделирования аналитические методы, такие как MathCad, MathLab, SAS, Statistica и другие. Основной недостаток этих средств – необходимость специальной подготовки исследователя.

2. Программные комплексы, специализирующиеся на моделировании узкого круга систем одной конкретной предметной области [3]. Например, пакеты имитационного моделирования производственных систем (AutoMod, WITNESS, QUEST, ProModel и другие); пакеты имитационного моделирования для медицинских учреждений (MedModel); пакеты имитационного моделирования се-

тей связи (COMNET, OPNET Modeler, IT Decision Guru). Недостаток, заключающийся в ограниченности применения таких программ, компенсируется такими преимуществами, как легкость их освоения специалистами предметной области, и эффективность применения, вследствие узкой специализации.

Подробный обзор современных программных продуктов имитационного моделирования систем приведен на сайте: <http://simulation.su>.

2.5. Принципы построения моделирующих алгоритмов

Задачей имитационного моделирования является получение траектории движения (изменения) рассматриваемой системы в n – мерном пространстве $(Z_0(t), Z_1(t), Z_2(t), \dots, Z_n(t))$, а также вычисление некоторых показателей, зависящих от выходных сигналов системы и характеризующих ее свойства.

Каждое состояние описывается набором фазовых переменных, количество фазовых переменных – k :

$$Z_0(t) = \{z_0^1(t), z_0^2(t), \dots, z_0^k(t)\}$$

$$Z_1(t) = \{z_1^1(t), z_1^2(t), \dots, z_1^k(t)\}$$

...

$$Z_i(t) = \{z_i^1(t), z_i^2(t), \dots, z_i^k(t)\}$$

Известны два принципа построения моделирующего алгоритма процесса функционирования систем: принцип Δt и принцип особых состояний (δz) [1, 24, 25].

Принцип Δt . Предположим, что начальное состояние системы соответствует значениям:

$$Z_0(0) = \{z_0^1(0), z_0^2(0), \dots, z_0^k(0)\}$$

Принцип Δt предполагает преобразование модели системы к такому виду, чтобы значения

$$Z_1(1) = \{z_1^1(1), z_1^2(t), \dots, z_1^k(1)\}$$

в момент времени $t_1 = t_0 + \Delta t$ можно было вычислить через начальные значения, а в момент $t_2 = t_1 + \Delta t$ через значения на предшествующем шаге и так для каждого i -ого шага ($\Delta t = const$; $i = 1, \dots, M$).

Таким образом, часы модельного времени продвигаются с постоянным шагом, точно на заданный интервал Δt . После каждого обновления часов выполняется проверка, чтобы определить, произошли какие-либо события в течение предыдущего интервала времени Δt или нет. Если на этот интервал запланированы одно или несколько событий, считается, что данные события происходят в конце интервала, после чего состояние системы и статистические счетчики соответствующим образом обновляются. Продвижение времени посредством постоянного шага показано на рисунке 2.2., где изогнутые стрелки показывают продвижение часов модельного времени, а e_i ($i = 1, 2, \dots$) – это действительное время возникновения события i любого типа, а не значение часов модельного времени.

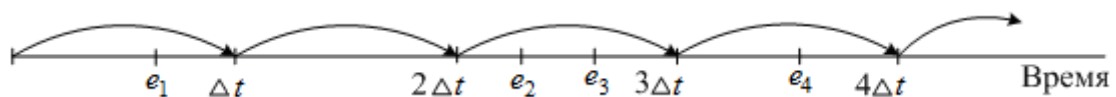


Рисунок 2.2 – Моделирование в соответствии с принципом Δt

На интервале $[0, \Delta t)$ событие происходит в момент времени e_1 , но оно рассматривается как произошедшее в момент времени Δt . На интервале $[\Delta t, 2\Delta t)$ события не происходят, но все же модель выполняет проверку, чтобы убедиться в этом. На интервале $[2\Delta t, 3\Delta t)$ события происходят в моменты времени e_2 и e_3 , однако считается, что они произошли в момент времени $3\Delta t$ и т.д. В ситуациях, когда принято считать, что два или несколько событий происходят в одно и то же время, необходимо применение ряда правил, позволяющих определять, в каком порядке обрабатывать события. Таким образом, продвижение времени посредством постоянного шага имеет два недостатка: возникновение ошибок, связанных с обработкой событий в конце интервала, в течение которого они происходят, а также необходимость решать, какое событие обрабатывать первым, если события, в действительности, происходящие в разное время, рассматриваются как одно-

временные. Подобного рода проблемы можно частично решить, сделав интервалы Δt менее продолжительными, но тогда возрастает число проверок возникновения событий, что приводит к увеличению времени выполнения задачи. Принимая во внимание это обстоятельство, продвижение времени с помощью постоянного шага не используют в дискретно-событийных имитационных моделях, когда интервалы времени между последовательными событиями могут значительно отличаться по своей продолжительности.

Принцип Δt эффективно использовать для моделирования систем, состояние которых меняется через определенный (постоянный) шаг времени. Пример системы, которая моделируется на основе использования принципа Δt – устройство, возводящее в квадрат входной сигнал каждый такт времени или дифференцирующий фильтр.

Пример. Процесс, происходящий в фильтре (рисунок 2.3), описывается дифференциальным уравнением:

$$\frac{dZ}{dt} = K(x(t) - Z(t)), \quad (2.1)$$

где K - коэффициент усиления; $x(t)$ – входной сигнал.

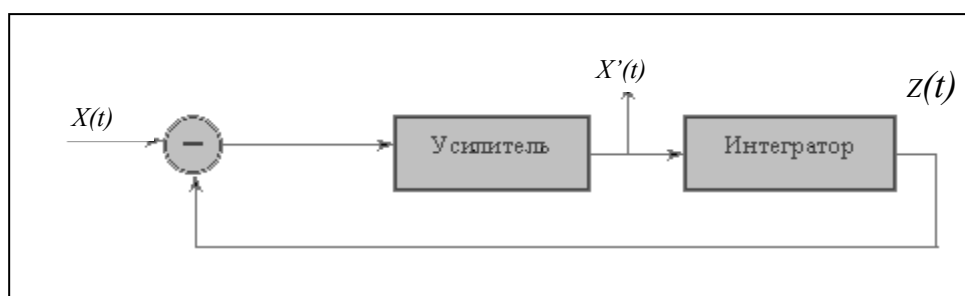


Рисунок 2.3 – Структурная схема дифференцирующего фильтра

Доказано, что:

$$\lim_{K \rightarrow \infty} \frac{dZ}{dt} = \frac{dx}{dt}. \quad (2.2)$$

Основываясь на определении производной:

$$\lim_{K \rightarrow \infty} \frac{dZ}{dt} = \frac{dx}{dt} = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t}. \quad (2.3)$$

При этом:

$$K = \frac{1}{\Delta t}. \quad (2.4)$$

Задав начальное условие $x(t_0)$ можно симитировать траекторию процесса, происходящего в фильтре, с целью получения текущего значения производной любой детерминированной функции $x(t)$, подаваемой на вход.

Воспользуемся принципом Δt построения моделирующего алгоритма. Пусть шаг $\Delta t=0,01$; $x(t) = \sin(t) + \cos(t)$; $x(t_0)=1$; $N=1000$.

В таблице 2.1 приведены результаты вычислений. В предпоследнем столбце таблицы 2.1 – приведены значения производной, вычисленной имитационным методом, в последнем столбце таблицы 2.1 приведены значения производной, вычисленные аналитически, по формуле: $x'(t) = \cos(t) - \sin(t)$.

Таблица 2.1.

i	t_i	$x(t_i)$	$x(t_i + \Delta t)$	$\frac{x(t_i + \Delta t) - x(t_i)}{\Delta t}$	$x'(t)$
0	0	1	1,00995	0,994983	1
1	0,01	1,00995	1,0198	0,98488	0,98995
2	0,02	1,0198	1,0295	0,97469	0,9798
3	0,03	1,0295	1,0392	0,96439	0,9696
4	0,04	1,0392	1,0487	0,95399	0,9592
5	0,05	1,0487	1,0582	0,94351	0,9488
6	0,06	1,0582	1,0675	0,93293	0,9382
7	0,07	1,0675	1,0767	0,92226	0,9276
8	0,08	1,0767	1,0858	0,91149	0,9169
9	0,09	1,0858		0,90063	0,9060
...					

На рисунке 2.4 приведены графики исходной функции $x(t)$ и производных $x'(t)$, рассчитанных имитационным методом и аналитическим. Очевидно, что погрешность расчета производной имитационным методом в сравнении с аналитическим методом не значительна и уменьшается с уменьшением шага Δt .

Значение производной, вычисленной имитационным методом, в нулевой момент времени может значительно отличается от значения производной, вычисленной аналитически и соответствовать переходному процессу в дифференциру-

ющем фильтре. Продолжительность переходного процесса в фильтре зависит от точности задания начального значения $x(t_0)$.

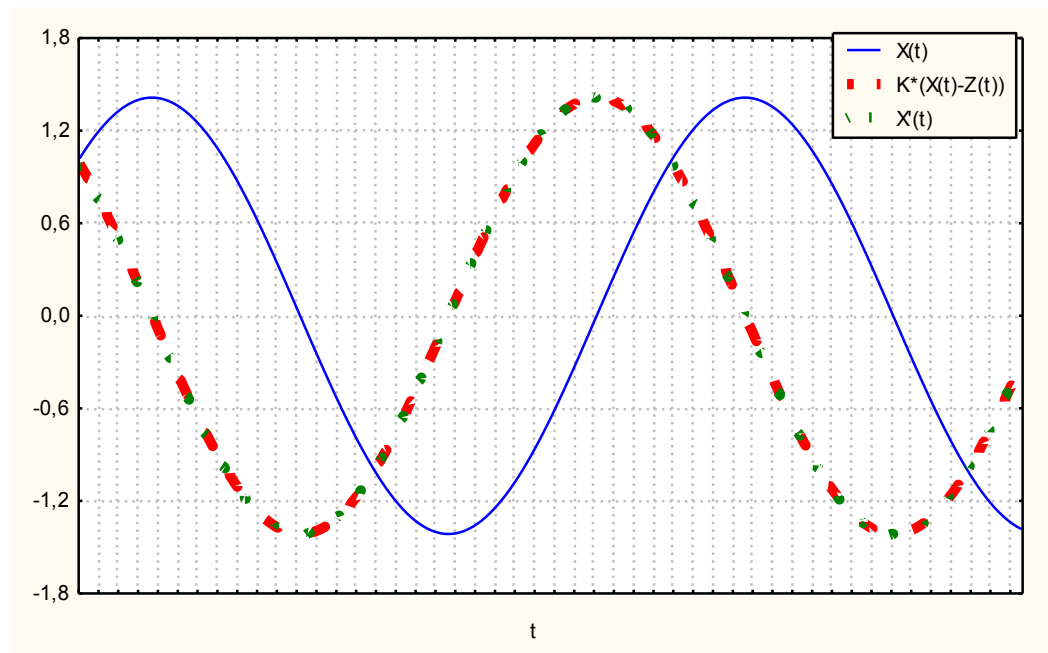


Рисунок 2.4

Принцип особых состояний (принцип δz). При рассмотрении некоторых видов систем можно выделить два вида состояний:

- обычное, в котором система находится большую часть времени, при этом $Z_i(t)$; $i = 1, \dots, n$ изменяются плавно;
- особое, характерное для системы в некоторые моменты времени, причем состояние системы изменяется в эти моменты скачком.

Принцип особых состояний отличается от принципа Δt тем, что шаг по времени в этом случае не постоянен, является величиной случайной и вычисляется в соответствии с информацией о предыдущем особом состоянии.

Примерами систем, имеющих особые состояния, являются системы массового обслуживания (СМО). Особые состояния появляются в моменты поступления заявок, в моменты освобождения каналов и т.д.

Подход, использующий продвижение времени в модели от одного особого состояния к другому (от события к событию), применяется в большинстве специализированных пакетов имитационного моделирования (например, в ExtendSim) и

используется большинством разработчиков, создающих свои модели на универсальных языках программирования (рисунок 2.5).

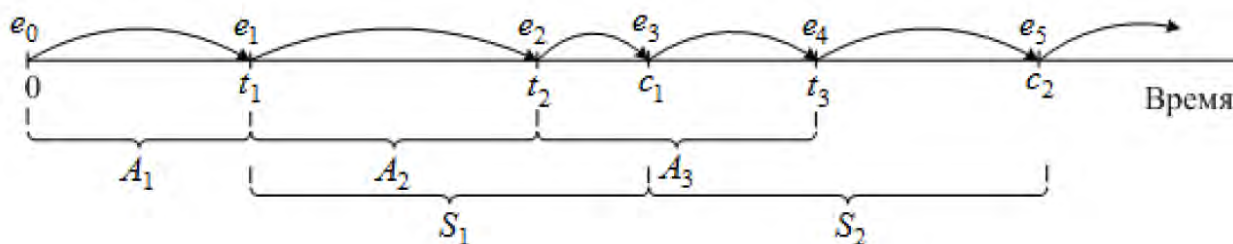


Рисунок 2.5 – Моделирование в соответствии с принципом δz

При использовании продвижения времени от события к событию часы модельного времени в исходном состоянии устанавливаются в 0, и определяется время возникновения будущих событий. После этого часы модельного времени переходят на время возникновения ближайшего события, и в этот момент обновляются состояние системы, с учетом произошедшего события, а также сведения о времени возникновения будущих событий. Затем часы модельного времени продвигаются ко времени возникновения, следующего нового ближайшего события, обновляется состояние системы и определяется время будущих событий и т.д. Процесс продвижения модельного времени от времени возникновения одного события ко времени возникновения другого продолжается до тех пор, пока не будет выполнено какое-либо условие останова, указанное заранее. Поскольку в дискретно-событийной имитационной модели все изменения происходят только во время возникновения событий, периоды бездействия системы просто пропускаются, и часы переводятся со времени возникновения одного события на время возникновения другого. При продвижении времени с постоянным шагом такие периоды бездействия не пропускаются, что приводит к большим затратам компьютерного времени.

Таким образом, принцип δz дает возможность для ряда систем существенно уменьшить затраты машинного времени на реализацию моделирующих алгоритмов по сравнению с принципом Δt . Логика построения моделирующего алгоритма, реализующего принцип δz , отличается от рассмотренной для принципа

Δt только тем, что включает в себя процедуру определения момента времени, соответствующего следующему особому состоянию системы. Для исследования процесса функционирования больших систем рационально использование комбинированного принципа построения моделирующих алгоритмов, сочетающего в себе преимущества каждого из рассмотренных принципов.

Пример. Пусть рассматривается вычислительная система, включающая один компьютер. Задан интервал поступления заданий на обработку (равномерный в диапазоне 2-6). Задан интервал выполнения заданий (равномерный в диапазоне 3-4). Реализовать моделирование работы вычислительной системы по обработке 100 заданий. Рассчитать среднее время ожидания в очереди и вероятность простоя компьютера.

1 шаг. Выберем входные, промежуточные и выходные переменные

входные и промежуточные переменные:

$T_{\text{инт. прих.}}$ – интервал между поступлением заданий;

$T_{\text{инт. obs.}}$ – интервал обработки задания;

$T_{\text{прих.}}$ – момент поступления очередного задания;

$T_{\text{нач. obs.}}$ – момент начала обработки очередного задания;

$T_{\text{ок. obs.}}$ – момент окончания обработки очередного задания;

выходные переменные:

$T_{\text{ож.}}$ – время ожидания задания в очереди;

$T_{\text{простоя}}$ – время простоя компьютера до поступления очередного задания;

$T_{\text{ож. ср.}}$ – среднее время ожидания задания в очереди;

$P_{\text{простоя}}$ – вероятность простоя компьютера.

В таблице 2.2 и на рисунке 2.6 приведены данные ручного моделирования обработки первых четырех заданий для иллюстрации принципа особых состояний. Интервалы поступления заданий и интервалы обработки заданы произвольно.

Таблица 2.2 – Имитационное моделирование обработки заданий

Номер шага	Т инт. прих.	Т инт. obs.	Т прих.	Т нач. obs.	Т ок. obs.	Т ож.	Т про-стоя
1	3	4	3	3	7	0	3
2	5	3	8	8	11	0	1
3	2	4	10	11	15	1	0
4	3	4	13	15	19	2	0
...							
100							

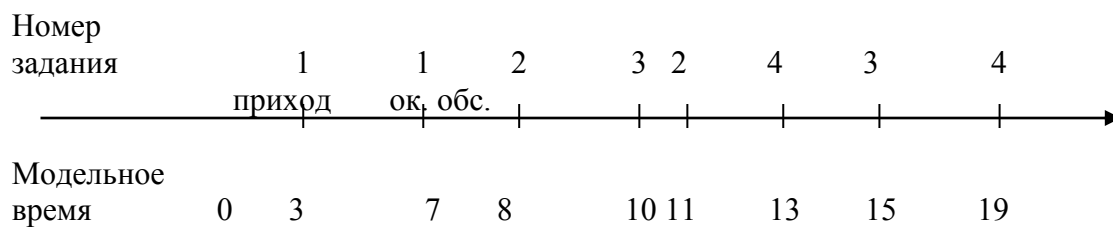


Рисунок 2.6

2 шаг. Разработка блок-схемы алгоритма имитации СМО (рисунок 2.7)

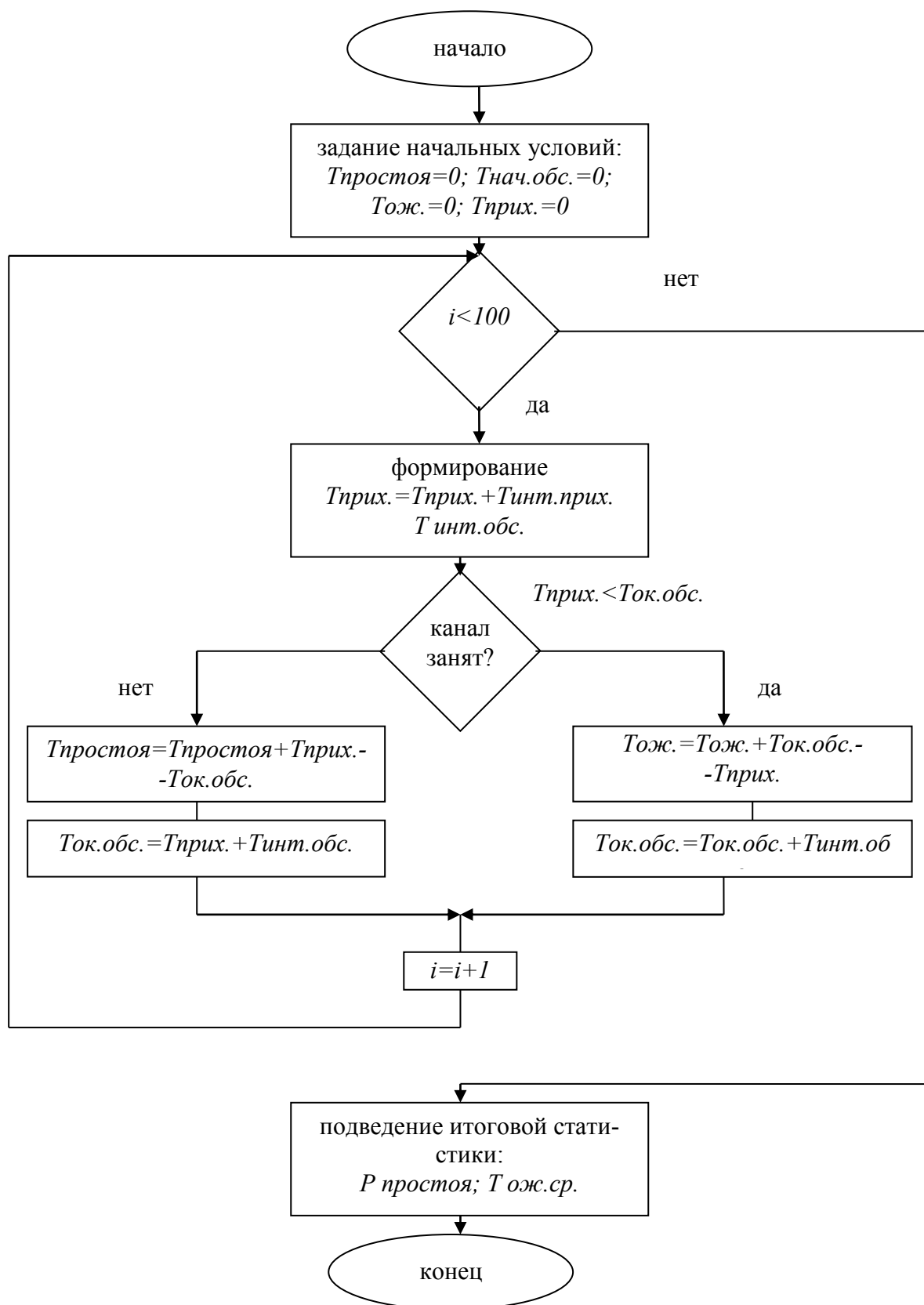


Рисунок 2.7 – Блок-схема алгоритма имитации работы СМО

3 шаг. В соответствии с разработанной блок-схемой составляется программа имитации работы СМО на любом языке программирования с использованием любых программных средств.

4 шаг. Расчет показателей эффективности СМО с помощью разработанной программы имитации.

5 шаг. Анализ полученных результатов, оценка эффективности СМО.

2.6. Методы генерации случайных последовательностей

При имитационном моделировании необходим учет случайных факторов и воздействий на систему, которые представляются случайными последовательностями чисел. Математически случайные последовательности описываются в виде случайных событий, значений дискретных и непрерывных случайных величин, значений случайных векторов и процессов. Для описания случайных последовательностей используется аппарат теории вероятностей и математической статистики.

Программная имитация случайной последовательности любого типа основана на генерации базовой последовательности случайных чисел (СЧ) $\{x_i\} = x_0, x_1, \dots, x_n$, представляющих собой выборку из равномерно распределенной на интервале $[0,1]$ генеральной совокупности значений величины ξ . Рассмотрим основные методы имитации случайных последовательностей.

2.6.1. Генерация случайных чисел

Существует три основных способа генерации случайных чисел (СЧ): аппаратный; табличный; алгоритмический [1, 3, 5, 25].

1. Аппаратный способ. При аппаратном способе случайные числа вырабатываются специальной приставкой (генератором или датчиком случайных чисел), служащей в качестве одного из устройств компьютера. В качестве физического эффекта (процесса), лежащего в основе таких генераторов СЧ, используются шу-

мы в электронных и полупроводниковых приборах, явления распада радиоактивных элементов, фотоэлектрический эффект, квантовые явления и т.д. Эти процессы в теории абсолютно непредсказуемы, на практике же получаемые на их основе случайные числа проверяются с помощью специальных статистических тестов.

Аппаратные генераторы случайных чисел главным образом применяются для проведения статистических испытаний и в криптографии, где они используются для создания криптографических ключей для зашифрованной передачи данных. Также такие аппараты широко используются в интернет-казино для имитации, например, рулетки. Но из-за сложности реализации и относительной медленности использование подобных генераторов зависит от потребностей конкретной предметной области и от устройства самого генератора.

Достоинства аппаратного способа: сравнительно небольшие затраты вычислительных ресурсов; запас СЧ теоретически не ограничен.

Недостатки аппаратного способа: отсутствует возможность проверки, а значит гарантии качества случайной последовательности во время моделирования; невозможно получить одинаковые последовательности СЧ, т.е. воспроизвести последовательность; необходимо дополнительное устройство, стабильность работы которого должна обеспечиваться; относительно медленная работа в сравнении с программным генератором.

2. Табличный способ. Случайные числа оформлены в виде таблицы в оперативной памяти или на внешнем носителе.

Достоинства табличного способа: можно воспроизвести последовательность СЧ.

Недостатки табличного способа: запас чисел ограничен; вычислительные ресурсы используются неэффективно, т.к. необходимо постоянное обращение к внешней или оперативной памяти и таблица занимает место в памяти.

Табличный способ также редко используют.

3. Алгоритмический (программный) способ. Случайные числа формируются на компьютере с помощью специальных алгоритмов и реализующих их программ. Каждое СЧ вычисляется программой по мере возникновения потребности

при моделировании системы на компьютере. Алгоритмический способ наиболее распространен.

Достоинства программного способа: возможность многократного воспроизведения последовательностей; не требуются специальные устройства.

Недостатки программного способа: погрешность в моделировании непрерывных распределений СЧ, вследствие того, что компьютер оперирует с n – рядными числами (т.е. дискретными); периодичность последовательностей СЧ, возникающую в силу их алгоритмической природы; сравнительно большие затраты вычислительных ресурсов.

Программная имитация СЧ. Рассмотрим алгоритмы имитации базовой последовательности случайных чисел $\{x_i\} = x_0, x_1, \dots, x_n$, представляющих собой выборку из равномерно распределенной на интервале $[0,1]$ генеральной совокупности значений случайной величины (СВ) ξ .

Кратко вспомним из курса теории вероятностей как математически описывается СВ [13, 22].

Случайная величина задается с помощью закона распределения вероятностей.

Закон распределения вероятностей дискретной случайной величины – правило сопоставления с каждым возможным значением x_i случайной величины X вероятности его появления $p_i = P\{X = x_i\}$:

$$x_i \rightarrow p_i = P\{X = x_i\}. \quad (2.5)$$

Закон распределения вероятностей непрерывной случайной величины – правило сопоставления с каждой областью Δx возможных значений случайной величины X вероятности события $p(\Delta x) = P\{X \in \Delta x\}$, заключающегося в том, что в эксперименте реализуется одно из возможных значений, принадлежащих заданной области Δx :

$$\Delta x \rightarrow p(\Delta x) = P\{X \in \Delta x\}. \quad (2.6)$$

Способы задания СВ: таблица частот (в дискретном случае), с помощью функций, графически.

Функцией распределения вероятностей $F_X(x)$ СВ X называют функцию, ставящую в соответствие любому заданному значению x величину вероятности события $F_X(x) = P\{X < x\}$, т.е. функция распределения задает вероятность того, что значения СВ меньше заданного значения x [22]. Функцию распределения вероятностей еще называют интегральной функцией. Свойства функции распределения:

Свойство 1. Значения функции распределения принадлежат отрезку $[0,1]$:

$$0 \leq F_X(x) \leq 1. \quad (2.7)$$

Свойство 2. Функция распределения есть неубывающая функция:

$$F_X(x_2) \geq F_X(x_1), \text{ если } x_2 \geq x_1. \quad (2.8)$$

Следствие 1. Вероятность того, что СВ X примет значение, заключенное в интервале (a,b) , равна приращению функции распределения на этом интервале:

$$P(a < X < b) = F_X(b) - F_X(a). \quad (2.9)$$

Следствие 2. Вероятность того, что непрерывная СВ X примет одно определенное значение, например x_1 , равна нулю:

$$P(X = x_1) = 0. \quad (2.10)$$

Свойство 3. Если все возможные значения СВ X принадлежат интервалу (a,b) , то

$$F_X(x) = 0, \text{ при } x \leq a; F_X(x) = 1, \text{ при } x \geq b. \quad (2.11)$$

Следствие. Справедливы следующие предельные соотношения:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow \infty} F_X(x) = 1. \quad (2.12)$$

Свойство 4. Функция распределения непрерывна слева:

$$\lim_{x \uparrow x_0} F_X(x) = F_X(x_0). \quad (2.13)$$

Функция плотности распределения вероятностей (дифференциальная функция) непрерывной СВ X – первая производная от функции распределения:

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x}. \quad (2.14)$$

Вероятность того, что непрерывная СВ примет значение, принадлежащее интервалу (a,b) , определяется равенством:

$$P(a < X < b) = \int_a^b f_X(x) dx. \quad (2.15)$$

Свойство 1. Функция плотности распределения неотрицательна.

Свойство 2. Несобственный интеграл от плотности распределения в пределах от $-\infty$ до ∞ равен единице:

$$\int_{-\infty}^{\infty} f_X(x) dx = 1. \quad (2.16)$$

В записях функции нижний индекс часто опускают, т.е. пишут $F(x)$ и $f(x)$.

Генератор случайных чисел имитирует выборку значений непрерывной случайной величины, распределенной по равномерному закону.

Непрерывная случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$, если на этом отрезке плотность распределения случайной величины постоянна, а вне него равна нулю.

$$f(x) = \begin{cases} 0, & x < a \\ C, & a \leq x \leq b \\ 0, & x > b \end{cases} \quad (2.17)$$

Постоянная величина C может быть определена из условия равенства единице площади, ограниченной кривой распределения (рисунок 2.8).

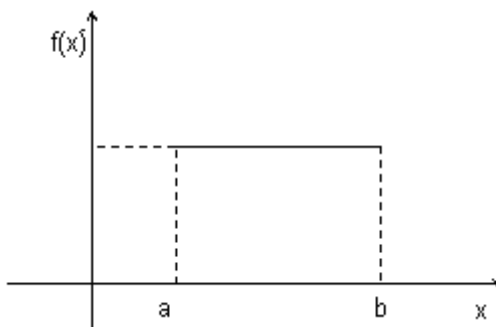


Рисунок 2.8 – График функции плотности распределения

Получаем $C = \frac{1}{b-a}$. Найдем функцию распределения $F(x)$ на отрезке $[a, b]$.

$$F(x) = \int_{-\infty}^x f(x) dx = \int_a^x \frac{1}{b-a} dx = \frac{x}{b-a} \Big|_a^x = \frac{x-a}{b-a}. \quad (2.18)$$

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

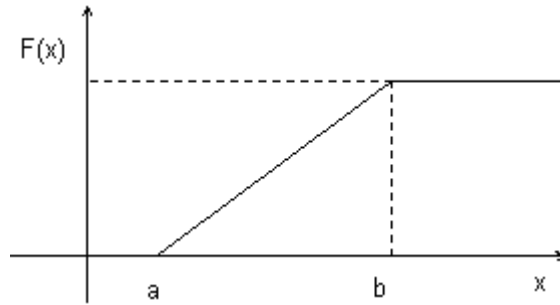


Рисунок 2.9 – Функция распределения

Вероятность попадания случайной величины в заданный интервал:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} \frac{dx}{b-a} = \frac{\beta - \alpha}{b-a}. \quad (2.19)$$

Для того чтобы случайная величина подчинялась закону равномерного распределения необходимо, чтобы ее значения лежали внутри некоторого определенного интервала, и внутри этого интервала значения этой случайной величины были бы равновероятны. Определим математическое ожидание и дисперсию случайной величины, подчиненной равномерному закону распределения:

$$\begin{aligned} M(\xi) &= \int_a^b xf(x)dx = \int_a^b xdx / (b-a) = (a+b)/2; \\ D(\xi) &= \int_a^b (x - M(\xi))^2 f(x)dx = (b-a)^2 / 12. \end{aligned} \quad (2.20)$$

В случае распределения значений СВ на отрезке [0,1]:

$$\begin{aligned} M(\xi) &= \int_0^1 xf(x)dx = \int_0^1 xdx = \frac{1}{2}; \\ D(\xi) &= \int_0^1 (x - M(\xi))^2 f(x)dx = \frac{1}{12} \end{aligned} \quad (2.21)$$

Рассмотрим способ формирования равномерно распределенной СВ ξ на компьютере. Пусть Z – дискретная СВ, принимающая значения:

$$z_i = \begin{cases} 1, p = 0.5; \\ 0, p = 0.5. \end{cases} \quad (2.22)$$

Представим бесконечную последовательность 0 и 1 и будем рассматривать эту последовательность как двоичные знаки некоторого числа ξ :

$$\xi = 2^{-1} z_1 + 2^{-2} z_2 + \dots + 2^{-n} z_n. \quad (2.23)$$

Очевидно, что ξ – СВ в интервале $[0,1]$. Вероятность попадания ξ в интервал $(0 < \xi \leq 0.5)$ равна: $p=0.5$; вероятность попадания ξ в интервал $(0 < \xi \leq 0.25)$ равна: $p=0.25$. Таким образом, вероятность попадания ξ в любой интервал вида $(\frac{k}{2^n} < \xi \leq \frac{k+1}{2^n})$ равна его длине $\frac{1}{2^n}$, $n=1,2,\dots$; $k=0,1,\dots,2^{n-1}$. Следовательно, ξ – равномерно распределенная СВ.

Отсюда, способ формирования ξ . Необходимо взять бесконечную последовательность чисел z_i и считать их двоичными знаками ξ . Так как на компьютере число разрядов ограничено, получить бесконечную последовательность нельзя. В цифровой вычислительной машине числа представляются n двоичными разрядами. Количество несовпадающих чисел, которое можно записать, 2^n . Поэтому вместо непрерывной совокупности СЧ с равномерным законом распределения в качестве исходной берут дискретную совокупность 2^n чисел с одинаковыми вероятностями попадания любого из них. Такое распределение СВ ξ называется *квазиравномерным*. СВ ξ в интервале $[0,1]$ принимает значения:

$$x_i = \frac{i}{2^n - 1}, i = 0,1,\dots,2^{n-1}; p_i = \frac{1}{2^n}. \quad (2.24)$$

Математическое ожидание и дисперсия квазиравномерной СВ ξ соответственно равны:

$$M(\xi) = 1/2;$$

$$D(\xi) = \frac{2^n + 1}{12(2^n - 1)}. \quad (2.25)$$

Последовательность x_i называется псевдослучайной, так как числа имеют квазиравномерное распределение и последовательность сформирована на основе алгоритмов, т.е. по сути – детерминирована.

Требования к идеальному генератору случайных чисел:

1. Последовательность должна состоять из квазиравномерно распределенных чисел.
2. Числа должны быть статистически независимыми.
3. Последовательности случайных чисел должны быть воспроизводимыми. Воспроизводимость СЧ позволяет упростить отладку компьютерной программы и проверить правильно ли она работает; можно использовать идентичные СЧ при моделировании различных структур системы и выполнять их более точное сравнение. Рекомендуются использовать разные потоки (генераторы) для потоков обслуживания и поступления заявок.
4. Последовательности должны иметь неповторяющиеся числа. Генератор должен иметь полный период, т.е. цикл повторения СЧ должен быть равен 2^n .
5. Последовательности должны получаться с минимальными затратами вычислительных ресурсов.
6. В генераторе должен быть предусмотрен простой способ получения отдельных потоков СЧ. Поток – часть последовательности СЧ, производимых генератором, очередной поток начинается там, где заканчивается предыдущий. Можно рассматривать потоки, как отдельные независимые генераторы. Один поток используется для генерации интервалов обслуживания, второй – для интервалов поступления.

Алгоритмы генерации случайных чисел. Наибольшее применение в практике моделирования на компьютере находят алгоритмы вида:

$$x_{i+1} = \Phi(x_i), \quad (2.26)$$

представляющие собой рекуррентные соотношения первого порядка, для которых начальное число x_0 и постоянные параметры заданы.

Одна из первых исторически сложившихся процедур – *метод серединных квадратов*, который был предложен фон Нейманом и Метрополисом в 1940-х годах. Далее приведен пример его использования.

Пример 1. Пусть имеется 2^n – разрядное число меньше 1: x_0 . Возведем его в квадрат: $(x_0)^2$, а затем отберем средние 2^n разрядов. Полученное таким образом число является следующим числом последовательности СЧ: x_1 и т.д.

Пусть $x_0 = 0,2152$, тогда $(x_0)^2=0,04631104$, $x_1 = 0,6311$, $(x_1)^2=0,39828721$, $x_2=0,8287$ и т.д.

Недостаток метода: наличие корреляции между числами последовательности, а иногда случайность вообще отсутствует. Например: $x_0 = 0,4500$, $(x_0)^2=0,20250000$, $x_1 = 0,2500$, $(x_1)^2=0,06250000$, $x_2=0,2500$ и последовательность вырождается в одно число.

Широкое применение получили *конгруэнтные процедуры* генерации псевдослучайных последовательностей, разработанные Лемером.

В основе лежит понятие *конгруэнтности*. Два целых числа α и β конгруэнтны (сравнимы) по модулю m , где m – целое число, тогда и только тогда, когда существует такое целое число k , что $\alpha-\beta=km$, т.е. если разность делится на m и если числа α и β дают одинаковые остатки от деления на абсолютную величину числа m .

Пример 2. $\alpha=9375$ и $\beta=6875$; $m=4$; $k=625$; $9375-6875=2500=4*625$; $9375:4=2343,75$; $6875:4=1718,75$.

Конгруэнтная процедура получения последовательности псевдослучайных чисел реализуется мультипликативным или смешанным методом.

Мультипликативный метод задает последовательность неотрицательных целых чисел $\{X_i\}$ не превосходящих M по формуле:

$$X_{i+1} = \lambda X_i \pmod{M}, \quad (2.27)$$

где λ, X_i, M – неотрицательные целые числа. Выражение (2.27) означает: X_{i+1} равно остатку от деления λX_i на M . В силу детерминированности метода получают воспроизводимые последовательности. В качестве X_0 выбирают произволь-

ное нечетное число; M – число, определяющее наибольшее значение получаемых СЧ, при машинной реализации: $M = p^g$, где p – основание системы счисления, g – число бит в машинном слове.

Для $p=2$ алгоритм сводится к выполнению следующих операций.

1. Вычислить $\lambda = 8t \pm 3$, где t – любое целое положительное число;
2. Найти произведение λX_0 взять g младших разрядов в качестве первого члена последовательности X_1 , а остальные отбросить;
3. Определить дробь $x_1 = X_1 / M$ из интервала $[0,1]$;
4. Присвоить $X_0 = X_1$ и вернуться к п.2.

Пример 3 (для двоичной системы счисления). $X_0=0111$ (7 в десятичной системе счисления); $t=1$; $\lambda=1$ или 5 (пусть $\lambda=0101$); $M=16$, $g=4$.

Вычисление первого числа последовательности x_1 :

$\lambda X_0 = (0101)(0111) = 00100011(35)$; $X_1 = 0011(3)$; $x_1 = 3/16 = 0.1875$ или $x_1 = 7*5(mod\ 16) = 0.1875$.

Вычисление второго числа последовательности x_2 :

$\lambda X_1 = (0101)(0011) = 00001111(15)$; $X_2 = 1111(15)$; $x_2 = 15/16 = 0.9375$ и т.д.

Полученные числа последовательности конгруэнтны.

Смешанный метод задает последовательность неотрицательных целых чисел $\{X_i\}$ не превосходящих M по формуле:

$$X_{i+1} = (\lambda X_i + \mu)(mod M), \quad (2.28)$$

где μ – неотрицательное целое число. С вычислительной точки зрения метод сложнее, но введение дополнительного параметра позволяет уменьшить возможную корреляцию между генерируемыми числами.

Пример 4 (для двоичной системы счисления). $X_0=0111$ (7); $\lambda=0101$ (5); $\mu=0011(3)$; $M=16$, $g=4$.

Вычисление первого числа последовательности x_1 :

$\lambda X_0 + \mu = (0101)(0111) + (0011) = 00100110(38)$; $X_1 = 0110(6)$; $x_1 = 6/16 = 0.375$ или $x_1 = (7*5+3)(mod\ 16) = 0.375$.

Вычисление второго числа последовательности x_2 :

$$\lambda X_1 + \mu = (0101)(0110) + (0011) = 00100001(33); X_2 = 0001(1); x_2 = 1/16 = 0,063 \text{ и т.д.}$$

Для того чтобы конгруэнтный генератор имел полный период равный 2^n , параметры λ, μ, M следует выбирать определенным образом. Правила определения параметров устанавливает теорема, доказанная Халлом и Добеллом [11].

Теорема. Конгруэнтный генератор, описанный формулой (2.28), имеет полный период при соблюдении трех условий:

- единственное положительное число, на которое без остатка делятся как M , так и μ , равно 1;
- если q является простым числом (делится только само на себя и на 1), на которое делится M , то $\lambda - 1$ делится на q без остатка;
- если M делится на 4, то $\lambda - 1$ тоже делится на 4.

В настоящее время практически все библиотеки стандартных программ для вычисления квазиравномерно распределенных СЧ основаны на конгруэнтной процедуре. Применяемые генераторы СЧ перед моделированием должны пройти тщательное предварительное тестирование на равномерность, стохастичность и независимость получаемых последовательностей СЧ.

Методы улучшения качества последовательностей СЧ. Одно из направлений улучшения качества последовательностей СЧ – использование рекуррентных формул порядка r :

$$x_{i+1} = \Phi(x_i, x_{i-1}, \dots, x_{i-r+1}). \quad (2.29)$$

Однако применение этих методов приводит к увеличению затрат вычислительных ресурсов на получение чисел. Приведем некоторые методы.

Метод смешанный (в общем случае):

$$X_{i+1} = (\lambda_1 X_i + \lambda_2 X_{i-1} + \dots + \lambda_p X_{i-p} + \mu) \pmod{M}, \quad x_{i+1} = X_{i+1} / M. \quad (2.30)$$

Параметры:

p – порядок, $X_i, X_{i-1}, \dots, X_{i-p}$ – стартовые значения; $\lambda_1, \lambda_2, \dots, \lambda_p$ – множители, период датчика $T \leq M^p - 1$, т.е. растет с увеличением M и p .

Пример 5:

$$M=16, \lambda_1 = 5, \lambda_2 = 11, \mu = 3, M = 16, p = 2, X_0 = 7.$$

$$X_1=7*5+3=38, X_1 / \text{mod } M = 6, x_1 = 6/16.$$

$$X_2=6*5+7*11+3=110, X_2 / \text{mod } M = 14, x_2 = 14/16.$$

$$X_3=14*5+6*11+3=139, X_3 / \text{mod } M = 11, x_3 = 11/16.$$

Метод, использующий нелинейные рекуррентные формулы (например, квадратичное соотношение):

$$X_{i+1} = (kX_i^2 + \beta X_i + \mu)(\text{mod } M), x_{i+1} = X_{i+1} / M. \quad (2.31)$$

Пример 6:

$$\beta = 5, k = 4, \mu = 3, M = 16, X_0 = 7$$

$$X_1=4*7*7+5*7+3=204, X_1 / \text{mod } M = 12, x_1 = 12/16.$$

Теорема, определяющая оптимальные параметры метода:

Если $M = 2^q, q \geq 2$, то наибольший период равный M достигается тогда и только тогда, когда β, μ – нечетные, k – четно, причем

$$\beta / \text{mod } M = (k + 1) / \text{mod } 4. \quad (2.32)$$

Метод возмущений. В основе метода формула:

$$x_{i+1} = \begin{cases} \Phi(x_i), & \text{если } i \neq 0(\text{mod } M) \\ \psi(x_i), & \text{если } i = 0(\text{mod } M) \end{cases} \quad (2.33)$$

В основном используется формула $\Phi(x_i)$ и только когда i кратно M последовательность «возмущается» и реализуется переход к формуле $\psi(x_i)$.

Целое число M называется периодом возмущения. Пример: $M=16$, для генерации каждого 16 числа используется вторая формула, а 15 чисел генерируются по первой формуле.

Сложные генераторы. Разработаны Мак-Лареном и Марсаль в 1965 г.

Идея метода: один или несколько отдельных генераторов комбинируются определенным образом для генерирования конечных СЧ.

Пример. Генератор А имитирует последовательность из k -случайных чисел (было предложено $k=128$). Генератор В используется для получения СЧ из последовательности А в диапазоне $[1, k]$ и полученное СЧ – первое число последовательности. Генератор А заменяет извлеченное число в последовательности на вновь сгенерированное и процесс продолжается.

В 1975 г. Нансе и Овертриста доказали, что таким способом из двух плохих генераторов СЧ можно получить сложный генератор с хорошими статистическими свойствами. Однако, если исходно два генератора хорошие, то построение

сложного генератора значимого эффекта не дает. Также было доказано, что вектор длиной $k=2$ дает такие же результаты, как более длинные векторы.

Генератор Таусворта. Идея генератора была предложена в 1965 г. Основан на криптографических методах и работают непосредственно с битами для получения СЧ.

Идея: Определим последовательность $b_1, b_2, \dots$ бинарных знаков с помощью формулы:

$$b_i = (c_1 b_{i-1} + c_2 b_{i-2} + \dots + c_q b_{i-q}) \pmod{2}, \quad (2.34)$$

где $c_j = 0$ или 1. Практически во всех приложениях генератора Таусворта только два $c_j \neq 0$ и выражение упрощается:

$$b_i = (b_{i-r} + b_{i-q}) \pmod{2}. \quad (2.35)$$

Сложение по модулю 2 эквивалентно применению к битам оператора исключающего ИЛИ, т.е.

$$b_i = \begin{cases} 0, & b_{i-r} = b_{i-q} \\ 1, & b_{i-r} \neq b_{i-q} \end{cases}. \quad (2.36)$$

Пример 7. Для инициализации b_i необходимо задать первые q значений. Пусть $r=3, q=5, b_1 = b_2 \dots = b_5 = 1$, тогда получаем последовательность:

1111 1000 1101 1101 0100 00

Период битов $31=2^q-1$. Для преобразования к СЧ в диапазоне $[0,1]$ последовательность делится на 2^l . Пусть $l=4$, тогда получаем следующую последовательность СЧ: 15/16, 8/16, 13/16, 13/16, ...

Достоинства генератора:

- генератор не зависит от разрядности компьютера, может быть получен период любой длины (например, $2^{521}-1$ и больше);
- генератор имеет хорошие статистические свойства.

Однако, генераторы Таусворта не нашли широкого применения, видимо это связано с тем, что они были разработаны позже конгруэнтных генераторов.

2.6.2. Генерация случайных событий

Имитация элементарного события. Необходимо реализовать случайное событие A , наступающее с заданной вероятностью p . Определим A как событие,

состоящее в том, что выбранное значение x_i равномерно распределенной на интервале $[0,1]$ случайной величины удовлетворяет неравенству:

$$x_i \leq p. \quad (2.37)$$

Тогда вероятность события A будет $P(A) = \int_0^p dx = p$.

Противоположное событие состоит в том, что $x_i > p$, его вероятность равна $1-p$.

Пример 8. Пусть нам известна вероятность отказа устройства $P_o = 0,1$. Смоделировать выпадение этого события можно, разыграв равномерно распределенное случайное число из диапазона от 0 до 1 и установив, в какой из двух интервалов (от 0 до 0,1 или от 0,1 до 1) оно попало (см. рисунок 2.10). Если число попадает в диапазон $(0; 0,1]$, то устройство отказало, то есть событие произошло, иначе — событие не произошло (устройство работает). При значительном числе экспериментов частота попадания чисел в интервал от 0 до 0,1 будет приближаться к вероятности $P_o = 0,1$, а частота попадания чисел в интервал $(0,1]$ будет приближаться к $P_p = 0,9$.

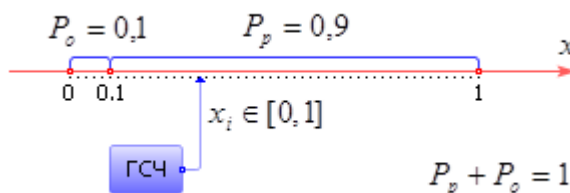


Рисунок 2.10 – Имитационное моделирование простого события

Имитация полной группы событий. Несовместные события составляют полную группу, если суммарная вероятность их появления равна единице.

Рассмотрим группу событий. Пусть $A_1, A_2, \dots, A_n$ — полная группа событий, наступающих с вероятностями $p_1, p_2, \dots, p_n$ соответственно. Определим событие A_{m+1} как событие, состоящее в том, что выбранное значение x_i случайной величины, равномерно распределенной на интервале $[0,1]$ удовлетворяет неравенству:

$$l_m < x_i \leq l_{m+1}, \quad 0 \leq m \leq n-1,$$

где $l_0 = 0$; $l_{m+1} = \sum_{i=1}^{m+1} p_i$.

Процедура моделирования испытаний в этом случае состоит в последовательном сравнении случайных чисел x_i со значениями l_m , l_{m+1} . Если условие выполняется, исходом испытания оказывается событие A_{m+1} .

Описанный алгоритм называют алгоритмом «розыгрыша по жребию» (см. рисунок 2.11).

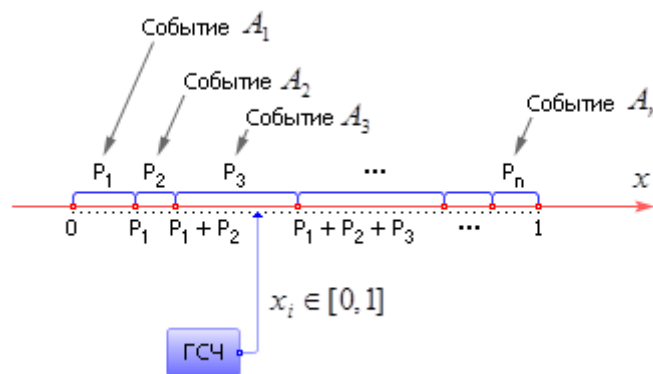


Рисунок 2.11 – Имитационное моделирование полной группы событий

Имитация сложного события, состоящего, например, из двух независимых элементарных событий A и B заключается в проверке неравенств:

$$\begin{cases} x_{i1} \leq p_A \\ x_{i2} \leq p_B \end{cases} \quad (2.37)$$

Здесь x_{i1} и x_{i2} – случайные числа, равномерно распределенные на интервале $[0, 1]$; p_A , p_B – вероятности наступления соответственно событий A и B .

В зависимости от исхода проверки неравенств делается вывод, какой из вариантов сложного события: AB , $\overline{AB}$, $A\overline{B}$, $\overline{A}B$ имеет место.

Имитация зависимых событий. В случае, когда сложное событие состоит из элементарных зависимых событий A и B имитация сложного события производится с помощью проверки следующих неравенств:

$$\left\{ \begin{matrix} x_{i1} \leq p_A \\ x_{i2} \leq p_{B/A} \end{matrix} \right\}; \quad \left\{ \begin{matrix} x_{i1} \leq p_A \\ x_{i2} > p_{B/A} \end{matrix} \right\}; \quad \left\{ \begin{matrix} x_{i1} > p_A \\ x_{i2} \leq p_{B/\bar{A}} \end{matrix} \right\}; \quad \left\{ \begin{matrix} x_{i1} > p_A \\ x_{i2} > p_{B/\bar{A}} \end{matrix} \right\}. \quad (2.38)$$

В зависимости от того, какая из этих четырех систем неравенств выполняется, делается вывод о том, какой из четырех возможных исходов: AB , $\bar{A}\bar{B}$, $\bar{A}B$, $A\bar{B}$ имеет место.

2.6.3. Имитация случайных величин

Моделирование последовательности значений случайных величин с заданным законом распределения реализуется на основе использования последовательности значений СВ, имеющей равномерное распределение в диапазоне $[0, 1]$. Разработан ряд методов моделирования значений непрерывных и дискретных СВ. Причем, разработаны как универсальные методы моделирования, на основе которых можно смоделировать разные законы, так и специальные, предназначенные для моделирования последовательности значений СВ, заданной определенным законом.

Рассмотрим универсальные методы моделирования непрерывной СВ.

Метод обратной функции.

Дано: Задан закон распределения СВ η в виде функции распределения вероятностей $F_\eta(y)$ и функции плотности распределения вероятностей $f_\eta(y)$.

Найти: последовательность значений СВ η : $y_1, y_2, \dots, y_n$.

Решение: Метод основан на теореме: если моделируемая СВ η имеет плотность распределения $f(y)$, то распределение СВ ξ :

$$\xi = F(y) = \int_0^y f(y) dy. \quad (2.39)$$

является равномерным в диапазоне $[0, 1]$.

Чтобы получить число, принадлежащее последовательности случайных чисел $\{y_i\}$, имеющих функцию плотности $f(y)$, необходимо разрешить относительно y_i уравнение:

$$x_i = \int_0^{y_i} f(y) dy, \quad (2.40)$$

где x_i – СЧ, равномерно распределенное на $[0,1]$. Для этого необходимо вычислить обратную функцию моделируемого распределения от x_i :

$$y_i = F^{-1}(x_i). \quad (2.41)$$

На рисунке 2.12 приведена графическая иллюстрация метода обратной функции.

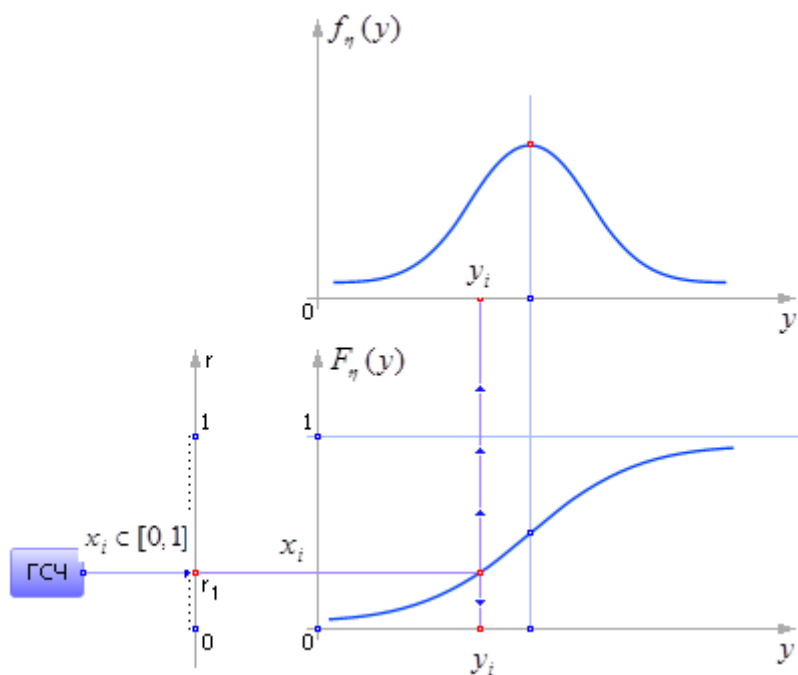


Рисунок 2.12 – Иллюстрация метода обратной функции

Этот способ получения СЧ с заданным законом распределения имеет ограниченную сферу применения, так как для многих законов распределения интеграл не берется, т.е. приходится прибегать к численным методам решения, что увеличивает затраты вычислительных ресурсов на получение каждого числа. Даже для случаев, когда интеграл берется в конечном виде получаются формулы, содержа-

щие действия логарифмирования, извлечения корня и т.д., что также резко увеличивает затраты машинного времени на получение каждого СЧ.

Пример 9. Необходимо получить последовательность случайных чисел с экспоненциальным законом распределения. Пусть, например, моделируется интервал времени между поступлением заявок на обслуживание. Функция плотности распределения интервала, заданная экспоненциальным законом, имеет вид:

$$f(t) = \lambda e^{-\lambda t}.$$

Выполним преобразования по методу обратной функции:

$$x_i = \int_0^t \lambda e^{-\lambda t} dt$$

$$x_i = 1 - e^{-\lambda t}$$

$$\ln(1 - x_i) = \ln e^{-\lambda t}$$

$$t = -\frac{1}{\lambda} \ln(1 - x_i)$$

Так как x_i – случайное число, имеющее равномерное распределение на $[0,1]$, то СЧ $(1 - x_i)$ также распределено равномерно на $[0,1]$. Поэтому, для моделирования последовательности значений СВ по экспоненциальному закону можно использовать выражение:

$$t = -\frac{1}{\lambda} \ln x_i.$$

На рисунках 2.13 и 2.14 приведены соответственно значения СВ и гистограмма с наложением функции плотности распределения экспоненциального закона.

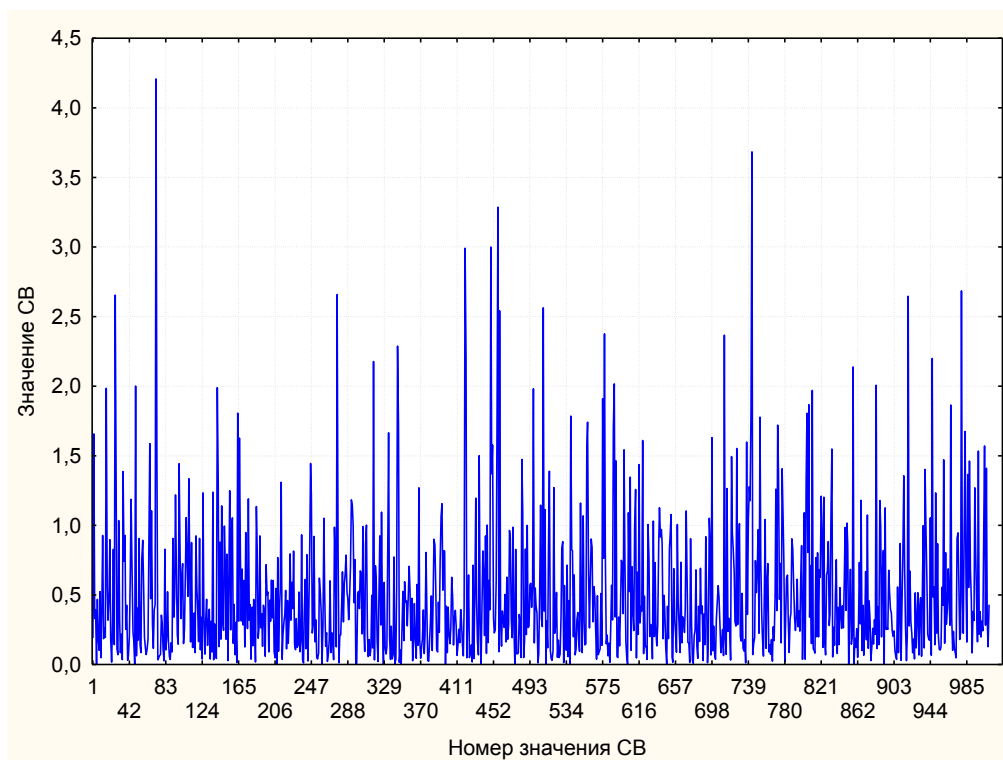


Рисунок 2.13 – График значений СВ

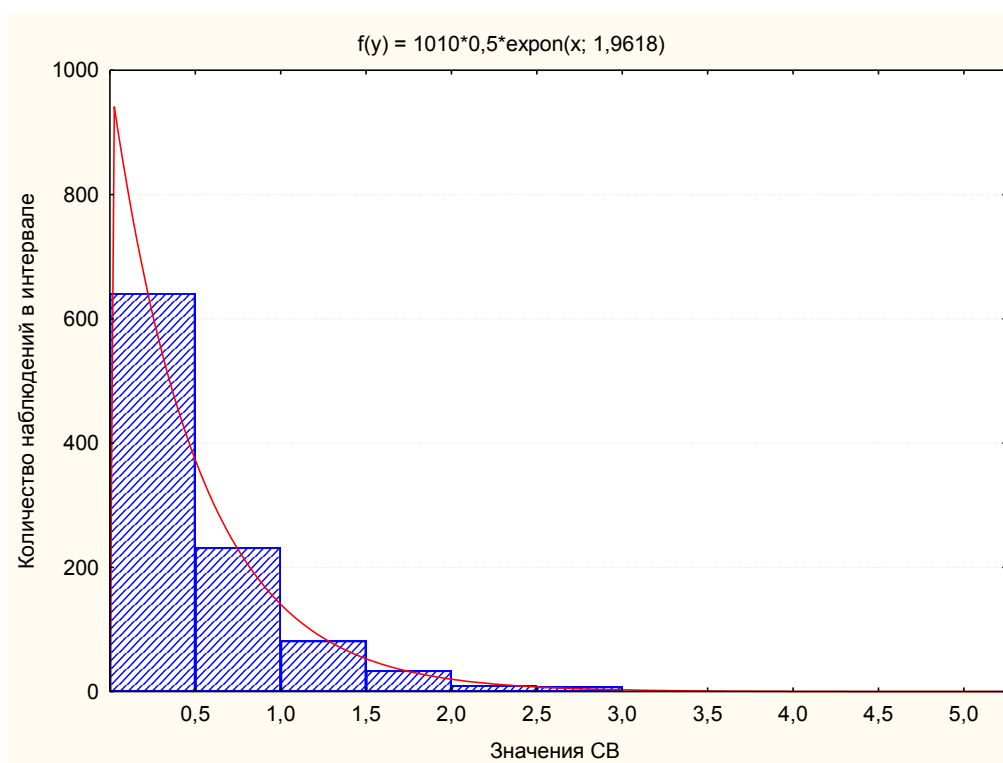


Рисунок 2.14 – Гистограмма СВ с наложением функции плотности распределения экспоненциального закона

Метод, основанный на кусочной аппроксимации функции плотности распределения.

Дано: Задан закон распределения СВ η в виде функции распределения вероятностей $F_\eta(y)$ и функции плотности распределения вероятностей $f_\eta(y)$.

Найти: последовательность значений СВ η : $y_1, y_2, \dots, y_n$.

Решение. СВ η задана на интервале (a, b) . Разобьем интервал (a, b) на m интервалов, таким образом, чтобы вероятность попадания СВ в любой интервал $(a_k; a_{k+1})$ была постоянной, т.е. (см. рисунок 2.15):

$$\int_{a_k}^{a_{k+1}} f_\eta(y) dy = 1/m. \quad (2.42)$$

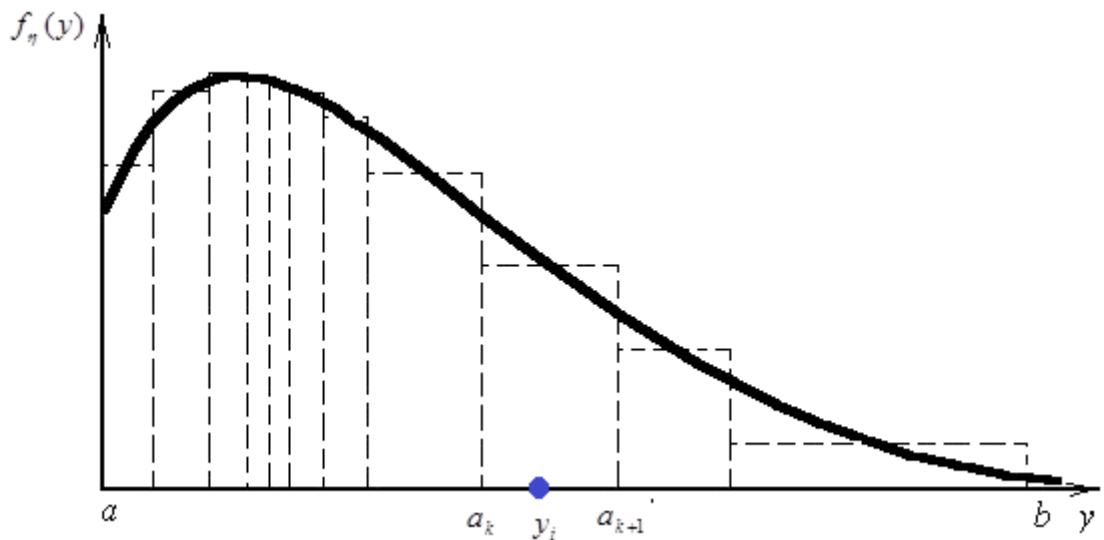


Рисунок 2.15 – Иллюстрация метода кусочной аппроксимации

В таком случае, алгоритм этого метода получения последовательности значений СВ сводится к выполнению следующих действий:

- генерируется СЧ x_{i1} , равномерно распределенное на интервале $[1, m]$;
- с помощью этого числа выбирается один из m интервалов: $(a_k; a_{k+1})$;
- генерируется СЧ x_{i2} , равномерно распределенное на интервале $[0, 1]$;
- вычисляется СЧ $y_i = a_k + x_{i2}(a_{k+1} - a_k)$ с требуемым законом распределения, т.е. выполняется операция приведения к интервалу $(a_k; a_{k+1})$.

Полученное СЧ y_i является искомым значением.

К достоинствам этого метода относятся сравнительно небольшие затраты вычислительных ресурсов при получении последовательности СЧ на компьютере, так как операция масштабирования (нахождения границ интервалов) выполняется один раз до начала моделирования. Количество вычислительных операций не зависит от точности аппроксимации, т.е. от количества интервалов разбиения m .

Недостаток метода: погрешность за счет замены непрерывной функции плотности распределения вероятностей – кусочно-постоянной. По-сути, метод моделирует последовательность значений СВ, заданной кусочно-постоянной функцией. Вычислительную погрешность метода можно уменьшить за счет увеличения количества интервалов. Для стандартных законов определено оптимальное число интервалов разбиения. Например, для экспоненциального закона – 24 интервала, для нормального закона – 25 интервалов. В GPSS описаны кусочно-постоянные функции для наиболее известных законов.

Метод Неймана. Метод Неймана является приближенным универсальным, применим только для моделирования СВ, заданных усеченным законом.

Дано: Задан закон распределения СВ η в виде функции распределения вероятностей $F_\eta(y)$ и функции плотности распределения вероятностей $f_\eta(y)$.

Найти: последовательность значений СВ η : $y_1, y_2, \dots, y_n$.

Решение: На рисунке 2.16 изображена функция плотности распределения СВ η , заданная на интервале (a, b) . Максимальное значение функции $f_\eta(y)$ обозначено W .

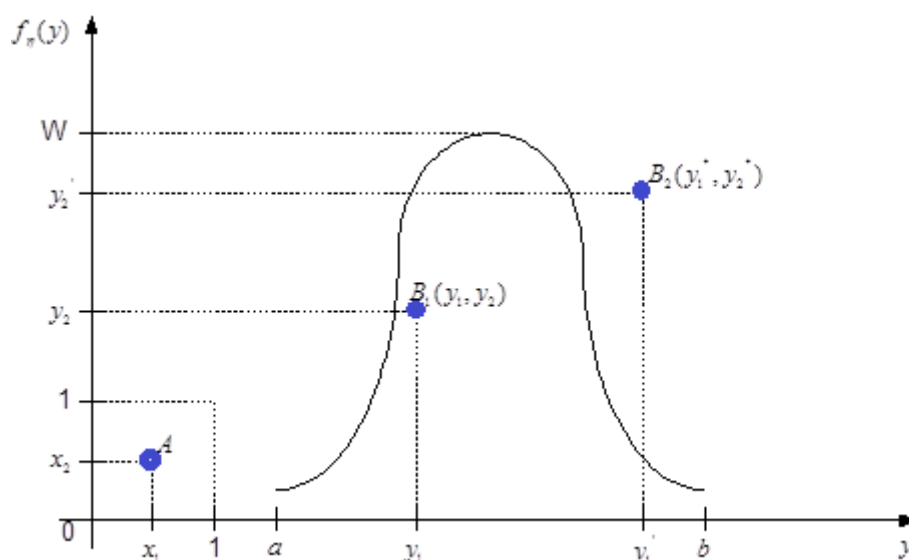


Рисунок 2.16 – Иллюстрация метода Неймана

Алгоритм метода сводится к следующему:

- с помощью датчика СЧ, равномерно распределенных на интервале $[0,1]$, выбирают пары чисел (x_1, x_2) (точка A на рисунке 2.16);
- формируется преобразованная пара чисел, равномерно распределенных на интервале соответственно $[a, b]$ и $[0, W]$: $y_1 = a + x_1(b - a)$; $y_2 = Wx_2$;
- проверяется выполнение неравенства: $y_2 \leq f(y_1)$.

Если неравенство выполнено, то y_1 и есть искомое значение СВ η . На рисунке 2.16 это соответствует первой координате точки B_1 (точка B_1 попадает под купол функции). В противном случае СЧ отбрасываются (например, точка B_2). Далее вновь генерируются СЧ (x_1, x_2) и алгоритм повторяется.

К достоинствам метода относится его сравнительная вычислительная простота. К недостаткам:

- метод применим для моделирования СВ, заданных только усеченным законом;
- часть сгенерированных значений отбрасывается. Количество отброшенных значений зависит от формы функции плотности распределения вероятностей. Чем больше часть площади прямоугольника со сторонами W и $b-a$, лежащая ниже кри-

вой функции плотности распределения вероятностей, тем меньше значений отбрасывается.

Методы генерации СВ, заданной нормальным законом распределения вероятностей.

Нормальным называется распределение вероятностей непрерывной случайной величины, которое описывается функцией плотности вероятностей:

$$f(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} \quad (2.43)$$

Нормальный закон распределения также называется законом Гаусса. Нормальный закон распределения занимает центральное место в теории вероятностей. Это обусловлено тем, что этот закон проявляется во всех случаях, когда случайная величина является результатом действия большого числа различных факторов. К нормальному закону приближаются все остальные законы распределения. Параметры m_y и σ_y , входящие в плотность распределения являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины. Функция распределения $F(y)$:

$$F(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} dy \quad (2.44)$$

Для моделирования последовательности значений СВ по нормальному закону разработано несколько специальных методов (алгоритмов).

Первый алгоритм основан на центральной предельной теореме (см. пункт 2.7). В основе алгоритма сложение СЧ, равномерно распределенных на интервале $[0,1]$:

$$y_i = \frac{1}{\sqrt{k/12}} \sum_{i=1}^k x_i - \frac{k}{2}, \quad (2.45)$$

где x_i – случайные числа, равномерно распределенные от 0 до 1. Наиболее удобно использовать значение $k=12$, т.е.

$$y_i = \sum_{i=1}^{12} x_i - 6. \quad (2.46)$$

Значения y_i будут иметь нормальное распределение с нулевым математическим ожиданием и единичной дисперсией. Для того, чтобы получить значения СВ, заданной нормальным законом с отличным от нуля математическим ожиданием m_y^* и σ_y^* используют преобразование:

$$y_i^* = \sigma_y^* y_i + m_y^*. \quad (2.47)$$

Однако этот алгоритм требует относительно больших затрат машинного времени и дает хорошее приближение к нормальному закону при генерации тысяч СЧ.

Второй алгоритм основан на известном факте, что распределение произведения двух независимых случайных величин, одна из которых имеет релеевское распределение, а другая распределена по закону арксинуса, является нормальным. Это позволяет использовать следующее выражение:

$$y_i = \sigma_y \sqrt{-2 \ln x_{i1}} \cdot \sin(2\pi \cdot x_{i2}), \quad (2.48)$$

где x_{i1}, x_{i2} – случайные числа, равномерно распределенные от 0 до 1. Значения

y_i будут иметь нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

Третий алгоритм заключается в следующем.

1. Генерируются два СЧ равномерно распределенные в интервале от 0 до 1:

x_{i1}, x_{i2} . Эти СЧ определяют координаты случайных точек на плоскости от 0 до 1.

2. СЧ преобразуются следующим образом:

$$x_{i1}^* = 2x_{i1} - 1;$$

$$x_{i2}^* = 2x_{i2} - 1.$$

В результате преобразования СЧ равномерно распределены от -1 до 1 (теперь числа определяют координаты случайных точек на плоскости от -1 до 1).

3. Вычисляется величина $d = (x_{i1}^*)^2 + (x_{i2}^*)^2$ - квадрат длины случайного вектора на плоскости.
4. Если выполняется условие $d \leq 1$ (проверка, вписывается ли случайный вектор в окружность с радиусом, равным 1), то вычисляется величина:

$$c = \frac{\sqrt{-2 \ln d}}{d}$$
. Если нет, то выполняется возврат к пункту 1.
5. Случайные величины $y_{i1} = c \cdot x_{i1}^*$ и $y_{i2} = c \cdot x_{i2}^*$ будут иметь нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

Достоинством алгоритма является то, что одновременно формируются два нормально распределенных СЧ: y_{i1} и y_{i2} .

На рисунках 2.17 и 2.18 приведены соответственно значения СВ и гистограмма с наложением функции плотности распределения нормального закона.

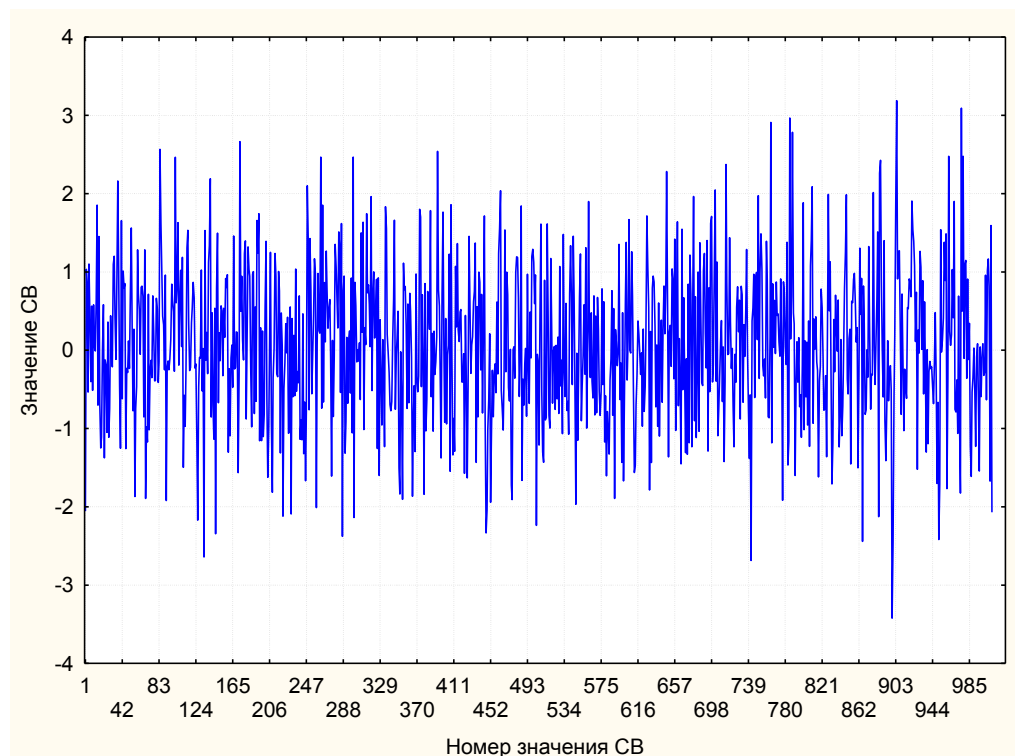


Рисунок 2.17 – График значений СВ

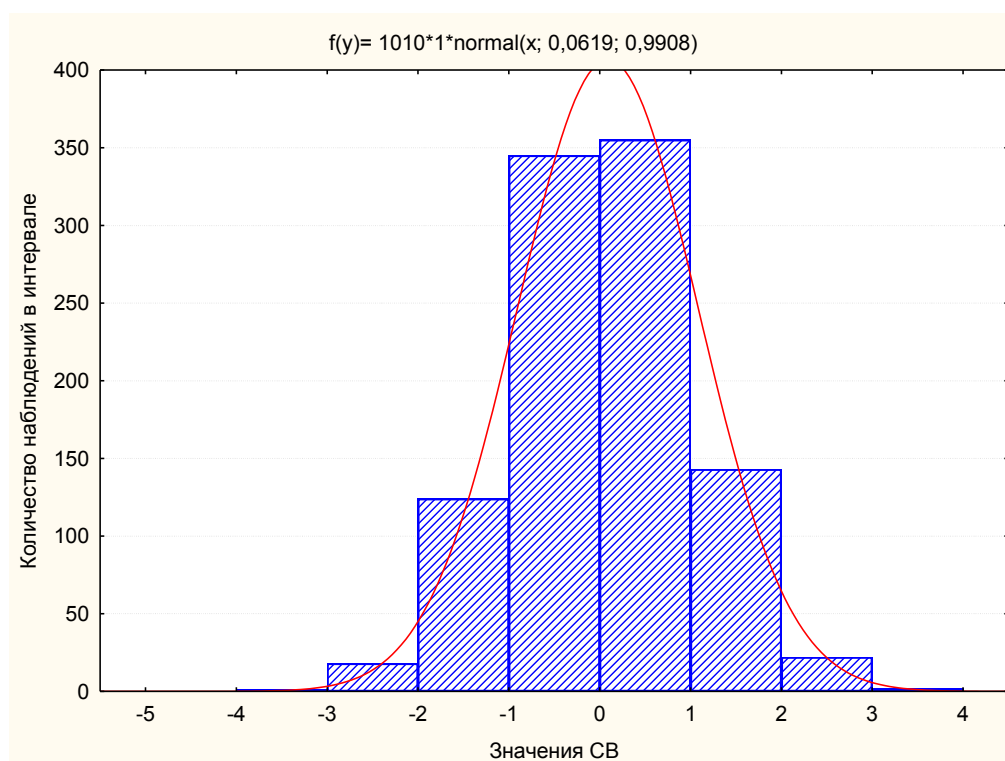


Рисунок 2.18 – Гистограмма СВ с наложением функции плотности распределения нормального закона

Описание разных методов генерации значений случайной величины по заданному закону распределения приведено в [3, 5, 9, 25, 28-29].

2.6.4. Имитация случайных векторов

Метод аналитических преобразований. Пусть система (вектор) непрерывных случайных величин $(\eta_1, \eta_2, \dots, \eta_n)$ задана условным законом распределения СВ $(\eta_i, i = \overline{1, n})$. По теореме умножения плотностей распределения совместная функция плотности распределения вероятностей случайного вектора $(\eta_1, \eta_2, \dots, \eta_n)$ может быть представлена в виде произведения [22]:

$$f(y_1, y_2, \dots, y_n) = f_1(y_1) \cdot f_2(y_2 | y_1) \cdot f_3(y_3 | y_1, y_2) \cdot \dots \cdot f_n(y_n | y_1, y_2, \dots, y_{n-1}). \quad (2.49)$$

Для системы двух случайных величин (η_1, η_2) алгоритм получения вектора ее значений сводится к следующему.

1. Вычисление частной функции плотности для η_1 :

$$f_1(y_1) = \int_{-\infty}^{+\infty} f_1(y_1, y_2) dy_2. \quad (2.50)$$

2. Получение значения $y_1 = \bar{y}_1$ в соответствии с функцией плотности $f_1(y_1)$ согласно любому методу генерации значений СВ, например, метод обратной функции или метод Неймана.
3. Вычисление частной функции плотности для второй компоненты η_2 системы. Она может быть получена на основании теоремы умножения законов распределения [22]:

$$f_2(y_2 | \bar{y}_1) = \frac{f(\bar{y}_1, y_2)}{f_1(\bar{y}_1)}. \quad (2.51)$$

4. Получение значения $y_2 = \bar{y}_2$ любым известным методом в соответствии с найденным законом ее распределения.

Алгоритм может быть обобщен для любого n . Однако использование этого метода связано с вычислительными трудностями, за исключением тех редких случаев, когда интегралы берутся. Поэтому на практике, как правило, применяют другие методы.

Метод разложения по координатным случайным величинам.

Пусть система СВ задана в рамках теории корреляций – математическими ожиданиями $(m_1, m_2, \dots, m_n)$ и матрицей корреляционных моментов:

$$K = \begin{vmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ k_{31} & k_{32} & \dots & k_{3n} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{vmatrix}, \quad (2.52)$$

где $k_{ij} = k_{ji}, j = i = 1, 2, \dots, n$.

Значения $(\eta_i, i = \overline{1, n})$ можно получить с помощью их разложения по координатам системы СВ $\xi_i, i = \overline{1, n}$:

$$\begin{aligned}
\eta_1 &= c_{11}\xi_1 + m_1, \\
\eta_2 &= c_{12}\xi_1 + c_{22}\xi_2 + m_2, \\
&\dots \\
\eta_n &= c_{1n}\xi_1 + c_{2n}\xi_2 + \dots + c_{nn}\xi_n + m_n,
\end{aligned} \tag{2.53}$$

где $\xi_i, i = \overline{1, n}$ - некоррелированные, центрированные, нормированные, нормально распределенные СВ.

Коэффициенты $c_{ij} (i, j = \overline{1, n}; i \leq j)$ могут быть достаточно просто получены решением системы уравнений:

$$k_{ij} = c_{1i}c_{1j} + c_{2i}c_{2j} + \dots + c_{ii}c_{ij}, (j, i = \overline{1, 2, \dots, n}; i \leq j). \tag{2.54}$$

Алгоритм получения значений системы непрерывных СВ сводится к следующему.

1. Решение системы уравнений (2.54).
2. Получение n значений $\xi_i, i = \overline{1, n}$ нормированных, центрированных СВ, распределенных нормально.
3. Вычисление значений $\eta_i, i = \overline{1, n}$ в соответствии с формулами (2.53).

Имитационное моделирование значений системы дискретных случайных величин.

Пусть компоненты случайного двумерного, дискретного вектора принимают значения соответственно a_i, b_j , где $i = \overline{1, n}; j = \overline{1, m}$. Дискретный вектор (η_1, η_2) задается матрицей вероятностей $\|P_{ij}\|$, P_{ij} – вероятность совместного появления i -го и j -го значения соответственно первой и второй компоненты, причем $\sum_{i=1}^n \sum_{j=1}^m P_{ij} = 1$. Моделирование двумерного дискретного случайного вектора производится по следующему алгоритму.

1. Вычисляют суммы $q_i = \sum_{j=1}^m P_{ij}, l_k = \sum_{i=1}^n q_i, k = \overline{1, n}$.
2. Пусть x – равномерно распределенное случайное число из интервала $[0, 1]$ такое, что $l_{k-1} \leq x \leq l_k$. Принимают, что первая компонента η_1 получила k -ое значение.

3. Выбирают k -ю строку в матрице $\|P_{ij}\|$, вычисляют

$$r_s = \sum_{j=1}^s P_{kj}, s = \overline{1, m}; r_0 = 0.$$

4. Если вновь полученное с помощью ДСЧ x такое, что $r_{s-1} < x \cdot r_m < r_s$, то считают, что вторая компонента случайного вектора η_2 получила s -ое значение.

Замечание. В алгоритме используется правило розыгрыша по жребью, однако надо иметь в виду, что $r_m \neq 1$. Обоснование этого алгоритма можно найти в [24].

2.6.5. Имитация случайных процессов

Случайной функцией называется функция, ординаты которой для любых фиксированных значений аргумента являются случайными величинами. Задачу моделирования случайных функций в общем случае нельзя свести к имитации СВ для каждого значения аргумента, так как между ординатами в общем случае существует корреляционная зависимость. Случайная функция, аргументом которой является время t , носит название случайного процесса (СП).

Целью имитационного моделирования СП на компьютере является воспроизведение различного рода сигналов и помех, математической моделью которых является СП. Нужно иметь в виду, что воспроизведение на компьютере процессов с непрерывным временем невозможно ввиду дискретной природы компьютера. Задача моделирования СП в дальнейшем понимается как задача разработки алгоритма, позволяющего формировать на компьютере реализации СП.

СП считается заданным, если задана функция дисперсии $d(t)$, математического ожидания $m(t)$ и корреляционная функция $k(t_i, t_j)$. Эти функции являются неслучайными, их определяют путем обработки опытных данных методами математической статистики.

Имитация нестационарных случайных процессов. Описываемый ниже алгоритм пригоден как для стационарных, так и нестационарных СП. Он предложен В.С. Пугачевым, называется методом канонических разложений и заключается в следующем.

Обозначим дискретные значения реализации СП через $F(t_1), F(t_2), \dots, F(t_n)$, где t принимает значения на конечном интервале времени $t = t_1, t_2, \dots, t_n$. В соответствии с методом:

$$\begin{aligned} F(t_1) &= m(t_1) + x_1 \phi_1(t_1), \\ F(t_2) &= m(t_2) + x_1 \phi_1(t_2) + x_2 \phi_2(t_2), \\ &\dots \\ F(t_n) &= m(t_n) + x_1 \phi_1(t_n) + x_2 \phi_2(t_n) + \dots + x_n \phi_n(t_n). \end{aligned} \quad (2.55)$$

Здесь $x_1, x_2, \dots, x_n$ – значения случайных некоррелированных, центрированных СВ ξ_i с заданным законом распределения; $\phi_i(t_k)$ – координатные функции, обладающие свойствами:

1. $\phi_i(t_j) = 0$, при $i > j$.
2. $\phi_i(t_i) = 1$.

Координатные функции и дисперсии D_i величин ξ_i можно вычислить в соответствии с рекуррентными уравнениями:

$$\begin{aligned} \phi_i(t_j) &= \frac{k(t_i, t_j) - \sum_{k=1}^{i-1} D_k \cdot \phi_k(t_i) \phi_k(t_j)}{D_i}, \\ D_i &= k(t_i, t_i) - \sum_{k=1}^{i-1} [\phi_k(t_i)]^2 D_k, i = \overline{1, n} \end{aligned} \quad (2.56)$$

Алгоритм имитации СП на интервале $[t_1, t_n]$ сводится к следующему.

1. Вычисляются значения координатных функций и дисперсий ξ_i в соответствии с формулами (2.56).
2. Получают значения СЧ $x_1, x_2, \dots, x_n$.
3. Производят вычисления по формулам (2.55).

Имитация стационарных случайных процессов. Для стационарных СП справедливы соотношения $m_i = m$; $d(t) = \sigma^2$; $k(t_i, t_j) = k(\tau)$, где $\tau = t_i - t_j$. Одним из методов имитации стационарных СП заключается в вычислении $F(t_i)$ по формулам:

$$\begin{aligned} F(t_1) &= m + c_1 x_1 + c_2 x_2 + \dots + c_n x_n, \\ F(t_2) &= m + c_1 x_2 + c_2 x_3 + \dots + c_n x_{n+1}, \\ &\dots \\ F(t_n) &= m + c_1 x_n + c_2 x_{n+1} + \dots + c_n x_{2n-1}. \end{aligned} \quad (2.57)$$

Здесь x_i – значения некоррелированных случайных величин ξ_i , для которых $M[\xi_i] = 0$, $D[\xi_i] = \sigma^2$, закон их распределения задан. Коэффициенты c_j ($j = \overline{1, n}$) вычисляются решением уравнений:

$$K(t_k - t_1) = (c_1 c_k + c_2 c_{k+1} + \dots + c_{n-k+1} c_n) \sigma, \quad k = \overline{1, n}. \quad (2.58)$$

Имитационное моделирование стационарных нормальных случайных процессов. Рассмотренные выше методы пригодны для моделирования СП, заданных на конечном интервале времени. При формировании реализаций большой длины эти методы трудоемки, что затрудняет их использование. На практике приходится моделировать СП, относящиеся к узкому классу СП, например, стационарный нормальный СП; стационарный СП, порожденный нормальным, нестационарным; СП со стационарными приращениями и т.д. Для таких классов СП существуют достаточно эффективные моделирующие алгоритмы [1].

В их основу положены линейные преобразования стационарной последовательности x_k независимых нормальных случайных чисел (белый шум) в последовательность $F(t_k)$, $k = 1, 2, \dots$; $t_k - t_{k-1} = \Delta t = \text{const}$, коррелированную по заданному закону. При этом оператор линейного преобразования записывается либо в виде формулы скользящего суммирования с некоторым весом d_i :

$$F(t_k) = \sum_{i=1}^N d_i x_{k-i}, \quad (2.59)$$

либо как рекуррентное уравнение вида:

$$F(t_k) = \sum_{i=0}^l d_i x_{k-i} - \sum_{j=1}^m b_j F(t_{k-j}), \quad k = 1, 2, \dots \quad (2.60)$$

Коэффициенты d_i, b_j в обеих формулах и их количество зависят от вида корреляционной функции. Первая из приведенных формул является математической моделью цифрового фильтра, называемого нерекурсивным; вторая – моделью рекурсивного цифрового фильтра. Имитация СП сводится к вычислению значений коэффициентов d_i, b_j по значениям корреляционной функции и вычислениям по выше приведенным формулам.

2.7. Метод статистического моделирования

Статистическое моделирование представляет собой метод получения с помощью компьютера статистических данных о процессах, происходящих в моделируемой системе. Это основной метод, применяемый в ходе имитационного моделирования стохастической системы.

Суть метода сводится к построению моделирующего алгоритма процесса функционирования системы, имитирующего поведение и взаимодействие элементов системы с учетом случайных входных воздействий и воздействий внешней среды и реализации этого алгоритма с использованием программно-технических средств. В его основе лежит метод статистических испытаний Монте – Карло [13], который базируется на использовании СЧ, распределенных в соответствии с заданным законом распределения.

Результат статистического моделирования – серия частных значений искомых величин или функций, их статистическая обработка.

Теоретической основой метода статистического моделирования служат предельные теоремы теории вероятностей [1, 22]. Принципиальное значение предельных теорем – гарантия высокого качества статистических оценок при числе испытаний (реализаций) $N \rightarrow \infty$. Практически приемлемые результаты часто могут быть получены уже при сравнительно небольших N .

Теорема Бернулли. Если проводится N независимых испытаний, в каждом из которых некоторое событие A осуществляется с вероятностью p , то относи-

тельная частота появления события m/N при $N \rightarrow \infty$ сходится по вероятности к p , т.е. при любом $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} P(|m/N - p| \geq \varepsilon) = 0, \quad (2.61)$$

где m - число положительных исходов испытания.

Теорема Пуассона. Если проводится N независимых испытаний, в каждом из которых некоторое событие A осуществляется с вероятностью p_i , то относительная частота появления события m/N при $N \rightarrow \infty$ сходится по вероятности к среднему из вероятностей p_i , т.е. при любом $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} P\left(\left|m/N - \frac{1}{N} \sum_{i=1}^N p_i\right| \geq \varepsilon\right) = 0. \quad (2.62)$$

Теорема Чебышева. Если в N независимых испытаниях наблюдаются значения $\{x_i\} = x_0, x_1, \dots, x_n$ СВ ξ , то при $N \rightarrow \infty$ среднее арифметическое значений СВ сходится по вероятности к ее математическому ожиданию a , т.е. при любом $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} P\left(\left|\frac{1}{N} \sum_{i=1}^N x_i - a\right| \geq \varepsilon\right) = 0. \quad (2.63)$$

Обобщенная теорема Чебышева. Если $\xi_1, \xi_2, \dots, \xi_N$ - независимые СВ с математическими ожиданиями $a_1, a_2, \dots, a_N$ и дисперсиями $D_1, D_2, \dots, D_N$, ограниченные сверху одним и тем же числом, то при $N \rightarrow \infty$ среднее арифметическое значений СВ сходится по вероятности к среднему арифметическому их математических ожиданий:

$$\lim_{N \rightarrow \infty} P\left(\left|\frac{1}{N} \sum_{i=1}^N x_i - \frac{1}{N} \sum_{i=1}^N a_i\right| \geq \varepsilon\right) = 0. \quad (2.64)$$

Теорема Маркова. Обобщенная теорема Чебышева справедлива и для зависимых случайных величин, если

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} D\left[\sum_{i=1}^N x_i\right] = 0. \quad (2.65)$$

Совокупность теорем, устанавливающих устойчивость средних показателей, называют *законом больших чисел*.

Центральная предельная теорема. Если $\xi_1, \xi_2, \dots, \xi_N$ – независимые одинаково распределенные СВ, имеющие математическое ожидание a и дисперсию D , то при $N \rightarrow \infty$ закон распределения суммы $\sum_{i=1}^N x_i$ неограниченно приближается к нормальному:

$$\lim_{N \rightarrow \infty} P \left\{ \alpha < \left(\sum_{i=1}^N x_i - Na \right) / \sqrt{ND} < \beta \right\} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2} dt. \quad (2.66)$$

2.8. Статистическая обработка результатов моделирования

В ходе статистического моделирования формируются выборочные данные, описывающие процесс и результаты функционирования системы. Выборочные данные являются исходным статистическим материалом для нахождения приближенных значений (оценок) показателей эффективности функционирования изучаемой системы.

Оценка статистических характеристик. По результатам имитационного моделирования рассчитывают оценки функциональных характеристик и оценки выборочных числовых характеристики показателей эффективности системы: оценка математического ожидания (выборочное среднее); выборочная медиана; выборочная мода; оценка дисперсии; оценка среднеквадратического отклонения; выборочные минимальное и максимальное значения; выборочный размах; коэффициент вариации.

Пусть $y_i, i = \overline{1, n}$ – выборочные данные (наблюдения), полученные в результате проведения n экспериментов с имитационной моделью системы.

Например, в качестве y_i могут выступать: время пребывания заявки в системе; время нахождения заявки в очереди на обслуживание; время обработки заявки и ряд других характеристик.

В терминах математической статистики показатель Y – случайная величина, а y_i – выборочные значения СВ, полученные в результате n наблюдений (измерений) СВ Y .

Оценка функциональных характеристик. По выборочным данным может быть построена эмпирическая функция распределения вероятностей и эмпирическая функция плотности распределения вероятностей, которые являются выборочными аналогами соответствующих теоретических характеристик.

Эмпирическая функция распределения вероятностей (кривая накопленных частот) – это функция $\hat{F}(y)$, определяющая для каждого значения y относительную частоту события:

$$\hat{F}(y) = \frac{n_y}{n}, \quad (2.67)$$

где n_y – число значений, меньших y ; n – объем выборки.

На рисунок 2.19 приведена эмпирическая функция распределения с наложением экспоненциального закона.

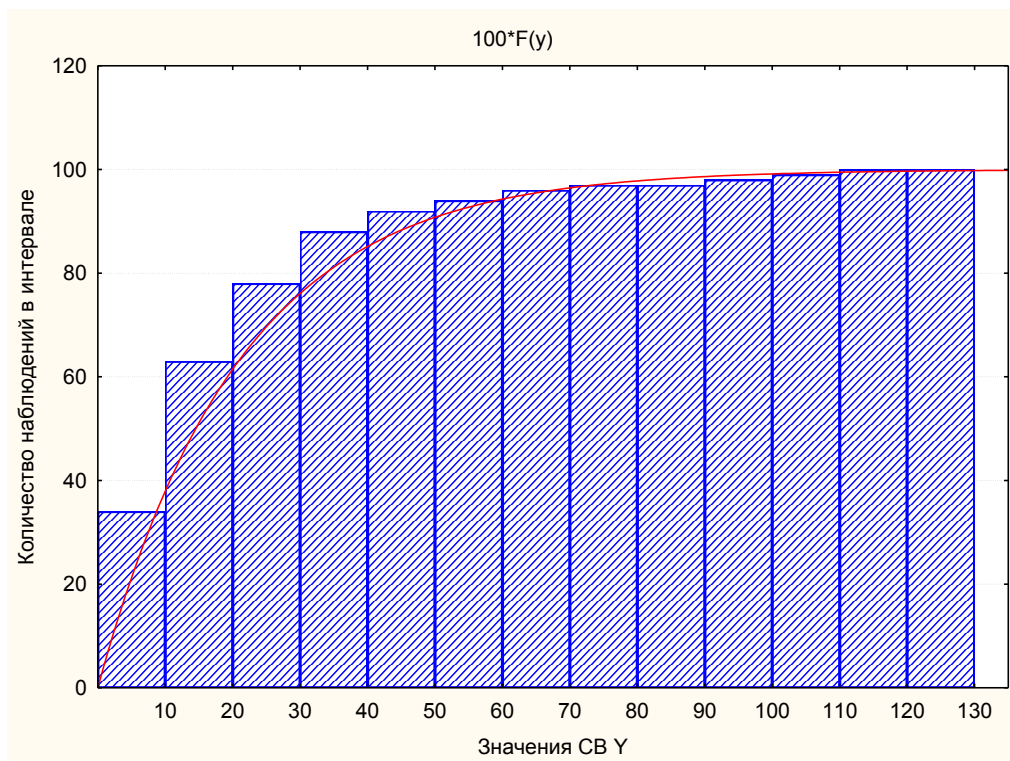


Рисунок 2.19. – Пример эмпирической функции распределения

Эмпирическая функция плотности распределения вероятностей (гистограмма) определяется по формуле:

$$\hat{F}(y) = \frac{n_j}{n \cdot h}, \quad (2.68)$$

где n_j – число наблюдений, h – ширина интервала группирования.

На рисунке 2.20 приведена эмпирическая функция плотности распределения с наложением экспоненциального закона.

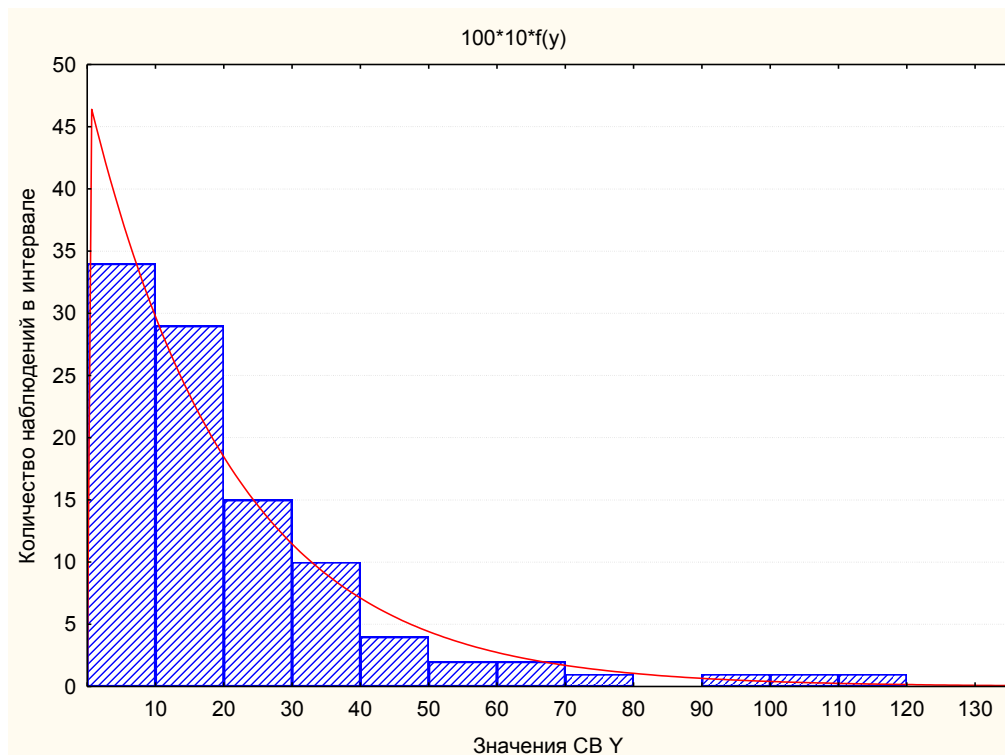


Рисунок 2.20. – Пример гистограммы

Теоретическое математическое ожидание непрерывной СВ Y , возможные значения которой принадлежат всей числовой оси, определяется по формуле:

$$M(Y) = \int_{-\infty}^{+\infty} yf(y)dy, \quad (2.69)$$

где $f(y)$ – функция плотности распределения вероятностей СВ Y .

Если возможные значения СВ принадлежат интервалу (a,b) , то теоретическое математическое ожидание находится по формуле:

$$M(Y) = \int_a^b yf(y)dy. \quad (2.70)$$

Математическое ожидание – это, пожалуй, основная и наиболее употребительная характеристика центра группирования СВ.

Оценку математического ожидания (выборочное среднее) получают как среднее арифметическое значение y_i :

$$\hat{m}_1 = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i . \quad (2.71)$$

Медиана y_{med} определяется как средневероятное значение СВ, т.е. вероятность того, что анализируемая СВ окажется больше медианы, равна вероятности того, что анализируемая СВ окажется меньше медианы и равна 0.5. Медиана также как и математическое ожидание относится к характеристикам центра группирования.

Выборочный аналог медианы $\hat{y}_{med}$ определяется как среднее (по местоположению) значение ранжированного (вариационного), т.е. расположенного в порядке возрастания ряда наблюдений $y_1 \leq y_2 \leq \dots \leq y_n$. Если количество наблюдений n нечетно, то

$$\hat{y}_{med} = y_{(n+1)/2} , \quad (2.72)$$

если n четно, то

$$\hat{y}_{med} = 1/2(y_{n/2} + y_{(n/2+1)}) . \quad (2.73)$$

Выборочная мода $\hat{y}_{mod}$ представляет собой наиболее часто встречающееся в выборочных наблюдениях значение, т.е. значение которое является «модным». Мода относится к характеристикам центра группирования.

Теоретическая дисперсия непрерывной СВ Y , возможные значения которой принадлежат интервалу (a,b) , вычисляется по формуле:

$$D(Y) = \int_a^b [y - M(Y)]^2 f(y) dy = \int_a^b y^2 f(y) dy - [M(Y)]^2 . \quad (2.74)$$

Оценку дисперсии можно найти по формуле:

$$\hat{D} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 . \quad (2.75)$$

Дисперсия относится к характеристикам степени рассеяния значений исследуемой СВ относительно центра группирования. Значение дисперсии дает представление о том, как сильно могут отклоняться от среднего значения выборочные наблюдения.

Выборочное среднеквадратическое отклонение (СКО):

$$\hat{\sigma} = \sqrt{\hat{D}}. \quad (2.76)$$

СКО используется наряду с выборочной дисперсией для характеристики степени отклонения наблюдаемых значений от среднего значения и оказывается в ряде случаев более удобным и естественным, так как СКО имеет ту же размерность, что и анализируемый показатель и соответственно выборочные среднее, медиана, мода.

Еще одна характеристика разброса (рассеяния) данных наблюдений – *выборочный размах*:

$$\hat{R} = y_{\max} - y_{\min}, \quad (2.77)$$

где $y_{\max}$ – максимальное выборочное значение;

$y_{\min}$ – минимальное выборочное значение.

Выборочный коэффициент вариации используется в тех случаях, когда степень рассеяния естественно описывать некоторой относительной характеристикой в соответствии со средним:

$$\hat{V} = \frac{\hat{\sigma}}{\bar{y}} 100\%. \quad (2.78)$$

Выборочный коэффициент вариации измеряется в процентах. По значению коэффициента вариации можно выдвинуть гипотезу о типе закона распределения СВ Y . Например, для экспоненциального распределения $\hat{V}=100\%$; для гамма распределения и распределения Вейбулла $\hat{V}$ может быть $<100\%$, $=100\%$, $>100\%$, когда параметр формы распределений соответственно <1 , $=1$, >1 .

Построение доверительных интервалов для математического ожидания и дисперсии.

Интервальной называют оценку, которая определяется двумя числами – концами интервала, покрывающего оцениваемый параметр.

Доверительным называют интервал, который с заданной вероятностью покрывает заданную характеристику.

Точность оценки математического ожидания характеристики определяется по формуле:

$$\varepsilon = M(y) - \hat{m}_1. \quad (2.79)$$

Доверительная вероятность оценки математического ожидания – вероятность того, что разница между теоретическим математическим ожиданием и выборочным средним не превзойдет заданной величины ε :

$$P(|M(y) - \hat{m}_1| < \varepsilon) = p. \quad (2.80)$$

Величина $\alpha = 1 - p$ называется уровнем значимости.

Пример. Пусть $\alpha = 0,05$ $p = 0,95$ $\varepsilon = 0,01$.

В 95 случаях из 100 разница между теоретическим и выборочным математическим ожиданием не превзойдет 0,01.

Доверительный интервал для математического ожидания задается формулой:

$$\hat{m}_1 - \varepsilon < M(y) < \hat{m}_1 + \varepsilon. \quad (2.81)$$

Точность оценки математического ожидания в случае известной дисперсии (известны теоретические характеристики) определяется формулой:

$$\varepsilon = \frac{x_{1+\alpha} \cdot \sigma}{\sqrt{n}}, \quad (2.82)$$

где $x_{1+\alpha}$ – квантиль стандартного нормального распределения уровня $1 + \alpha$.

Для $\alpha = 0,05$ $x_{1+\alpha} = 1,96$. Для $\alpha = 0,01$ $x_{1+\alpha} = 2,58$.

Точность оценки математического ожидания в случае неизвестной дисперсии (теоретические характеристики заменяются выборочными аналогами) определяется формулой:

$$\varepsilon = \frac{t_{(1-\alpha/2)} \cdot \hat{\sigma}}{\sqrt{n}}, \quad (2.83)$$

где $t_{(1-\alpha/2)}(n-1)$ – квантиль распределения Стьюдента с числом степеней свободы $(n-1)$ уровня $(1-\alpha/2)$.

Например, при $n=100$ и $\alpha=0,05$ квантиль распределения Стьюдента $t_{(0,975)}(99)=1,98$.

Аналогичные формулы для построения доверительного интервала для дисперсии имеют вид:

$$\hat{D} - \varepsilon < D(y) < \hat{D} + \varepsilon. \quad (2.84)$$

$$\varepsilon = \frac{x_{1+\alpha} \cdot \sqrt{\mu_4 - \sigma^4}}{\sqrt{n}}, \quad (2.85)$$

где μ_4 – центральный момент четвертого порядка.

Теоретический центральный момент четвертого порядка непрерывной СВ Y , возможные значения которой принадлежат интервалу (a,b) , вычисляется по формуле:

$$\mu_4 = \int_a^b [y - M(Y)]^4 f(y) dy. \quad (2.86)$$

Оценку центрального момента четвертого порядка можно найти по формуле:

$$\hat{\mu}_4 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^4. \quad (2.87)$$

Определение объема выборки. Объем выборки – это количество реализаций (объем экспериментов), которое необходимо при имитационном моделировании, чтобы обеспечить требуемую статистическую точность результата. При определении количества реализаций обычно учитывается вид характеристики эффективности. Характеристикой эффективности может быть вероятность выполнения той или иной задачи или некоторая скалярная функция параметров, алгоритма, струк-

туры системы. Например, среднее время пребывания заявки в системе, среднее время ожидания в очереди, дисперсия времени пребывания заявки в системе, дисперсия времени ожидания в очереди и т.п.

При вычислении оценки математического ожидания выбранной характеристики (параметра) с заданной точностью и достоверностью объем экспериментов можно вычислить по формуле, выведенной на основе (2.82):

$$n = \frac{x_{1+\alpha}^2 \sigma^2}{\varepsilon^2} \quad (2.88)$$

Также объем экспериментов можно вычислить на основе дисперсии выбранной характеристики (параметра):

$$n = \frac{x_{1+\alpha}^2 (\mu_4 - \sigma^4)}{\varepsilon^2}. \quad (2.89)$$

Для практического использования формул (2.88, 2.89) необходимо знать теоретические D и μ_4 , которые, как правило, бывают неизвестными. В этом случае используют оценки $\hat{D}$ и $\hat{\mu}_4$, получаемые предварительно, при некотором n (50-100 реализаций). Если при этом заданная точность ε не достигнута, то n увеличивают, после чего проводится уточнение оценок. Этот процесс повторяется до тех пор, пока не будет обеспечена заданная точность.

Проверка гипотезы о типе закона распределения СВ [3, 13, 17]. При обработке результатов машинного эксперимента с моделью системы часто возникает задача определения эмпирического закона распределения СВ. Для этого строят гистограмму (эмпирическую оценку функции плотности распределения вероятностей СВ, и на основе ее анализа выдвигают гипотезу о типе закона распределения СВ. Проверка гипотезы реализуется с помощью критериев согласия (Колмогорова – Смирнова, χ^2 Пирсона, Крамера – Мизеса – Смирнова и других).

Рассмотрим один из часто используемых критериев. Критерий согласия χ^2 Пирсона проверяет основную гипотезу о согласии выборочного распределения с теоретическим распределением:

$$H_0 : \hat{F}(x; \hat{\Theta}) = F_{теор}(x; \Theta);$$

$$H_1 : \hat{F}(x; \hat{\Theta}) \neq F_{теор}(x; \Theta).$$

При этом параметры теоретического закона могут как известны, т.е. проверяется простая гипотеза, так и не известны (оцениваются по выборке), т.е. проверяется сложная гипотеза.

Для использования критерия выборочные данные представляются в группированном виде, строится гистограмма распределения. Объем выборочных данных должен быть достаточно большим ($n \geq 200$).

Критическая статистика критерия имеет вид:

$$\psi_{кр} = \chi^2(k-s-1) = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} \quad (2.90)$$

где k – число интервалов группирования;

s – число неизвестных параметров распределения;

n_i – число наблюдений в интервале,

p_i – теоретическая вероятность попадания в интервал.

Для вычисления n_i необходимо: разбить диапазон изменения СВ Y на k интервалов так, чтобы соблюдались следующие условия:

- число интервалов должно быть не меньше 8 ($k \geq 8$);
- в каждый интервал должно попасть не менее 7-10 значений.

Если число наблюдений в интервале меньше 5, то происходит объединение интервалов.

Теоретическая вероятность попадания в интервал вычисляется по формуле:

$$p_i = \int_{y_i}^{y_{i+1}} f(y) dy = F(y_{i+1}) - F(y_i), \quad (2.91)$$

где y_i и y_{i+1} – границы интервалов.

Критическая статистика критерия подчиняется распределению χ^2 с $(k-s-1)$ числом степеней свободы в случае проверки сложной гипотезы и $(k-1)$ числом степеней свободы в случае проверки простой гипотезы.

Вычисляется расчетное значение критерия $\chi^2_{расч}$ на основе формулы (2.90) и сравнивается с двумя критическими границами:

$$\chi^2_{\left(1-\frac{\alpha}{2}\right) \cdot 100\%}(k-s-1) \leq \chi^2_{расч} \leq \chi^2_{\frac{\alpha}{2} \cdot 100\%}(k-s-1). \quad (2.92)$$

Критические границы – процентные точки или квантили распределения χ^2 , которые находятся по статистической таблице. Если неравенство (2.92) выполняется, то гипотеза о согласии выборочного распределения с моделью теоретического закона распределения не отвергается при уровне значимости α .

2.9. Сравнение альтернативных конфигураций системы

Задача имитационного моделирования сводится не только к оценке эффективности функционирования системы и ее оптимизации, но и к сравнению альтернативных конфигураций системы, выбору лучшего варианта реализации системы. Сравниваться могут конкурирующие проекты реализации системы, или варианты имитационной модели с образцом, который представляет существующую систему, или выходные характеристики реальной системы и имитационной с целью доказательства адекватности модели. Рассмотрим пример.

Пример 1. В компании рассматривается вопрос о приобретении нового оборудования, которое позволит сократить время обслуживания заявок на 10%. Будет ли сокращение времени статистически значимым? Известно, что заявки поступают в систему по экспоненциальному закону со средним равным 10 единиц. Обслуживаются заявки также по экспоненциальному закону со средним равным 9 единиц (в исходной системе). Необходимо сравнить характеристики эффективности функционирования двух вариантов систем (среднее время пребывания заявки в системе).

Пусть по результатам n_i прогонов имитационной модели i -ого варианта реализации системы ($i=1,2$) получены выборки выходных характеристик: $\bar{y}_{ij}$, $j=1, n_i$. Например, выборки средних времен пребывания заявки в системе

для двух вариантов системы (пример). Полученные значения приведены в таблице 2.3. Было выполнено 5 прогонов $n_1 = n_2 = 5$ с имитационными моделями системы, каждый прогон состоял из 100 испытаний (экспериментов)).

Таблица 2.3 – Результаты моделирования

Номер прогона, j	Первый вариант реализации $\bar{y}_{1j}$	Второй вариант реализации $\bar{y}_{2j}$
1	49,6	26,3
2	55,6	42,2
3	48,5	31,3
4	60,8	36,7
5	57,7	33,8

Один из вариантов сравнения – построение доверительного интервала для разности выходной характеристики для двух вариантов системы на основе парного t – критерия Стьюдента. Данный метод используется, если число прогонов для первого и второго вариантов системы одинаково: $n = n_1 = n_2$.

Границы доверительного интервала для разности средних значений выходных характеристик для двух вариантов реализации системы вычисляются по формуле:

$$\bar{Z}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\hat{D}[\bar{Z}(n)]}, \quad (2.93)$$

где

$$\hat{D}[\bar{Z}(n)] = \frac{\sum_{j=1}^n [Z_j - \bar{Z}(n)]^2}{(n-1)},$$

$$\bar{Z}(n) = \frac{\sum_{j=1}^n Z_j}{n},$$

$$Z_j = \bar{y}_{1j} - \bar{y}_{2j}, \text{ при } j = 1, 2, \dots, n,$$

$t_{n-1, 1-\alpha/2}$ – квантиль распределения Стьюдента с $(n-1)$ степенями свободы уровня $1 - \alpha/2$.

В результате вычислений получены следующие значения:

$$\alpha = 0,05; \bar{Z}(n) = 20,38; \hat{D}[\bar{Z}(n)] = 23,397;$$

$$t_{n-1, 1-\alpha/2} = 2,77;$$

$$20,38 \pm 2,77 \cdot 4,83.$$

Таким образом, 95% доверительный интервал Стьюдента: $[7; 33,8]$.

Границы доверительного интервала имеют один знак, поэтому делается вывод о статистической значимости отличий выходных характеристик для двух вариантов системы при заданном уровне значимости 0,05. Приобретение нового оборудования приведет к уменьшению времени обработки заявки в системе на величину от 7 до 33,8 единиц.

В случае, если доверительный интервал Стьюдента содержит ноль, делается вывод о статистической незначимости отличий выходной характеристики в двух вариантах реализации системы.

Для применения t – критерия Стьюдента необходимо, чтобы дисперсии выборок значений выходных характеристик для первого и второго вариантов системы, полученные по n прогонам, были равны: $\hat{D}_1 = \hat{D}_2$. При моделировании реальных систем – это ненадежное допущение, поэтому на практике, как правило используют доверительный интервал Велча. Рассмотрим применение доверительного интервала Велча на примере.

Пример 2. Моделируется работа производственного участка, изготавливающего детали определенного вида. Процесс изготовления детали требует последовательного выполнения двух операций соответственно на токарном и фрезерном станках. На участке имеется 2 токарных станка и 1 фрезерный. Известно, что поток заказов на изготовление деталей пуассоновский с интенсивностью $\lambda=0.05$ (деталей в минуту). Допустимая очередь заказов на изготовление деталей перед токарным станком равна 5. Если очередь равна 5-ти, то очередной заказ передается на другой производственный участок. Допустимая очередь заказов перед фре-

зерным станком равна 2. Если в очереди 2 заказа, то очередной заказ не принимается токарными станками, пока не появится место в очереди. Время обработки на токарных станках в среднем: $m_{c1}=15$ минут (1-ый станок); $m_{c2}=20$ минут (2-ой станок), на фрезерном станке в среднем: $m_{c3}=5$ минут (закон распределения экспоненциальный).

В производственной компании оценивается целесообразность покупки нового фрезерного станка, позволяющего уменьшить время обработки деталей на 10%. Изучение системы посредством имитационного моделирования поможет определить, позволит ли это значительно уменьшить среднее время пребывания детали на производственном участке. Задача заключается в сравнении средних значений выходных характеристик для двух вариантов реализации системы с помощью статистических процедур.

В случае сравнения выборок значений выходных характеристик (для двух вариантов системы) с неравными и неизвестными дисперсиями строят доверительный интервал Велча, основанный на t -критерия Стьюдента [3]. Критерий допускает неравное число прогонов двух вариантов реализации системы: $n_1 \neq n_2$.

Границы доверительного интервала Велча для разности средних значений выходных характеристик для двух вариантов реализации системы вычисляются по формуле:

$$\bar{y}_1 - \bar{y}_2 \pm t_{\hat{f}, 1-\alpha/2} \sqrt{\frac{\hat{D}_1}{n_1} + \frac{\hat{D}_2}{n_2}}, \quad (2.94)$$

где $\bar{y}_1, \bar{y}_2$ – среднее выборок значений выходных характеристик для первого и второго вариантов системы, полученные по n_1, n_2 прогонам;

$\hat{D}_1, \hat{D}_2$ – дисперсия выборок значений выходных характеристик для первого и второго вариантов системы, полученные по n_1, n_2 прогонам;

n_1, n_2 – количество прогонов первого и второго вариантов реализации системы;

$t_{\hat{f}, 1-\alpha/2}$ – квантиль распределения Стьюдента с $\hat{f}$ степенями свободы уровня $1 - \alpha/2$.

Оценка числа степеней свободы $\hat{f}$ вычисляется по формуле:

$$\hat{f} = \frac{[\hat{D}_1/n_1 + \hat{D}_2/n_2]^2}{[\hat{D}_1/n_1]^2/(n_1-1) + [\hat{D}_2/n_2]^2/(n_2-1)}. \quad (2.95)$$

В результате вычислений получены следующие значения: $n_1 = n_2 = 5$ (выполнено 5 прогонов с имитационными моделями системы, каждый прогон состоял из 100 испытаний (экспериментов)); $\bar{y}_1 = 28.9$ минут и $\bar{y}_2 = 27.8$ минут (среднее время пребывания детали на производственном участке для первого и второго вариантов системы); $\hat{D}_1 = 2.16$ и $\hat{D}_2 = 1.7$ (дисперсия времени пребывания детали на производственном участке для первого и второго вариантов системы); $\hat{f} = 7.89$; $t_{7.89, 0.95} = 1.86$. Следовательно, 90-процентный доверительный интервал Велча составляет: $[-2.69; 0.58]$.

Так как доверительный интервал содержит нулевое значение, делается вывод о статистической незначимости отличий между средним временем пребывания детали на производственном участке для двух вариантов системы при уровне значимости 0.1. Таким образом, приобретение нового фрезерного станка не приведет к значимому уменьшению времени обработки детали и поэтому нецелесообразно.

На заключительном этапе подводятся итоги исследования. Выбирается и обосновывается наилучший вариант функционирования системы. Даются рекомендации по улучшению характеристик работы системы.

2.10. Контрольные вопросы и упражнения

1. В чем сущность имитационного моделирования? Какие основные проблемы связаны с имитационным моделированием системы?
2. Дана имитационная модель простейшего цифрового фильтра:

$$y_{i+1} = y_i + \frac{1}{2} \Delta t (x_{i+1} - x_i),$$

где x_i и y_i , $i=0,1,2,\dots$ соответственно входной и выходной процессы, $\Delta t = 1$. Какой принцип должен быть использован для построения моделирующего алгоритма? Почему? Перечислите достоинства и недостатки этого принципа.

3. Каковы основные этапы разработки имитационной модели системы?

4. Какие способы генерации последовательностей СЧ используются при моделировании систем? Что собой представляют конгруэнтные процедуры генерации последовательностей СЧ?

5. В чем сущность метода статистического моделирования систем на компьютере? Что является теоретической основой метода?

6. В результате имитационного моделирования вычислительной системы получены выборочные данные времени пребывания заданий в системе (в сек.): 17.2; 22.1; 22.5; 23.7; 21.4; 24.2; 24.3; 23.7; 24.9; 24.5; 23.2; 25.1; 25.4; 25.2; 24.9; 25.8; 26.2. Рассчитайте статистические характеристики показателя «время пребывания заявки в системе», выполните их содержательную интерпретацию.

7. Оцените объем выборки, необходимый для оценки математического ожидания и дисперсии времени пребывания заявки в системе (см. задачу п. 6) с точностью 0.05 и достоверностью 0.99.

3. Отдельные разделы теории моделирования систем

3.1. Элементы теории Марковских случайных процессов, применяемые при моделировании систем

3.1.1. Потоки событий

Одно из основных понятий теории моделирования систем – понятие потока событий, поступающего на вход системы.

Потоком событий называется последовательность однородных событий, появляющихся одно за другим в случайные моменты времени [27]. Например, поток заданий на обработку в вычислительной системе; поток вызовов на телефонной станции; поток автомашин, подъезжающих на заправочную станцию; поток заболеваний гриппом в зимний сезон; поток забитых шайб при игре в хоккей; по-

ток заявок на ремонт, поступающих в ремонтную организацию; поток отказов (сбоев) компьютера в ходе его работы и т.д.

Математическое представление потока событий. Поток событий представляет собой в общем случае просто последовательность случайных точек $\theta_1, \theta_2, \dots, \theta_n, \dots$ на оси времени $0t$ с разделяющими их случайными интервалами $\tau_1, \tau_2, \dots, \tau_{n-1}, \tau_n, \dots$, так что $\tau_1 = \theta_2 - \theta_1, \tau_2 = \theta_3 - \theta_2, \dots, \tau_n = \theta_{n+1} - \theta_n, \dots$ (рисунок 3.1).



Рисунок 3.1.

Потоки событий различаются между собой по внутренней структуре: по законам распределения интервалов $\tau_1, \tau_2, \dots, \tau_{n-1}, \tau_n, \dots$ между событиями (например, нормальный, равномерный, экспоненциальный законы и другие), их взаимной зависимости или независимости и т.д.

С первого взгляда наиболее простым представляется поток событий, в котором интервалы между событиями строго одинаковы и равны определенной неслучайной величине τ . Такой поток событий называется регулярным. Примеры регулярных потоков представляют собой поток изменений минутной цифры на вокзальных электронных часах, поток изменений состояний компьютера, определяемый тактом его работы и т.п.

Регулярный поток событий редко встречается на практике и представляет определенный интерес только как предельный случай для других потоков. Несмотря на видимую простоту, регулярный поток не имеет преимуществ при математическом анализе, так как намного уступает по простоте проведения расчетов другим типам потоков.

Математическая модель простейшего Пуассоновского потока. На практике чаще всего ограничиваются рассмотрением простейшего (Пуассоновского) потока заявок.

Поток событий, обладающий свойствами ординарности, стационарности и отсутствия последействия, называется *простейшим* (или *стационарным Пуассоновским*) *поток*ом. Простейшим этот поток назван потому, что исследование систем, находящихся под воздействием простейших потоков, проводится самым простым образом.

Для простейшего потока событий вероятность того, что на участке времени длины τ наступит ровно k событий, имеет распределение Пуассона с параметром $\alpha = \lambda\tau$:

$$P\{X(t, \tau = k)\} = \alpha^k e^{-\alpha} / k!, \quad (k=0, 1, 2, \dots), \quad (3.1)$$

где λ – интенсивность потока событий.

Физический смысл λ – это среднее число событий, приходящееся на единицу времени (число заявок в единицу времени); размерность – 1/время.

Распределение интервалов между заявками для простейшего потока будет экспоненциальным (показательным) с функцией распределения вероятностей и функцией плотности распределения вероятностей соответственно:

$$F(\tau) = P(t \leq \tau) = 1 - e^{-\lambda\tau}, \quad (3.2)$$

$$f(\tau) = \lambda e^{-\lambda\tau}.$$

Математическое ожидание и дисперсия длины интервала времени между последовательными моментами поступления событий соответственно:

$$M[\tau] = \int_0^{\infty} \tau f(\tau) d\tau = 1/\lambda, \quad (3.3)$$

$$D[\tau] = \int_0^{\infty} \tau^2 f(\tau) d\tau - (M[\tau])^2 = 1/\lambda^2. \quad (3.4)$$

Свойства простейшего пуассоновского потока [27].

1. Ординарность. Поток называется ординарным, если события в нем происходят по одному, а не группами по 2, 3 и т.д. Ординарность потока означает, что вероятность попадания на элементарный участок Δt двух или более событий пренебрежимо мала по сравнению с вероятностью попадания на него ровно одно-

го события, т.е. при $\Delta t \rightarrow 0$ эта вероятность представляет собой бесконечно малую высшего порядка:

$$\lim_{\Delta t \rightarrow 0} \frac{p_{>1}(t, \Delta t)}{p_1(t, \Delta t)} = 0. \quad (3.5)$$

Для ординарного потока можно пренебречь возможностью совместного появления на элементарном участке двух и более событий. В каждый момент времени в систему может поступать не более одной заявки.

Примерами ординарных потоков событий могут служить поток деталей, поступающих на конвейер для сборки; поток отказов технического устройства и т.д. Пример неординарного потока – поток пассажиров, прибывающих в лифте на данный этаж. Если в неординарном потоке события происходят только парами, тройками и т.д., то рассматривают ординарный поток пар, троек и т.д.

2. Отсутствие последействия. Для любых не перекрывающихся участков времени $\tau_1, \tau_2, \dots, \tau_n$, (рисунок 3.2) числа событий $X_1=X(t_1, \tau_1), X_2=X(t_2, \tau_2), \dots, X_n=X(t_n, \tau_n)$, попадающих на эти участки, представляют собой независимые случайные величины, т.е. вероятность попадания любого числа событий на один из участков не зависит от того, сколько их попало на другие.

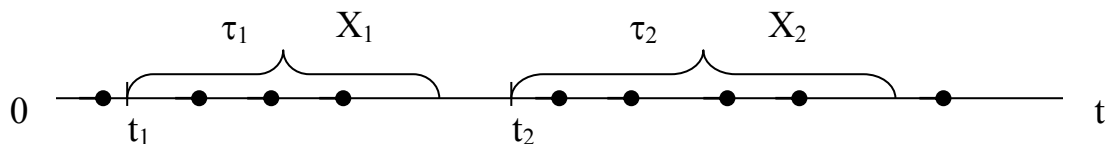


Рисунок 3.2

Отсутствие последействия означает, что для любого момента времени t_0 , будущие моменты наступления события потока (при $t > t_0$) не зависят от того, в какие моменты наступали события в прошлом (при $t < t_0$).

Ординарный поток событий, в котором отсутствует последействие, называется пуассоновским потоком.

3. Стационарность. Поток событий называется стационарным, если все его вероятностные характеристики не меняются со временем. В частности, для

стационарного потока событий вероятность попадания того или иного числа событий на участок длины τ зависит только от длины этого участка и не зависит от того, где именно на оси времени $0t$ этот участок расположен. Это значит, что числа событий $X_1=X(t_1, \tau_1)$ и $X_2=X(t_2, \tau_2)$, попадающих на два участка одинаковой длины, будут иметь одинаковые распределения. Отсюда следует, в частности, что для стационарного потока событий его интенсивность $\lambda(t)$ постоянна.

Кроме того, к достоинствам простейшего потока можно также отнести:

1. Сумма N независимых, ординарных и стационарных потоков заявок с интенсивностями $\lambda_i (i = \overline{1, N})$ сходится к простейшему потоку с интенсивностью

$$\lambda = \sum_{i=1}^N \lambda_i, \text{ при условии, что складываемые потоки оказывают более или менее}$$

одинаково малое влияние на суммарный поток.

2. Поток заявок, полученный путем случайного разрежения исходного потока, когда каждая заявка с определенной вероятностью p исключается из потока независимо от того, исключены другие заявки или нет, образует простейший поток с интенсивностью $\lambda_i = p\lambda$, где λ – интенсивность исходного потока. В отношении исходного потока заявок делается предположение лишь об ординарности и стационарности.

Поток с ограниченным последствием (рекуррентный поток) – поток, у которого случайные интервалы $\tau_1, \tau_2, \dots, \tau_n$ между соседними по времени событиями представляют собой независимые случайные величины. При его моделировании применяется последовательная (рекуррентная процедура): сначала разыгрывается величина τ_1 , затем τ_2 и т.д. Поток без последствия – частный случай потока с ограниченным последствием.

Поток Пальма. Стационарный поток с ограниченным последствием называется потоком Пальма. Для такого потока интервалы $\tau_1, \tau_2, \dots, \tau_n$ между соседними событиями представляют собой последовательность независимых, одинаково распределенных СВ.

Простейший пуассоновский поток – поток Пальма, так как у простейшего потока интервалы $\tau_1, \tau_2, \dots, \tau_n$ распределены одинаково, по показательному закону и независимы между собой. Поток Пальма отличный от простейшего получается, если интервалы между событиями распределены по другому закону.

Потоки Эрланга, их свойства и применение. Поток Эрланга k -го порядка называется поток Пальма, у которого интервалы времени между событиями распределены по закону Эрланга k -го порядка. Поток Эрланга k -го порядка может быть получен из простейшего с помощью его прореживания. В простейшем потоке сохраняется каждое k -е событие, остальные отбрасываются. На рисунке 3.3 приведен пример формирования потока Эрланга 2-го порядка.

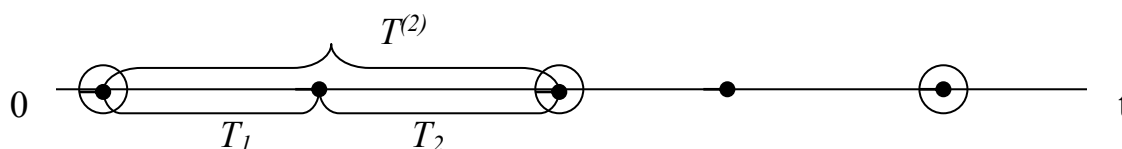


Рисунок 3.3.

Интервал между соседними событиями в потоке Эрланга k -го порядка:

$$T^{(k)} = \sum_{i=1}^k T_i. \quad (3.6)$$

Функция плотности распределения интервала времени между двумя событиями описывается законом Эрланга k -го порядка:

$$f^{(k)}(\tau) = \lambda(\lambda\tau)^{k-1} e^{-\lambda\tau} / (k-1)!. \quad (3.7)$$

Математическое ожидание и дисперсия длины интервала времени между последовательными моментами поступления событий соответственно:

$$M[T^{(k)}] = k / \lambda, \quad (3.8)$$

$$D[T^{(k)}] = k / \lambda^2.$$

Свойства потока Эрланга. При $k=1$ получается обычное экспоненциальное распределение, а при $k \rightarrow \infty$ поток Эрланга приближается к регулярному потоку с постоянным интервалом между событиями. Это свойство потоков Эрланга удобно в практических применениях, так как дает возможность, задаваясь различными k , получать потоки, обладающие различным последствием – от полно-

го отсутствия последствия ($k=1$), до жесткой функциональной связи между моментами появления событий ($k=\infty$).

3.1.2. Случайный процесс

Второе фундаментальное понятие теории моделирования систем – это понятие случайного процесса.

Пусть имеется некоторая система S , которая в процессе функционирования может принимать различные состояния S_i , $i=1, 2, \dots, n$. Если состояния системы меняются случайным образом, то последовательность состояний системы образует случайный процесс. Полное множество состояний исследуемой системы S_i может быть либо конечным, либо бесконечно большим. Большинство реальных систем имеют дискретное конечное пространство состояний. Последовательность состояний такой системы S_i и сам процесс переходов из состояния в состояние называется *цепью* [6].

Различают системы с непрерывным и дискретным временем.

Системы с непрерывным временем предполагают, что переход системы из одного состояния в другое может осуществляться в любой момент времени, т.е. время пребывания системы в каждом состоянии представляет непрерывную случайную величину. Для *систем с дискретным временем* время пребывания системы в каждом состоянии фиксированное, а моменты переходов $t_1, t_2, \dots, t_k$ размещаются на временной оси через равные промежутки и называются «шагами» или «этапами». Время нахождения системы в некотором состоянии представляет дискретную случайную величину.

На практике случайные процессы с дискретными состояниями и дискретным временем встречаются довольно редко. Как правило, рассматривают случайные процессы с дискретными состояниями и непрерывным временем.

3.1.3. Непрерывные цепи Маркова. Расчет вероятностей состояний

Пусть имеется некоторая система S , переходы системы из состояния в состояние происходят в случайные моменты времени под воздействием каких-либо потоков событий (например, поток отказов, поток восстановлений). Для системы такими событиями могут быть: поступление очередной заявки (событие входного потока); окончание обслуживания очередной заявке (событие потока обслуживания заявок) и т.д.

Процесс, протекающий в системе S , называется *Марковским процессом с дискретными состояниями и непрерывным временем* (или *непрерывной Марковской цепью*), если выполняется условие: для любого фиксированного момента времени условные вероятности состояния системы в будущем зависят только от состояния системы в настоящем и не зависят от того, когда (на каком шаге) и откуда система перешла в это состояние [16, 27].

Схематично систему удобно представлять в виде размеченного графа состояний. *Граф состояний* – это схема, отражающая переход системы из состояния в состояние. Вершины графа соответствуют состояниям, дуги – переходам из состояния в состояние. *Размеченный граф состояний* – граф состояний с проставленными у стрелок интенсивностями соответствующих потоков событий, переводящих систему из состояния в состояние.

Пример 1. Рассматриваемая система – компьютер, обрабатывающий поток заданий. В очереди на обработку может стоять не более двух заданий. Поток заданий Пуассоновский с интенсивностью λ , поток обслуживания Пуассоновский с интенсивностью μ . Размеченный граф состояний системы имеет вид:

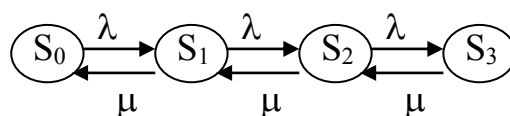


Рисунок 3.4.

S_0 – в системе нет заявок, компьютер простаивает;

S_1 – в системе одна заявка, которая обрабатывается компьютером;

S_2 – одна заявка обрабатывается, одна стоит в очереди на обработку;

S_3 – одна заявка обрабатывается, две стоят в очереди на обработку.

Здесь и в дальнейшем предполагается ординарность суммарного потока всех событий в системе, то есть одновременно не могут произойти два и более событий (невозможно одновременное окончание обработки заявки и поступление в систему заявки). Другое предположение – нулевое время перехода заявки из очереди на канал обслуживания и нулевое время освобождения канала.

Основная задача анализа системы: вычислить вероятности состояний в любой момент времени t , т.е. $p_0(t), p_1(t), \dots, p_n(t)$, где n – число состояний системы. Для нахождения вероятностей состояний необходимо знать характеристики переходного процесса. Введем в рассмотрение плотность вероятности перехода λ_{ij} :

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{p_{ij}(t)}{\Delta t}, \quad (3.9)$$

где Δt – элементарный участок времени, примыкающий к t ;

$p_{ij}(\Delta t)$ – вероятность того, что система, находящаяся в S_i , перейдет в S_j за Δt .

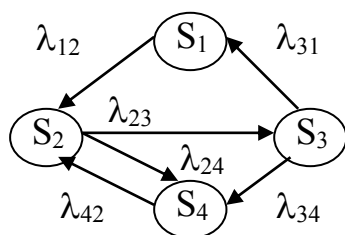
При малом Δt :

$$p_{ij}(\Delta t) \approx \lambda_{ij} \Delta t. \quad (3.10)$$

Если все плотности вероятностей λ_{ij} не зависят от времени, Марковский процесс называют *однородным*.

Для нахождения вероятностей $p_0(t), p_1(t), \dots, p_n(t)$ составляют систему дифференциальных уравнений Колмогорова. Покажем методику вывода уравнений Колмогорова на примере.

Пример 2. На рисунке 3.5 приведен размеченный граф состояний системы. Необходимо найти вероятность первого состояния системы в момент t : $p_1(t)$.



Решение. Придадим t малое приращение Δt и найдем вероятность того, что в момент $t + \Delta t$ система будет находиться в S_1 .

Это событие может произойти двумя способами:

Рисунок 3.5.

1 способ. В момент времени t система была в S_1 и за Δt не вышла из S_1 .

Вероятность этого события: $p_1(t)(1 - \lambda_{12}\Delta t)$.

2 способ. В момент времени t система была в S_3 и за Δt перешла в S_1 .

Вероятность этого события: $p_3(t)\lambda_{31}\Delta t$.

По правилу сложения вероятностей:

$$p_1(t + \Delta t) = p_1(t)(1 - \lambda_{12}\Delta t) + p_3(t)\lambda_{31}\Delta t. \quad (3.11)$$

Перенесем $p_1(t)$ в левую часть и разделим на Δt :

$$\frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} = -p_1(t)\lambda_{12} + p_3(t)\lambda_{31}. \quad (3.12)$$

Теперь устремим Δt к нулю:

$$\lim_{\Delta t \rightarrow 0} \frac{p_1(t + \Delta t) - p_1(t)}{\Delta t} = -p_1(t)\lambda_{12} + p_3(t)\lambda_{31}. \quad (3.13)$$

В результате в левой части производная функции $p_1(t)$ по времени:

$$\frac{dp_1(t)}{dt} = -p_1(t)\lambda_{12} + p_3(t)\lambda_{31}. \quad (3.14)$$

Данному дифференциальному уравнению удовлетворяет функция $p_1(t)$.

Аналогично можно вывести дифференциальные уравнения для вероятностей состояний $p_2(t), p_3(t), p_4(t)$ и составить систему дифференциальных уравнений.

Одно из уравнений в системе заменяется на нормировочное условие: для любого момента t сумма вероятностей состояний равна единице:

$$\sum_{i=1}^4 p_i(t) = 1, \quad (3.15)$$

так как события, состоящие в том, что в момент t система S находится в состояниях S_1, S_2, S_3, S_4 несовместны и образуют полную группу.

Уравнения Колмогорова имеют определенную структуру и построены по правилу, справедливому для любой непрерывной Марковской цепи.

В левой части уравнения стоит производная вероятности состояния, а правая часть содержит столько членов, сколько стрелок связано с данным состоянием. Если стрелка направлена из состояния, то соответствующий член имеет знак «минус»; если стрелка направлена в состояние – знак «плюс». Каж-

дый член равен произведению интенсивности потока событий, соответствующего данной стрелке, умноженной на вероятность того состояния, из которого исходит стрелка.

Для *примера 1*, система уравнений Колмогорова будет иметь вид:

$$\begin{aligned}\frac{dp_0}{dt} &= -\lambda p_0(t) + \mu p_1(t), \\ \frac{dp_1}{dt} &= -(\lambda + \mu)p_1(t) + \mu p_2(t) + \lambda p_0(t), \\ \frac{dp_2}{dt} &= -(\lambda + \mu)p_2(t) + \mu p_3(t) + \lambda p_1(t), \\ p_0(t) + p_1(t) + p_2(t) + p_3(t) &= 1.\end{aligned}\tag{3.16}$$

Если все потоки событий, переводящие систему из состояния в состояние пуассоновские и независимые, то процесс, протекающий в системе – Марковский. Действительно, Пуассоновский поток обладает отсутствием последействия при заданном состоянии системы в данный момент. Переходы системы в другие состояния обусловлены только появлением каких-либо событий в Пуассоновских потоках, а вероятности появления событий не зависят от «предыстории» процесса.

Стационарный режим для непрерывной цепи Маркова. Предельные значения $p_0(t), p_1(t), \dots, p_n(t)$ при $t \rightarrow \infty$:

$$\lim_{t \rightarrow \infty} p_i(t) = p_i, i = \overline{1, n}\tag{3.17}$$

называются *предельными вероятностями состояний*. Режим функционирования системы называют *стационарным*, если вероятности состояний не зависят от времени, и система случайным образом меняет состояния [16, 27].

Условия существования стационарного режима:

- цепь Маркова должна быть однородной;
- множество состояний системы должно быть эргодическим, т.е. из любого состояния S_i можно за конечное число шагов перейти в состояние S_j .

Предельные вероятности состояний представляют собой среднее время пребывания системы в данном состоянии. Например, если у системы S три возможных состояния: S_1, S_2, S_3 , причем их предельные вероятности равны 0.2, 0.3, 0.5, то это означает, что после перехода к установившемуся режиму система S в среднем две десятых времени будет находиться в состоянии S_1 , три десятых – в состоянии S_2 , половину времени – в S_3 .

Для вычисления предельных вероятностей необходимо в системе уравнений Колмогорова положить все левые части (производные) равными нулю. Действительно, в предельном (стационарном) режиме все вероятности состояний постоянны, а значит, их производные равны нулю. Следовательно, система дифференциальных уравнений превращается в систему линейных алгебраических уравнений.

Для *примера 1*, система уравнений Колмогорова будет иметь вид:

$$\begin{aligned} 0 &= -\lambda p_0 + \mu p_1, \\ 0 &= -(\lambda + \mu)p_1 + \mu p_2 + \lambda p_0, \\ 0 &= -(\lambda + \mu)p_2 + \mu p_3 + \lambda p_1, \\ p_0 + p_1 + p_2 + p_3 &= 1. \end{aligned} \tag{3.18}$$

3.1.4. Марковский случайный процесс «гибели и размножения».

Расчет вероятностей состояний

В *примере 1* состояния системы образуют цепь: каждое состояние, кроме исходного и последнего, связано прямой и обратной связью с двумя соседними состояниями; крайние состояния связаны с одним соседним. Такая схема процесса, протекающего в системе, называется схемой «гибели и размножения», а сам процесс – *процессом «гибели и размножения»* [16].

Если в такой системе все потоки, переводящие систему из состояния в состояние Пуассоновские, то процесс называется *Марковским случайным процессом «гибели и размножения»*. Термин ведет начало от биологических задач, процесс

описывает изменение численности популяции. Переход из состояния в состояние происходит в момент гибели или рождения особи.

На практике значительная часть систем может описываться в рамках схемы «гибели и размножения». Рассмотрим эту схему в общем виде и решим соответствующую систему алгебраических уравнений для нахождения предельных вероятностей состояний системы в стационарном режиме.

На рисунке 3.6 представлен граф состояний системы, описываемой процессом «гибели и размножения».

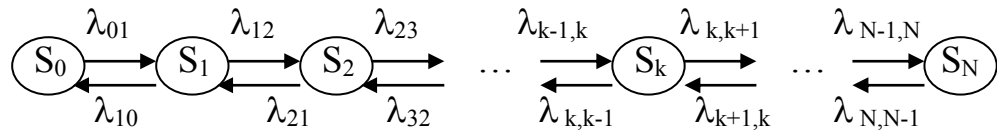


Рисунок 3.6.

Для нулевого состояния алгебраическое уравнение Колмогорова имеет вид:

$$\lambda_{01}p_0 = \lambda_{10}p_1. \quad (3.19)$$

Для первого состояния алгебраическое уравнение Колмогорова имеет вид:

$$\lambda_{12}p_1 + \lambda_{10}p_1 = \lambda_{01}p_0 + \lambda_{21}p_2. \quad (3.20)$$

В силу (3.19) можно сократить равные друг другу члены и получим:

$$\lambda_{12}p_1 = \lambda_{21}p_2. \quad (3.21)$$

Аналогичные уравнения получаются для всех состояний системы. В результате система уравнений Колмогорова имеет вид:

$$\begin{aligned} \lambda_{01}p_0 &= \lambda_{10}p_1, \\ \lambda_{12}p_1 &= \lambda_{21}p_2, \\ &\dots \\ \lambda_{N-1,N}p_{N-1} &= \lambda_{N,N-1}p_N, \\ p_1 + p_2 + \dots + p_N &= 1. \end{aligned} \quad (3.22)$$

Для решения системы (3.22) из первого уравнения выразим p_1 :

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0. \quad (3.23)$$

Из второго уравнения выразим p_2 :

$$p_2 = \frac{\lambda_{12}}{\lambda_{21}} p_1 = \frac{\lambda_{12}}{\lambda_{21}} \frac{\lambda_{01}}{\lambda_{10}} p_0. \quad (3.24)$$

Аналогично вероятности всех состояний выражаются через p_0 . Подставим полученные выражения в нормировочное условие:

$$p_0 + \frac{\lambda_{01}}{\lambda_{10}} p_0 + \frac{\lambda_{12}}{\lambda_{21}} \frac{\lambda_{01}}{\lambda_{10}} p_0 + \dots + \frac{\lambda_{N-1,N} \lambda_{N-2,N-1} \dots \lambda_{01}}{\lambda_{N,N-1} \lambda_{N-1,N-2} \dots \lambda_{10}} p_0 = 1, \quad (3.25)$$

откуда

$$p_0 = \left[1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}}{\lambda_{21}} \frac{\lambda_{01}}{\lambda_{10}} + \dots + \frac{\lambda_{N-1,N} \lambda_{N-2,N-1} \dots \lambda_{01}}{\lambda_{N,N-1} \lambda_{N-1,N-2} \dots \lambda_{10}} \right]^{-1}. \quad (3.26)$$

Остальные вероятности находятся через p_0 .

Пример 3. На рисунке 3.7 приведен граф вероятностей системы. Необходимо рассчитать предельные вероятности состояний.

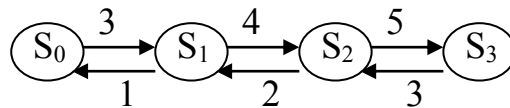


Рисунок 3.7.

На основе выведенных соотношений (3.23 – 3.26) получим:

$$p_0 = \left[1 + \frac{3}{1} + \frac{3 \cdot 4}{1 \cdot 2} + \dots + \frac{3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 3} \right]^{-1} = 0.05, \quad p_1 = 0.15, \quad p_2 = 0.3, \quad p_3 = 0.5.$$

3.2. Элементы теории массового обслуживания,

применяемые при моделировании систем

3.2.1. Основные определения и понятия

Под *системой массового обслуживания* (СМО) понимают динамическую систему, предназначенную для эффективного обслуживания потока заявок (требований на обслуживание) при ограничениях на ресурсы системы.

Первые задачи теории массового обслуживания (ТСМО) были рассмотрены сотрудниками Копенгагенской телефонной компании, датским ученым А.К. Эрлангом (1878 г. – 1929г.) в период между 1908 и 1922 гг. Эти задачи были обу-

словлены стремлением упорядочить работу телефонной сети и разработать методы, позволяющие заранее повысить качество обслуживания потребителей в зависимости от числа используемых устройств. Оказалось, что ситуации, возникающие на телефонных станциях, являются типичными не только для телефонной связи. Работа аэродромов, работа морских и речных портов, магазинов, библиотек, радиолокационных станций и комплексов, терминальных классов, вычислительных систем и сетей, автоматизированных систем управления и т.д. может быть описана в рамках ТСМО.

Пример 4. В качестве примера СМО приведем информационно-поисковую систему. Система включает распределенный банк данных, организованный на базе трех удаленных друг от друга вычислительных центров A , B и C . Все центры связаны между собой каналами передачи информации, работающими в дуплексном режиме независимо друг от друга. В каждый из центров поступают заявки на проведение информационного поиска, где они предварительно обрабатываются, в результате чего формируются запросы для центров A , B и C . В центре, получившем заявку от пользователя, начинается поиск информации по запросу, а на другие центры по соответствующим каналам передаются тексты запросов, после чего там также может начаться поиск информации. Тексты ответов передаются соответствующим каналом в центр, получивший заявку на поиск. Заявка считается выполненной, если получены ответы от всех трех центров.

Необходимо рассчитать характеристики эффективности функционирования информационно-поисковой системы: число заявок, поступивших и обслуженных в каждом центре; вероятность загрузки компьютеров центров; характеристики очереди запросов к компьютерам каждого центра; среднее время обслуживания заявок в каждом центре и т.д.

Характерным для приведенной задачи и для других подобных задач из различных предметных областей являются:

- условие двойной случайности: случайный момент времени поступления заявки на обслуживание (в информационно-поисковую систему, на телефонную

станцию, на пункт скорой помощи, на вход процессора и т.д.); случайная длительность времени обслуживания;

- наличие очередей (задач к компьютерам вычислительного центра, судов перед шлюзами, деталей на обработку и т.д.).

Реальные системы, как правило, очень сложны и включают в себя ряд этапов (стадий) обслуживания (рисунок 3.8). Причем на каждом этапе может существовать вероятность отказа в выполнении или возможна ситуация приоритетного обслуживания по отношению к другим требованиям. При этом отдельные звенья обработки заявок могут прекратить свою работу (например, для ремонта) или могут быть подключены дополнительные средства. Могут быть такие обстоятельства, когда требования, получившие отказ, вновь возвращаются в систему (подобная ситуация характерна для информационных систем).

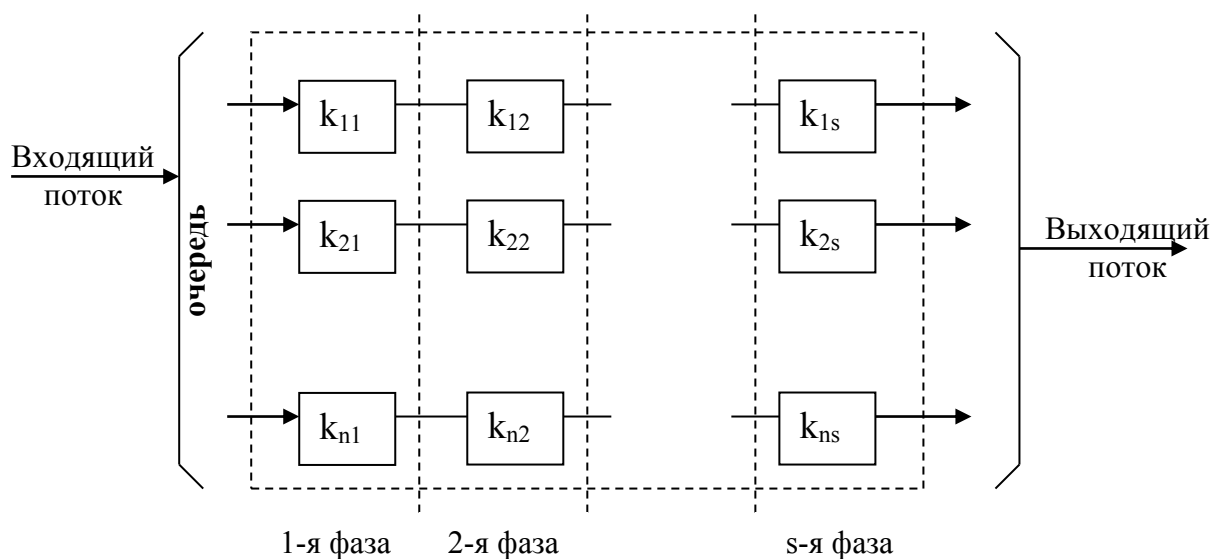


Рисунок 3.8. – Структура типовой n – канальной СМО

При описании СМО задают: входной поток (поток требований, поступающих в систему); выходной поток (поток обслуженных требований); систему обслуживания (совокупность очередей и приборов (каналов) обслуживания); правила формирования очереди и дисциплину обслуживания.

Классификация СМО. Выделяют различные признаки классификации СМО: системы с отказами и системы без отказов; замкнутые и разомкнутые; мно-

гоканальные и одноканальные; с равными (с точки зрения порядка обслуживания) заявками и с приоритетными заявками.

В *системе с отказами (с потерями, с конечной длиной очереди)* заявка, пришедшая в момент, когда все каналы обслуживания заняты или заняты все места в очереди, получает отказ и покидает систему. В *системе без отказов (без потерь, с бесконечной длиной очереди)* такая заявка не покидает систему, а становится в очередь и ждет, пока не освободится какой-нибудь канал. Время ожидания в общем случае неограниченно.

В *замкнутых СМО* в системе циркулирует определенное конечное число заявок (конечное число требований). В *разомкнутых СМО* количество заявок, поступающих в систему – теоретически бесконечно.

В n – канальной СМО одновременно могут обслуживаться n заявок (рисунке 3.8), в одноканальной системе – только одна заявка. В простейшем случае каждый канал характеризуется своей производительностью (интенсивностью обслуживания заявок). Если в СМО поступают заявки нескольких типов, то для каждого типа заявок может быть задана соответствующая интенсивность обслуживания. Интервал между поступлением заявок в систему и длительность обслуживания заявки каналом в общем случае – это случайные величины с заданными законами распределения.

Заявки в СМО могут иметь *одинаковый приоритет* и *разный приоритет*. При поступлении в СМО нескольких типов заявок с разными приоритетами могут быть организованы отдельные очереди для заявок каждого типа. Кроме размера, для каждой такой очереди указывается приоритет находящихся в ней заявок. Приоритеты обычно кодирует целыми числами $0, 1, 2, 3, \dots$, причем, чем меньше число, тем меньше приоритет соответствующих заявок. При наличии приоритетной организации в СМО на обслуживание в первую очередь выбираются заявки с высшими приоритетами.

Различают заявки с относительным и абсолютным приоритетами. Если заявка с абсолютным приоритетом поступила в СМО в тот момент, когда на обслуживании находятся заявка с меньшим приоритетом, то поступившая

заявка сразу начинает обслуживаться, прерывая на время своего обслуживания находящуюся там заявку. Вытесненная таким образом заявка возвращается в начало своей очереди и ожидает продолжение обслуживания (дообслуживания), либо выводится из системы. Для заявок с относительным приоритетом их приоритет вступает в действие не в момент их поступления в СМО, а в момент выбора следующей заявки из очереди (из очередей) на обслуживание. Прерываний в этом случае нет. Приоритеты заявок также бывают статические (постоянные) и динамические (изменяющиеся во времени).

Основные задачи ТСМО заключаются:

- в расчете выходных характеристик СМО, характеризующих эффективность функционирования системы (вероятность простоя системы, вероятность отказа в обслуживании, среднее число заявок в системе, среднее число занятых каналов, среднее время ожидания в очереди и т.д.);
- в оценке и анализе эффективности функционирования СМО на основе рассчитанных характеристик;
- в оптимизации характеристик функционирования СМО.

3.2.2. Типы систем массового обслуживания, описываемые Марковским случайным процессом «гибели и размножения»

Рассмотрим основные типы СМО, соответствующие схеме «гибели и размножения» [16].

1. Многоканальные (одноканальные) системы без потерь с неограниченным ожиданием и бесконечным потоком требований на входе (разомкнутые системы). На рисунке 3.9 приведен размеченный граф состояний системы.

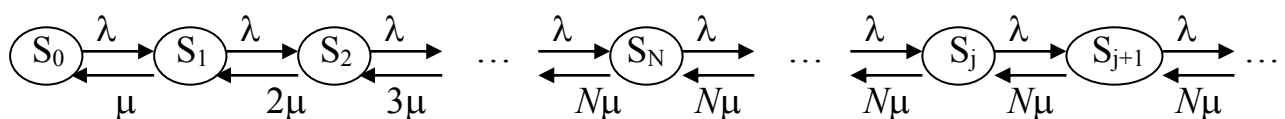


Рисунок 3.9.

μ – интенсивность потока обслуживания;

λ – интенсивность входного потока;

N – число каналов обслуживания;

2. Многоканальные (одноканальные) системы с отказами и бесконечным потоком требований на входе (разомкнутые системы). На рисунке 3.10 приведен размеченный граф состояний системы.

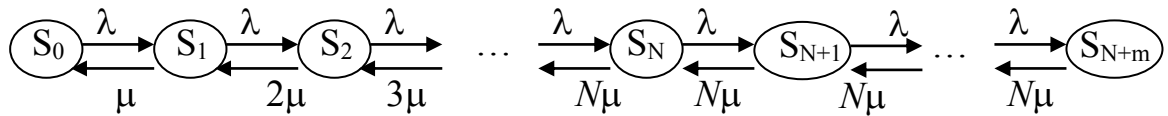


Рисунок 3.10.

m – величина, ограничивающая длину очереди (в другой терминологии – m – количество мест в накопителе очереди).

3. Многоканальные (одноканальные) системы без потерь с источником конечного числа требований (замкнутые системы). На рисунке 3.11 приведен размеченный граф состояний системы.

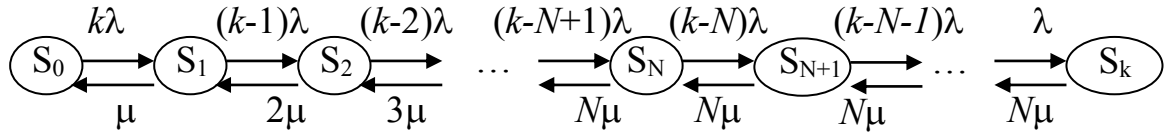


Рисунок 3.11.

k – максимальное число заявок в СМО.

3.2.3. Расчет характеристик типовых систем массового обслуживания

Приведем формулы для расчета вероятностей состояний и характеристик эффективности функционирования типовых систем массового обслуживания, рассмотренных в п. 3.2.2.

Для нахождения предельных вероятностей состояний составляется система уравнений Колмогорова. Ниже приведены конечные формулы для расчета предельных вероятностей состояний, выведенные из системы уравнений Колмогорова

ва. Характеристики эффективности функционирования системы рассчитываются на основе предельных вероятностей состояний.

1. Расчет многоканальных (одноканальных) систем без потерь с неограниченным ожиданием и бесконечным потоком требований на входе (разомкнутые системы)

Вероятности состояний системы:

$$p_0 = \left[1 + \sum_{j=1}^N a^j \frac{1}{j!} + a^{N+1} \frac{1}{N!(N-a)} \right]^{-1}, \quad a = \frac{\lambda}{\mu}, \quad (3.27)$$

$$p_j = p_0 \times \begin{cases} \frac{a^j}{j!}, & 1 \leq j \leq N; \\ \frac{a^j}{N! \cdot N^{j-N}}, & j > N. \end{cases} \quad (3.28)$$

Вероятность загрузки системы:

$$p_{\text{заг.}} = 1 - p_0. \quad (3.29)$$

Вероятность отказа в обслуживании: $p_{\text{отк.}} = 0$.

Относительная пропускная способность (средняя доля поступивших заявок, обслуживаемая системой – отношение среднего числа заявок, обслуживаемых системой в единицу времени, к среднему числу поступающих за это время заявок): $q = 1$, так как каждая заявка будет рано или поздно обслужена.

Абсолютная пропускная способность (среднее число заявок, которые может обслужить система в единицу времени):

$$A = \lambda q = \lambda. \quad (3.30)$$

Среднее число требований в очереди:

$$n_0 = \frac{a^{N+1} p_0}{N N! (1 - a_\Sigma)^2}, \quad a_\Sigma = \frac{\lambda}{N\mu}. \quad (3.31)$$

Среднее время ожидания в очереди:

$$t_{\text{ож.}} = \frac{a^N p_0}{N\mu N! (1 - a_\Sigma)^2}. \quad (3.32)$$

Среднее число занятых каналов:

$$N_3 = \frac{\lambda}{\mu}. \quad (3.33)$$

Среднее число заявок в системе:

$$J = n_0 + N_3. \quad (3.34)$$

Среднее время пребывания требования в системе:

$$t_c = t_{ож} + 1/\mu. \quad (3.35)$$

2. Расчет многоканальных систем с отказами и бесконечным потоком требований на входе (разомкнутые системы)

Вероятности состояний системы:

$$p_0 = \left[1 + \sum_{j=1}^N a^j \frac{1}{j!} + a^N \left(\frac{a_\Sigma - a_\Sigma^{m+1}}{N!(1 - a_\Sigma)} \right) \right]^{-1}, \quad a_\Sigma = \frac{\lambda}{N\mu}, \quad a = \frac{\lambda}{\mu}, \quad (3.36)$$

$$p_j = p_0 \times \begin{cases} \frac{a^j}{j!}, & 1 \leq j \leq N; \\ \frac{a^j}{N!N^{j-N}}, & N < j \leq N + m. \end{cases} \quad (3.37)$$

Вероятность загрузки системы:

$$p_{заг.} = 1 - p_0. \quad (3.38)$$

Вероятность отказа в обслуживании:

$$p_{отк.} = p_{N+m}. \quad (3.39)$$

Относительная пропускная способность:

$$q = 1 - p_{отк.}. \quad (3.40)$$

Абсолютная пропускная способность:

$$A = \lambda q. \quad (3.41)$$

Среднее число требований в очереди:

$$n_0 = p_{N+1} + 2p_{N+2} + \dots + mp_{N+m}. \quad (3.42)$$

Среднее время ожидания в очереди:

$$t_{ож.} = \frac{1}{N\mu} p_N + \frac{2}{N\mu} p_{N+1} + \dots + \frac{m}{N\mu} p_{N+m-1}. \quad (3.43)$$

Среднее число занятых каналов:

$$N_з = a(1 - \frac{a^{N+m}}{N^m N!} p_0). \quad (3.44)$$

Среднее число заявок в системе:

$$J = n_0 + N_з. \quad (3.45)$$

Среднее время пребывания требования в системе:

$$t_c = t_{ож.} + (1 - p_{отк.}) / \mu. \quad (3.46)$$

3. Расчет многоканальных систем без потерь с источником конечного числа требований (замкнутые системы)

Вероятности состояний системы:

$$p_0 = \left[1 + \sum_{j=1}^N a^j C_k^j + \sum_{j=N+1}^k a^j C_k^j \frac{j!}{N! N^{j-N}} \right]^{-1} \quad (3.47)$$

$$p_j = p_0 \times \begin{cases} \frac{a^j k!}{(k-j)! j!}, 1 \leq j \leq N; \\ \frac{a^j k!}{(k-j)! N! N^{j-N}}, N < j \leq k. \end{cases} \quad (3.48)$$

Вероятность загрузки системы:

$$p_{заг.} = 1 - p_0. \quad (3.49)$$

Вероятность отказа в обслуживании: $p_{отк.} = 0$.

Относительная пропускная способность: $q = 1$, так как каждая заявка будет рано или поздно обслужена.

Абсолютная пропускная способность:

$$A = (1 - p_0) \mu = \lambda. \quad (3.50)$$

Среднее число требований в очереди:

$$n_0 = \sum_{j=N+1}^k (j - N) p_j = \sum_{j=N+1}^k (j - N) C_k^j \frac{a^j j!}{N! N^{j-N}} p_0. \quad (3.51)$$

Среднее время ожидания в очереди:

$$t_{ож.} = \frac{J}{\lambda(k-J)} - \frac{1}{\mu}. \quad (3.52)$$

Среднее число заявок в системе:

$$J = n_0 + N_3. \quad (3.53)$$

Среднее число занятых каналов:

$$N_3 = N - \sum_{j=0}^{N-1} (N-j)p_j. \quad (3.54)$$

Среднее время пребывания требования в системе:

$$t_c = \frac{J}{\lambda(k-J)}. \quad (3.55)$$

3.2.4. Системы массового обслуживания с «взаимопомощью» между каналами

Системы массового обслуживания с «взаимопомощью» – системы, в которых одна и та же заявка может обслуживаться несколькими каналами. Взаимопомощь может быть организована в системе любого типа (замкнутой, разомкнутой).

При анализе таких систем необходим учет двух факторов:

1. Насколько убыстряется дисциплина обслуживания заявки при работе нескольких каналов. Самый простой вариант – увеличение интенсивности обслуживания пропорционально количеству обслуживающих каналов, т.е.: μk , где k – число каналов, занятых обслуживанием заявки.

2. Дисциплина взаимопомощи. Самый простой вариант – «все как один». Заявку обслуживают сразу все каналы.

Выгодно или нет вводить «взаимопомощь» зависит от реальной системы и ее параметров.

Пример 5. Имеется двухканальная система с потерями и бесконечным потоком заявок на входе. В очереди на обслуживание может стоять не более одной заявки. В системе протекают Пуассоновские потоки: интенсивность входного потока: $\lambda=6$; интенсивность потока обслуживания: $\mu=3$.

Выгодно ли вводить взаимопомощь по типу «все как один» с точки зрения увеличения загрузки системы? Выгодно ли это с точки зрения уменьшения времени пребывания заявки в системе?

На рисунке 3.12 приведен граф состояний системы без «взаимопомощи» каналов:

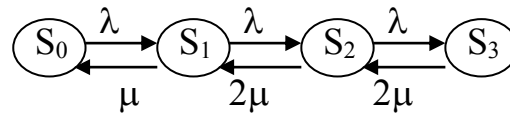


Рисунок 3.12.

В результате расчета по формулам (3.36 – 3.37) имеем:

$$p_0 \approx 0.145, \quad p_1 \approx 0.285, \quad p_2 \approx 0.285, \quad p_3 \approx 0.285.$$

Вероятность загрузки системы (3.38): $p_{заг.} \approx 0.855$.

Вероятность отказа в обслуживании (3.39): $p_{отк.} \approx 0.285$.

Среднее время пребывания заявки в системе (3.46): $t_c \approx 0.286$.

На рисунке 3.13 приведен граф состояний системы с «взаимопомощью» каналов:

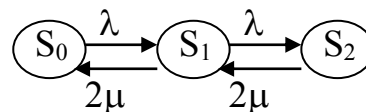


Рисунок 3.13

В результате расчета по формулам (3.36 – 3.37) имеем:

$$p_0 \approx 0.333, \quad p_1 \approx 0.333, \quad p_2 \approx 0.334.$$

Вероятность загрузки системы (3.38): $p_{заг.} \approx 0.667$.

Вероятность отказа в обслуживании (3.39): $p_{отк.} \approx 0.334$.

Среднее время пребывания заявки в системе (3.46): $t_c \approx 0.167$.

Таким образом, введение «взаимопомощи» между каналами привело к снижению загрузки системы. Это объясняется увеличением вероятности отказа: за то время, пока все каналы заняты обслуживанием одной заявки, могут прийти другие заявки, и, естественно, получить отказ. Среднее время пребывания заявки в системе, как и следовало ожидать, уменьшилось. В общем случае введение взаимопомощи целесообразно, когда необходимо уменьшить время пребывания заявки в

системе любыми средствами (например, пребывание в системе опасно для заявки).

3.2.5. Системы массового обслуживания с ошибками

На практике встречаются случаи, когда заявка, принятая к обслуживанию в СМО, выполняется не с полной достоверностью, а с некоторой вероятностью $p \neq 1$.

Примеры таких СМО: справочное бюро (не всегда выдает верные справки); корректор (может неверно исправить ошибку); телефонная станция (не всегда соединяет абонента с нужным номером) и т.д.

Пример 6. Добавим к задаче примера 5 условие: заявка правильно обслуживается с вероятностью $p = 2/3$. Граф состояний системы для этого случая имеет вид (рисунок 3.14):

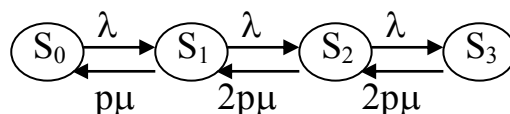


Рисунок 3.14.

Для разомкнутых СМО учет ошибок в обслуживании сводится к уменьшению относительной и абсолютной пропускных способностей системы в p раз. Характеристики системы, связанные с очередью, не изменяются, так как поток заявок на входе бесконечен и добавление к входному потоку заявок, обслуженных с ошибками, не изменяет его интенсивности.

В замкнутых системах заявка, обслуженная с ошибкой, вновь становится в очередь на обслуживание, и, следовательно, увеличивает загрузку СМО. Изменяются также все характеристики, связанные с очередями.

Еще одну разновидность СМО с ошибками в обслуживании представляют системы, в которых характер обслуживания зависит от длины очереди. При увеличении длины очереди канал начинает «спешить». Время обслуживания уменьшается, но увеличивается вероятность ошибки в обслуживании. Интенсивность обслуживания в этом случае вычисляется по формуле:

$$\mu^*(r) = \mu(r)p(r), \quad (3.56)$$

где $\mu(r)$ – интенсивность обслуживания при длине очереди r ;

$p(r)$ – вероятность обслуживания заявки при длине очереди r .

Пример 7. На рисунке 3.15 приведен граф состояний для одноканальной СМО с ограниченной длиной очереди равной двум, с ошибками в обслуживании в зависимости от длины очереди.

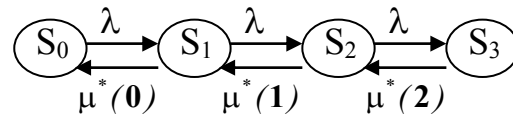


Рисунок 3.15.

Очевидно, что ошибки в обслуживании характерны для систем, в которых канал обслуживания – человек.

3.3. Контрольные вопросы и упражнения

1. Каковы основные понятия теории моделирования систем?
2. Какими основными свойствами обладает простейший Пуассоновский поток заявок? Поток Пальма? Поток Эрланга?
3. Что понимают под Марковским случайным процессом, под Марковским случайным процессом «гибели и размножения»? Приведите примеры.
4. Какова взаимосвязь между Марковским случайным процессом и Пуассоновским потоком заявок?
5. При каких условиях в непрерывной цепи Маркова существует стационарный режим. Приведите примеры систем, для которых стационарный режим возможен и для которых этот режим исключен.
6. Как рассчитать вероятности состояний системы в любой момент времени t , вероятности состояний СМО в стационарном режиме? Поясните на примере.
7. При каких условиях (допущениях) можно применять формулы п. 3.2.3 для расчета вероятностей состояний и характеристик эффективности функционирования системы?

8. В трехканальную СМО с отказами поступает простейший поток с плотностью $\lambda=0,4 \text{ с}^{-1}$. Интервал времени обслуживания $T_{об.}$ распределен по показательному закону, математическое ожидание $M[T_{об.}]=5\text{с}$. Постройте размеченный граф состояний системы. Вычислите вероятности состояний и выходные характеристики системы. Оцените эффективность функционирования системы.

9. К задаче п.8 добавлено условие: введена взаимопомощь между каналами типа «все как один». Постройте размеченный граф состояний системы. Как изменятся выходные характеристики системы? Сравните два варианта реализации СМО с точки зрения эффективности функционирования системы.

10. В замкнутой СМО обслуживаются заявки от 5-ти источников, число каналов обслуживания равно трем. Интенсивность входного потока заявок: $\lambda=5$; интенсивность потока обслуживания: $\mu=6$. Все потоки, протекающие в системе, – Пуассоновские. Постройте размеченный граф состояний системы. Вычислите вероятности состояний и выходные характеристики системы. Оцените эффективность функционирования системы.

11. К задаче п. 10 добавлено условие: вероятность безошибочного обслуживания заявки: $p=0.8$. Постройте размеченный граф состояний системы. Как изменятся выходные характеристики системы? Сравните два варианта реализации СМО с точки зрения эффективности функционирования системы.

4. Планирование машинных экспериментов с моделями систем

Имитационное моделирование является по своей сути – это машинный эксперимент с моделью исследуемой или проектируемой системы. План имитационного эксперимента на компьютере представляет собой метод получения с помощью эксперимента необходимой пользователю информации.

Цель проведения машинного эксперимента с моделью системы – получение информации о характеристиках процесса функционирования рассматриваемой

системы для анализа или оптимизации характеристик (параметров) системы при заданных ограничениях на ресурсы (затраты машинного времени, памяти и т.п.).

Пример. В качестве объекта моделирования рассматривается СМО с одним каналом обслуживания, бесконечным потоком требований на входе и ограниченной очередью. Необходимо определить взаимосвязь между интенсивностью входного потока заявок – λ , интенсивностью обслуживания – μ , количеством мест в очереди – L и средней длиной очереди заявок или вероятностью загрузки системы, проанализировать и подобрать оптимальные параметры системы λ , μ , L . Для решения этой задачи строится имитационная модель СМО, с которой проводится машинный эксперимент.

К достоинствам машинных экспериментов относят: возможность полного воспроизведения условий эксперимента с моделью исследуемой СМО; простоту прерывания и возобновления эксперимента.

Основной недостаток – наличие корреляции в выходных последовательностях, т.е. i – ое значение выходной характеристики зависит от предыдущих значений: $i-1$, $i-2$, ..., как следствие, в них содержится меньше информации, чем в независимых наблюдениях.

Эффективность машинных экспериментов с имитационными моделями систем существенно зависит от выбора плана эксперимента, так как именно план определяет объем и порядок проведения вычислений на компьютере, приемы накопления и статистической обработки результатов моделирования системы и в целом влияет на эффективность использования ресурсов компьютера при моделировании. Выделяют два вида планирования: стратегическое и тактическое планирование.

Стратегическое планирование ставит своей целью решить задачу получения необходимой информации о системе с помощью модели, реализованной на компьютере, с учетом ограничений имеющихся ресурсов. По своей сути стратегическое планирование аналогично внешнему проектированию при создании системы, только здесь в качестве объекта выступает процесс моделирования системы.

Тактическое планирование представляет собой определение способа проведения каждой серии испытаний машинной модели, предусмотренных планом эксперимента. Тактическое планирование аналогично внутреннему проектированию системы, но опять в качестве объекта рассматривается процесс работы с моделью.

4.1. Основные понятия теории планирования экспериментов

Математические методы планирования экспериментов основаны на модели типа «черного ящика», объект исследования представляется в виде структурной схемы:

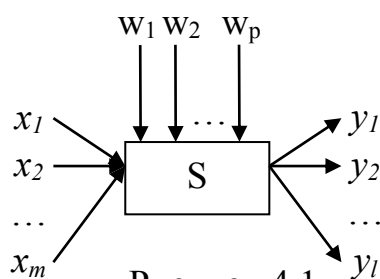


Рисунок 4.1

Имеем следующие группы параметров:

- 1.) $y_i, i = 1, l$ – зависимые выходные переменные, называемых реакциями или откликами (для машинного эксперимента эти переменные эндогенные);
- 2.) $x_i, i = 1, m$ – входные независимые переменные, называемых факторами (экзогенные переменные);
- 3.) $w_i, i = 1, p$ – случайные факторы (воздействия), не поддающиеся учету и контролю.

Факторы делятся на количественные и качественные; управляемые и неуправляемые (наблюдаемые); фиксированные и случайные. *Количественные факторы* измеряются числовыми значениями (например, интенсивности входного потока заявок и потока обслуживания, количество мест в очереди), а *качественным факторам* соответствует классификационная шкала или условно порядковая шкала (например, дисциплина постановки заявок в очередь или дисциплина обслуживания). Фактор называется *управляемым*, если его значения выби-

раются исследователем в процессе эксперимента и *неуправляемым*, если его значения наблюдаются и регистрируются. Фактор – *фиксированный*, если в ходе эксперимента исследуются все значения фактора, а если используется только случайная выборка из совокупности интересующих значений фактора, то фактор – *случайный*.

Каждый фактор принимает в ходе эксперимента значения, называемые *уровнями*: x_{ij} – j -ое значение i -го фактора, $i = \overline{1, m}$; $j = \overline{1, k}$. Каждому фиксированному набору уровней факторов соответствует точка в многомерном пространстве, называемом *факторным пространством*. Уровни факторов определяются до проведения эксперимента.

На практике целью многофакторного эксперимента является нахождение функции реакции, связывающей выходную переменную с совокупностью входных переменных:

$$y = f(x_1, x_2, \dots, x_m). \quad (4.1)$$

Функция реакции описывает поведение объекта, ее значение должно быть оптимизировано в результате эксперимента. Как правило, план эксперимента строится относительно одной выходной переменной y .

4.2. Этапы планирования и проведения эксперимента

1. Определение набора факторов и переменных отклика (реакции).

Выбирается переменная отклика, значение которой необходимо оптимизировать в результате эксперимента и отбираются факторы x_i , $i = \overline{1, m}$, влияющие на исследуемую реакцию системы. В процессе эксперимента обычно одновременно изменяется несколько факторов, поэтому совокупность факторов должна быть совместима и независима. Совместимость факторов означает, что все их комбинации осуществимы, а независимость соответствует возможности установления фактора на любом уровне независимо от уровней других.

Для примера 5 набор факторов: $x_1 = \lambda$ (интенсивность поступления заявок), $x_2 = \mu$ (интенсивность обслуживания заявок), $x_3 = L$ (число мест в очереди); переменная отклика: y – средняя длина очереди.

2. Определение вида функции реакции (структурная идентификация модели). Под структурой модели понимают вид функции f , связывающей совокупность факторов $x_1, x_2, \dots, x_m$ с реакцией системы y с точностью до ее неизвестных коэффициентов (параметров). Выбор структуры модели осуществляется на основе априорной информации. Этот этап трудно формализуем, связан с эвристическими решениями, в основном решается только с участием человека.

При планировании и проведении экспериментов наибольшее применение нашли модели в виде алгебраических полиномов. Чем выше степень полинома, тем точнее модель. Например, аппроксимация полиномом второго порядка функции реакции в однофакторной модели планирования может быть представлена в виде:

$$y = b_0 + b_1 x_1 + b_2 x_1^2. \quad (4.2)$$

Аппроксимация полиномом второго порядка функции реакции в m – факторной модели может быть представлена в виде:

$$y = b_0 + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{l=1}^m b_{ij} x_i x_l. \quad (4.3)$$

На практике стремятся свести нелинейную модель к линейной для упрощения и использования методов линейной регрессии при оценке параметров модели. Для этого вводят фиктивную переменную $x_{i,l} = x_i x_l$ и преобразуют модель к виду:

$$y = b_0 + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{l=1}^m b_{ij} x_{i,l}. \quad (4.4)$$

В результате получают модель множественной линейной регрессии. Выбранная модель включает в себя линейные члены полинома и их произведения,

т.е. учитывает эффект взаимодействия каждой пары факторов. В принципе можно построить модели для учета взаимодействия 3-х, 4-х, ..., m – факторов.

3. Определение уровней варьирования факторов. Выбирается нулевой уровень фактора, представляющий собой центр эксперимента (заданное значение) и каждый фактор варьируется на нескольких уровнях k относительно центра.

Эксперимент, в котором реализуются все возможные значения сочетания уровней факторов, называется *полным факторным экспериментом* (ПФЭ).

Если выбранная модель планирования включает в себя только линейные члены полинома и их произведения, то для оценки коэффициентов модели используется план эксперимента с варьированием всех m факторов на двух уровнях, т.е. $k=2$. Такие планы называют планами 2^m , где $N=2^m$ – число всех возможных испытаний.

В таблице 4.1, для примера, приведены уровни факторов и интервалы их варьирования. Для каждого фактора выбираются нижний и верхний уровни: $x_{in} = x_{i\min}$ (кодируется: -1) и $x_{iv} = x_{i\max}$ (кодируется: +1).

Таблица 4.1. – Уровни варьирования факторов

Факторы	Уровни факторов			Интервалы варьирования
	-1	0	+1	
x_1	45	50	55	5
x_2	41	45	49	4
x_3	2	3	4	1

4. Составление матрицы планирования эксперимента. В таблице 4.2 приведена матрица планирования для проведения полного факторного эксперимента. Полный факторный эксперимент дает возможность определить коэффициенты регрессии, соответствующие не только линейным эффектам, но и всем эффектам взаимодействий. При проведении эксперимента используют либо исходные значения факторов, либо стандартизованные значения (для удобства). Для перехода к стандартизованным значениям применяют преобразование:

$$\tilde{x}_i = \frac{(x_i - x_{i0})}{\Delta x_i}, \quad (4.5)$$

где x_{i0} – значение, соответствующее нулевому уровню фактора;

Δx_i – интервал варьирования i – го фактора.

Таблица 4.2. – Матрица планирования эксперимента

Номер опыта	План ПФЭ							y
	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_1\tilde{x}_2$	$\tilde{x}_2\tilde{x}_3$	$\tilde{x}_1\tilde{x}_3$	$\tilde{x}_1\tilde{x}_2\tilde{x}_3$	
1.	-1	-1	-1	+1	+1	+1	-1	y_1
2.	-1	-1	+1	+1	-1	-1	+1	y_2
3.	-1	+1	-1	-1	-1	+1	+1	y_3
4.	-1	+1	+1	-1	+1	-1	-1	y_4
5.	+1	-1	-1	-1	+1	-1	+1	y_5
6.	+1	-1	+1	-1	-1	+1	-1	y_6
7.	+1	+1	-1	+1	-1	-1	-1	y_7
8.	+1	+1	+1	+1	+1	+1	+1	y_8

Важны общие свойства матрицы планирования:

1. **Симметричность** матрицы относительно центра эксперимента:

$$\sum_{j=1}^N x_{ij} = 0. \quad (4.6)$$

2. Выполнение условия **нормировки**, т.е. сумма квадратов элементов каждого столбца равна числу опытов:

$$\sum_{j=1}^N x_{ij}^2 = N. \quad (4.7)$$

3. **Ротатабельность**. Это означает, что точки (значения факторов) в матрице планирования подбираются так, что точность предсказания выходной переменной должна быть одинакова на равных расстояниях от центра эксперимента (нулевого уровня) и не зависеть от направления.

Количество испытаний в эксперименте должно быть больше или равно количеству определяемых коэффициентов полиномиальной модели. Если в модели необходим учет только линейных эффектов, достаточно провести $N=m+1$ испытаний, где m – число факторов (+1 для определения свободного члена в уравне-

нии). Поэтому можно построить *дробный факторный эксперимент* (ДФЭ), включающий всего четыре испытания для приведенного примера. Также ДФЭ строится в случае, если необходимо в модели учесть не все взаимодействия. Правило проведения ДФЭ формулируется так: для сокращения числа испытаний новому фактору присваивается значение вектор – столбца матрицы, принадлежащего взаимодействию, которым можно пренебречь.

При проведении эксперимента из четырех испытаний для оценки влияния трех факторов пользуются половиной ПФЭ типа 2^3 , так называемой «полуреplikой». При построении «полуреплики» 2^{3-1} $\tilde{x}_3$ можно приравнять к $\tilde{x}_1\tilde{x}_2$ или к $\tilde{x}_1\tilde{x}_2$.

Кроме планов 2^m применяют также многоуровневые планы, в которых факторы варьируются на 3-х, 4-х, ..., k – уровнях и обозначаются соответственно 3^m , 4^m , ..., k^m – планы.

Пример. Графическое представление плана эксперимента первого порядка для двух переменных показано на рисунке 4.2. Искомая функция $y = f(x_1, x_2)$ описывается моделью в виде плоскости:

$$y = b_0 + b_1x_1 + b_2x_2. \quad (4.8)$$

или гиперboloида:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2. \quad (4.9)$$

Расположение этой модели в пространстве показано на рисунке 4.2 поверхностью, проходящей через точки 1-2-3-4.

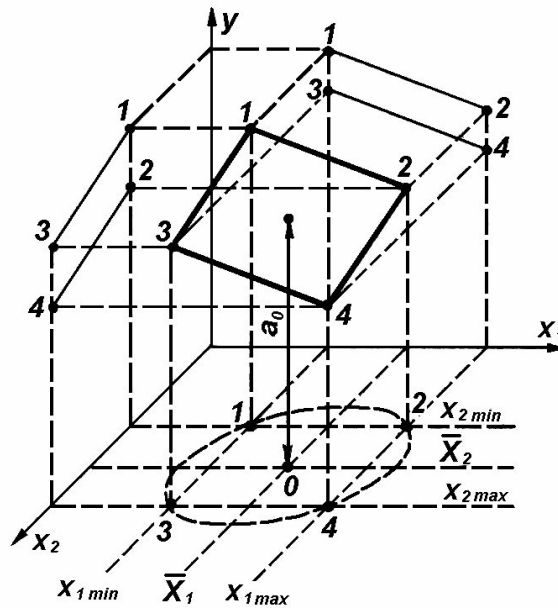


Рисунок 4.2.

5. Проведение эксперимента. Проводится эксперимент в соответствии с составленной матрицей планирования, измеряется реакция системы для каждого сочетания уровней фактора, т.е. в заданных точках факторного пространства. В результате получают исходные данные для определения коэффициентов регрессионной модели. Для более точного определения коэффициентов функции реакции эксперимент с матрицей планирования проводят неоднократно. Т.е. проводится серия n параллельных экспериментов (n прогонов), каждая серия состоит из N испытаний. В теории планирования рекомендуется реализовать не менее пяти серий экспериментов: $n \geq 5$.

6. Определение однородности серий экспериментов.

Для каждой строки матрицы планирования по результатам n параллельных экспериментов находят среднее арифметическое значение выходной переменной:

$$\bar{y}_j = \frac{1}{n} \sum_{v=1}^n y_{jv}, \quad (4.10)$$

где v – номер параллельного эксперимента;

y_{jv} – значение v -м параллельном эксперименте j -ой строки матрицы.

С целью оценки отклонений выходной переменной от ее среднего значения для каждой строки матрицы планирования вычисляют дисперсию s_j^2 эксперимен-

та по данным n параллельных экспериментов. Статистической дисперсией называют среднее значение квадрата отклонений случайной величины от ее среднего значения:

$$s_j^2 = \frac{1}{n-1} \sum_{v=1}^n (y_{jv} - \bar{y}_j)^2. \quad (4.11)$$

Ошибка эксперимента s_j определяется как корень квадратный из дисперсии:

$$s_j = + \sqrt{\frac{1}{n-1} \sum_{v=1}^n (y_{jv} - \bar{y}_j)^2}. \quad (4.12)$$

В этом случае ошибка при большом рассеянии будет значительной. Рассеяние результатов эксперимента определяется влиянием неуправляемых факторов, погрешностями измерений и другими причинами. Большое рассеяние изучаемой величины может произойти из-за наличия в эксперименте сомнительных результатов.

Проверяют гипотезу однородности дисперсий. Проверка однородности двух дисперсий производится с помощью F -критерия Фишера. Критическая статистика критерия представляет собой отношение большей дисперсии к меньшей:

$$F_{кр.} = \frac{s_1^2}{s_2^2}, \quad (4.13)$$

где $s_1^2 > s_2^2$.

Критическая статистика подчиняется распределению Фишера с $(n-1, n-1)$ числом степеней свободы. Для проверки гипотезы равенства дисперсий расчетное значение критерия $F_{кр.}$ сравнивается с критической границей F_t (см. таблицу 4.3). Критическая граница — это квантиль распределения Фишера уровня α : $F_t = F_\alpha(n-1, n-1)$. Если $F_{кр.} < F_t$ то гипотеза не отвергается при заданном уровне значимости α и дисперсии можно считать однородными.

Таблица 4.3 – Критические точки распределения Фишера при уровне значимости 0,05

$\alpha=0.05$		число степеней свободы для большей дисперсии ν_1																		
		1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120
число степеней свободы для меньшей дисперсии ν_2	1	161	200	216	225	230	234	237	239	271	242	243	244	246	248	249	250	251	252	253
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,71	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,42	2,35	2,32	2,29	2,25	2,21	2,16
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,35	2,28	2,24	2,21	2,17	2,13	2,08
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,31	2,23	2,19	2,15	2,11	2,06	2,01
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
	120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,87	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35

Однородность ряда дисперсий проверяют по критерию Кохрена или по критерию Бартлета. При равномерном дублировании экспериментов однородность ряда дисперсий проверяют с помощью G -критерия Кохрена, представляющего собой отношение максимальной дисперсии к сумме всех дисперсий:

$$G_p = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_N^2} = \frac{s_{\max}^2}{\sum_{j=1}^N s_j^2} \quad (4.14)$$

Дисперсии однородны, если расчетное значение G_p - критерия не превышает табличного значения G_t - критерия. В таблице 4.4. N показывает число сравниваемых дисперсий, n – число параллельных экспериментов. Если $G_p > G_t$, то дисперсии неоднородны, а это указывает на то, что исследуемая величина y не подчиняется нормальному закону. В этом случае нужно попытаться заменить y случайной величиной $q=f(y)$, достаточно близко следующей нормальному закону. Если дисперсии экспериментов однородны, то вычисляют дисперсию воспроизводимости:

$$s_y^2 = \frac{1}{N} \sum_{j=1}^N s_j^2, \quad (4.15)$$

где N – число экспериментов или число строк матрицы планирования.

Таблица 4.4. Критические точки G -распределения при уровне значимости 0,05

N	$n-l$								
	1	2	3	4	5	6	7	8	9
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439
12	0,5410	0,3924	0,3624	0,2880	0,2624	0,2439	0,2299	0,2187	0,2098
15	0,4709	0,3346	0,2758	0,2419	0,2195	0,2034	0,1911	0,1815	0,1736
20	0,3894	0,2705	0,2205	0,1921	0,1735	0,1602	0,1501	0,1422	0,1357

По результатам эксперимента вычисляют коэффициенты модели.

7. Определение главных эффектов факторов и эффектов взаимодействия. Главным эффектом фактора i (ГЭФ) называется средняя величина изменения в реакции, обусловленная переходом i – го фактора с нижнего уровня («-») на верхний уровень («+») при условии, что другие факторы остаются без изменений.

Для факторного плана 2^3 (см. таблицу 4.2) главный эффект фактора 1 опре-

деляется как:

$$e_1 = \frac{(y_5 - y_1) + (y_6 - y_2) + (y_7 - y_3) + (y_8 - y_4)}{4} \quad (4.16)$$

или аналогично:

$$e_1 = \frac{-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8}{2^{m-1}}. \quad (4.17)$$

ГЭФ интерпретируется следующим образом: средний эффект от увеличения интенсивности входного потока с 45 до 55 состоит в увеличении длины очереди на e_1 единиц.

Эффектом взаимодействия между факторами называется разность между средней реакцией системы, когда факторы на одинаковом уровне («+» или «-») и средней реакцией системы, когда факторы на разных уровнях. Например, эффект взаимодействия между первым и вторым факторами определяется выражением:

$$e_{12} = \frac{y_1 + y_2 - y_3 - y_4 - y_5 - y_6 + y_7 + y_8}{2^{m-1}}. \quad (4.18)$$

Так как реакции системы y – случайная величина, то эффекты также являются случайными. Для оценки статистической значимости эффектов необходимо выполнить следующее:

- провести эксперимент в соответствии с матрицей планирования n раз и получить n независимых значений каждого эффекта;
- построить $100(1 - \alpha)\%$ доверительный интервал для ожидаемого эффекта с использованием t – распределения Стьюдента с $(n-1)$ степенями свободы по формуле:

$$e_i \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{D}{n}}, \quad (4.19)$$

где $t_{n-1, 1-\alpha/2}$ – квантиль распределения Стьюдента уровня $1 - \alpha/2$.

Если доверительный интервал для эффекта, соответствующего i – ому фактору, не содержит ноль, то эффект статистически значим при уровне α .

В результате проведения этого этапа структурная модель уравнения реакции может быть упрощена путем исключения факторов или их взаимодействий,

для которых эффекты статистически незначимы.

8. Определение параметров функции реакции (параметрическая идентификация). Параметрическая идентификация сводится к оценке параметров функции φ на основе апостериорной (после опыта) информации. Задача обычно ставится как задача математического программирования, так как при этом минимизируется некоторый функционал, являющийся ущербом, наносимым процессом идентификации.

Функция реакции $Y = f(X)$ описывает изменение условного среднего значения выходной переменной Y (при условии, что значения входных переменных X зафиксированы на уровнях X^*) в зависимости от изменения значений X .

Математически:

$$f(X^*) = M\{Y | X = X^*\} = \bar{f}(X^*) \quad (4.20)$$

или сокращенно:

$$f(X) = M\{Y | X\} = \bar{f}(X). \quad (4.21)$$

В сокращенной записи не различают обозначения входных переменных X и их возможных значений X^* .

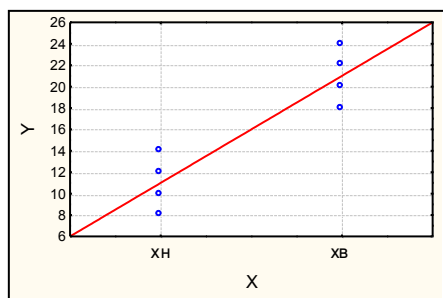


Рисунок 3.3

Пример 6:

$$M\{Y | x_n\} = \bar{\varphi}(x_n) = \frac{1}{m_1} (y_{11} + y_{12} + \dots + y_{1m_1}),$$

$$M\{Y | x_6\} = \bar{\varphi}(x_6) = \frac{1}{m_2} (y_{21} + y_{22} + \dots + y_{2m_2}),$$

где m_1, m_2 – количество наблюдений при x_n, x_6 .

Для нахождения оценок коэффициентов функции реакции, как правило, используют метод наименьших квадратов (МНК). В основе МНК лежит критерий: сумма квадратов отклонений экспериментальных наблюдений от построенной линии функции реакции минимальна. В случае линейной зависимости между реакцией и факторами системы критерий МНК имеет вид:

$$Q = \sum_{j=1}^N (b_0 + \sum_{i=1}^m b_i x_{ij} - y_j)^2 = \sum_{j=1}^N \varepsilon_j^2 \rightarrow \min, \quad (4.22)$$

где ε_j – случайный остаток (отклонение);

$b_i, i = \overline{0, m}$ – неизвестные коэффициенты функции реакции.

По сути, критерий МНК – это обобщенный показатель разброса экспериментальных наблюдений относительно искомой линейной функции реакции.

Оценки коэффициентов $\hat{b}_j$ находят из условия равенства частных производных функции Q нулю в точке минимума:

$$\frac{\partial Q}{\partial b_0} = \frac{\partial Q}{\partial b_1} = \dots = \frac{\partial Q}{\partial b_m} = 0. \quad (4.23)$$

Приведем конечные формулы, полученные в результате применения МНК. Оценку свободного члена определяют по формуле:

$$\hat{b}_0 = \frac{1}{N} \sum_{j=1}^N \bar{y}_j. \quad (4.24)$$

Оценки коэффициентов регрессии, характеризующие линейные эффекты, вычисляют по зависимости:

$$\hat{b}_i = \frac{1}{N} \sum_{j=1}^N x_{ij} \bar{y}_j. \quad (4.25)$$

Оценки коэффициентов регрессии, характеризующие эффекты взаимодействия, определяют по формуле:

$$\hat{b}_{il} = \frac{1}{N} \sum_{j=1}^N x_{ij} x_{lj} \bar{y}_j, \quad (4.26)$$

где i, l – номера факторов; x_{ij}, x_{lj} – кодированные значения факторов i и l в j -м эксперименте.

В результате функция реакции, учитывающая только линейные эффекты, имеет вид:

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^m \hat{b}_i x_i + \varepsilon. \quad (4.27)$$

Функция реакции, учитывающая линейные эффекты и эффекты взаимодействия имеет вид:

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^m \hat{b}_i x_i + \sum_{i=1}^m \sum_{l=1}^m \hat{b}_{il} x_{i,l} + \varepsilon \quad (4.28)$$

В случае построения нескольких моделей функции реакции критерий МНК может послужить основой для выбора окончательного вида модели. Лучшей считается модель, для которой критерий МНК минимален. Это один из статистических приемов выбора вида функции реакции.

В случае нелинейной модели используют преобразования для сведения к линейному виду: например, введение фиктивных переменных, логарифмирование и т.д. Далее применяют МНК к преобразованной модели.

9. Анализ точности и адекватности построенной функции реакции. На этом этапе рассчитываются характеристики точности построенной функции реакции, и оценивается ее адекватность экспериментальным наблюдениям. Приведем основные статистические характеристики и критерии, которые используют для оценки функции реакции.

Проверяется статистическая значимость коэффициентов модели на основе сравнения абсолютной величины коэффициента с доверительным интервалом или с помощью критерия Стьюдента.

1. Сравнение абсолютной величины коэффициента с доверительным интервалом.

Вычисляют дисперсии коэффициентов регрессии для построения доверительного интервала. Дисперсию $s^2\{b_i\}$ i -го коэффициента определяют по зависимости:

$$s^2\{b_i\} = \frac{1}{nN} s_y^2. \quad (4.29)$$

Доверительный интервал Δb_i находят по формуле:

$$\Delta b_i = \pm t_{\tau} s\{b_i\}, \quad (4.30)$$

где t_{τ} – критическая точка распределения Стьюдента (квантиль распределения) при принятом уровне значимости и числе степеней свободы f , с которым определялась дисперсия s_y^2 . При равномерном дублировании экспериментов число степеней свободы находится по зависимости:

$$f = (n - 1)N,$$

где N – число экспериментов в матрице планирования, n – число параллельных экспериментов, $s\{b_i\}$ – ошибка в определении i -го коэффициента регрессии, вычисляемая по формуле:

$$s\{b_i\} = +\sqrt{s^2\{b_i\}}. \quad (4.31)$$

Значения квантилей распределения Стьюдента приведены в таблице 4.5.

Коэффициент значим, если его абсолютная величина больше доверительного интервала.

- Критерий Стьюдента применяется для оценки статистической значимости коэффициентов функции реакции. Если проверяется гипотеза $H_0: \hat{b}_i = 0$, то критическая статистика критерия имеет вид:

$$t_p = \frac{|b_i|}{s\{b_i\}}. \quad (4.32)$$

Расчетное значение критической статистики сравнивают с табличным t_t (см. таблицу 4.5). Коэффициент значим, если выполняется условие: $t_p > t_t$ для принятого уровня значимости и числа степеней свободы, с которым определялась дисперсия s_y^2 .

Таблица 4.5. Критические точки t -распределения при уровне значимости 0,05

Число степеней свободы	1	2	3	4	5	6	7	8
Значение	12,71	4,3	3,18	2,78	2,57	2,45	2,37	2,3
Число степеней свободы	9	10	11	12	13	14	15	16
Значение	2,26	2,23	2,2	2,18	2,16	2,14	2,13	2,12
Число степеней свободы	17	18	19	20	21	22	23	24
Значение	2,11	2,1	2,09	2,09	2,08	2,07	2,07	2,06
Число степеней свободы	25	26	27	28	29	30	31	32
Значение	2,06	2,06	2,05	2,05	2,05	2,04	2,02	2,02

Критерий Стьюдента вычисляют для каждого коэффициента регрессии. Статистически незначимые коэффициенты могут быть исключены из уравнения.

3. *Критерий Фишера* позволяет оценить статистическую значимость построенной функции реакции и проверяет основную гипотезу

$$H_0 : \hat{b}_1 = \hat{b}_2 = \dots = \hat{b}_k = 0.$$

Критическая статистика критерия имеет вид:

$$F_p = \frac{s_{ад}^2}{s_y^2}. \quad (4.33)$$

и подчиняется распределению Фишера с $(k, n - k - 1)$ степенями свободы, где k – количество факторов в модели.

В формуле критической статистики критерия используется дисперсия адекватности $s_{ад}^2$. Дисперсия адекватности или остаточная дисперсия, характеризует рассеяние эмпирических (наблюдаемых) значений y относительно расчетных $\hat{y}$, определенных по найденному уравнению регрессии. Дисперсию адекватности определяют по формуле:

$$s_{ад}^2 = \frac{n \sum_{j=1}^N (\bar{y}_j - \hat{y}_j)^2}{f} = \frac{n \sum_{j=1}^N (\bar{y}_j - \hat{y}_j)^2}{N - (k + 1)}, \quad (4.34)$$

где $\bar{y}_j$ – среднее арифметическое значение выходной характеристики в j -м эксперименте; $\hat{y}_j$ – значение выходной характеристики, вычисленное по модели для условий j -го опыта;

Если выполняется условие:

$$F_{p.} < F_m, \quad (4.35)$$

где F_m – квантиль распределения Фишера для принятого уровня значимости α и числом степеней свободы $(k, n - k - 1)$, то гипотеза H_0 не отвергается при уровне значимости α . По сути, критерий позволяет оценить в какой мере замена $\bar{y}_j$ на $\hat{y}_j$ позволяет увеличить точность прогнозирования y по заданным значениям x_{ij} . Если гипотеза H_0 не отвергается, то использовать уравнение реакции нецелесообразно, зависимость Y от X точнее описывается просто средним.

10. Оптимизация параметров исследуемой системы. Анализ и интерпретация полученных результатов. На заключительном этапе в простейшем случае решается оптимизационная задача вида:

$$\begin{aligned} y &= f(x_1, x_2, \dots, x_m) \rightarrow \min(\max) \\ x_{in} &\leq x_i \leq x_{ie}, i = \overline{1, m} \\ &\dots \\ x_i &\geq 0 \end{aligned} \quad (4.36)$$

для нахождения оптимального набора параметров системы, максимизирующего (минимизирующего) исследуемую реакцию системы. Также подводятся итоги проведенного исследования, анализируются и оцениваются полученные результаты, формулируются предложения по улучшению процесса функционирования системы.

Рассмотренные этапы планирования машинных экспериментов с моделью системы итерационно взаимодействуют, последовательность этапов жестко не задана.

4.3. Контрольные вопросы и упражнения

1. Каковы цели проведения машинных экспериментов с моделью системы? Перечислите характерные особенности машинного эксперимента по сравнению с другими видами экспериментов.

2. Какие существуют виды машинных экспериментов? В чем их особенности?

3. Каковы основные этапы планирования и проведения машинных экспериментов? Какие этапы на Ваш взгляд являются ключевыми?

4. В структурную модель функции реакции включено пять факторов. Сколько испытаний будет проведено в ПФЭ? Если в модели необходимо учесть: 1) только линейные эффекты; 2) линейные эффекты и все парные эффекты взаимодействия факторов, сколько следует провести испытаний? Составьте матрицу планирования соответствующую ДФЭ для этих случаев.

5. Приведите пример системы, соответствующей схеме М/М/3/2 (Пуассоновский поток поступления и обслуживания заявок; три канала; два места в очереди). Для этой системы сформулируйте возможную цель (цели) проведения машинного эксперимента, определите набор факторов и реакцию (реакции) системы, продумайте схему планирования и проведения эксперимента.

5. Список литературы

Основная литература

1. Советов Б.Я., Яковлев С.А. Моделирование систем (7-е изд.). – М.: Издательство Юрайт, 2012. – 343 с.
2. Советов Б.Я., Яковлев С.А. Моделирование систем. Практикум: Учеб. пособие для вузов (4-е изд., перераб. и доп.) – М.: Издательство Юрайт, 2012. – 295 с.
3. Кельтон В., Лоу А. Имитационное моделирование. Классика CS. 3-е изд. – СПб.: Питер; Киев: 2004. – 847 с.
4. Алгазинов Э.К., Сирота А.А. Анализ и компьютерное моделирование информационных процессов и систем / под ред. А.А. Сирота. – М.: Диалог-МИФИ, 2009. – 416 с.
5. Максимей И.В. Имитационное моделирование сложных систем: учебное пособие. 1 часть. Математические основы / В.И. Максимей. – Минск: Изд. Центр БГУ, 2009. – 263 с.
6. Моделирование систем: учебник для ВУЗов / С.И. Дворецкий, Ю.Л. Муромцев, В.А. Погонин, А.Г. Схиртладзе. – М.: Издательский центр «Академия», 2009. – 320 с.
7. Боев В.Д. Моделирование систем. Инструментальные средства GPSS World: учеб. пособие. – СПб.: БХВ-Петербург, 2004. – 368 с.
8. Дигрис А.В. Дискретно-событийное моделирование, Минск, 2011. [электронный ресурс]: <http://simulation.su/uploads/files/default/2011-uch-posob-digris-diskr-sob-mod.pdf>.

9. Расширенный редактор GPSS World: учеб. пособие. [электронный ресурс]: http://www.elina-computer.ru/uploads/files/GPSS_EE_Help.pdf.
10. Учебное руководство по GPSS World. [электронный ресурс]: http://open.ifmo.ru/images/5/56/33071_gpss_world_tutorial.pdf.
11. Русаков А.М. Исследование и моделирование сложных систем. Москва, 2012. [электронный ресурс]: <http://simulation.su/uploads/files/default/2014-res-and-sim-sys.pdf>.

Дополнительная литература

12. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 487 с.
13. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. – М.: «Финансы и статистика», 1983. – 471 с.
14. Альсова О.К. Моделирование систем: учебное пособие / О.К. Альсова. – Новосибирск: Изд-во НГТУ, 2007. – 72 с.
15. Арсеньев Б.П., Яковлев С.А. Интеграция распределенных баз данных. – СПб.: Лань, 2001. – 420 с.
16. Вентцель Е.С. Исследование операций. – М.: Высшая школа, 2000. – 550 с.
17. Губарев В.В. Вероятностные модели / Новосиб. электротехн. ин-т. – Новосибирск, 1992. – Ч.1. – 198 с; Ч.2. – 188 с.
18. Губарев В.В. Системный анализ в экспериментальных исследованиях. – Новосибирск: Изд-во НГТУ, 2000. – 99 с.
19. Денисов А.А., Колесников Д.Н. Теория больших систем управления: Учеб. пособие для вузов. – Л. Энергоиздат, 1982. – 288 с.
20. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1973.
21. Карпов Ю. Имитационное моделирование систем. Введение в моделирование с AnyLogic 5. – СПб.: БХВ-Петербург, 2005. – 400 с.

- 22.Колемаев В.А. и др. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1991.
- 23.Кудрявцев Е.М. GPSS World. Основы имитационного моделирования различных систем. – М.: ДМК Пресс, 2004. – 320 с.
- 24.Потапов В.Д., Яризов А.Д. Имитационное моделирование. – М.: Высшая школа, 1981.
- 25.Рыжиков Ю.И. Имитационное моделирование. Теория и технологии. – СПб.: КОРОНА принт; М.: Альтекс-А, 2004. – 384 с.
- 26.Сирота А.А. Компьютерное моделирование и оценка эффективности сложных систем. – М.: Техносфера, 2006. – 280 с.
- 27.Теория случайных процессов и ее инженерные приложения: Учеб. пособие для вузов/Е.С. Вентцель, Л.А. Овчаров. – 3-е изд. перераб. и доп. – М.: Издательский центр «Академия», 2003. – 432 с.
- 28.Томашевский В., Жданова Е. Имитационное моделирование в среде GPSS. – М.: Бестселлер, 2003. – 416 с.
- 29.Шеннон Р. Имитационное моделирование систем – искусство и наука. – М.: Мир, 1978.