

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ОБРАЗОВАНИЯ

Государственное образовательное учреждение
высшего профессионального образования
«Оренбургский государственный университет»

В.В. ПАНИЧЕВ,
Н.А. СОЛОВЬЕВ

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Рекомендовано Ученым советом государственного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по специальности «Программное обеспечение вычислительной техники и автоматизированных систем»

Оренбург 2008

УДК 519.876.5
ББК 22.18
С 56

Рецензент
кандидат технических наук И. Шевченко

С 56 **Паничев В.В.**
Компьютерное моделирование: учебное пособие / В.В. Паничев,
Н.А. Соловьев – Оренбург: ГОУ ОГУ, 2008. – 130 с.

ISBN 5-06-004087-9

В пособии рассмотрены основы теории моделирования систем; принципы и методы математического и имитационного моделирования динамических систем и систем массового обслуживания. Теоретический материал дополнен примерами и задачами с решениями, вопросами для самопроверки.

Учебное пособие предназначено для студентов, обучающихся по программам высшего профессионального образования по специальности 230105.65 «Программное обеспечение вычислительной техники и автоматизированных систем», при изучении дисциплины «Компьютерное моделирование».

УДК 519.876.5
ББК 22.18

© Паничев В.В., 2008
© ГОУ ОГУ, 2008

ISBN

Содержание

	стр.
Введение.....	5
1 Принципы построения математических моделей систем.....	6
1.1 Основные понятия теории моделирования систем.....	6
1.1.1 Понятие модели. Виды моделирования.....	6
1.1.2 Этапы компьютерного моделирования.....	8
1.1.3 Принципы моделирования.....	10
1.1.4 Математическая модель.....	11
1.1.5 Классификация математических моделей.....	14
1.2 Моделирование сложных систем.....	18
1.2.1 Основные понятия и определения.....	18
1.2.2 Модель сложной системы.....	20
1.2.3 Ограничения на параметры и характеристики модели.....	22
1.2.4 Общий подход к формированию математических моделей.....	23
1.3 Типовые математические схемы моделирования.....	25
Вопросы для самоконтроля.....	29
2 Принципы имитационного моделирования систем.....	30
2.1 Статистическое моделирование систем.....	30
2.1.1 Характеристика методов моделирования вероятностных объектов.....	30
2.1.2 Формирование базовой последовательности случайных чисел.....	33
2.1.3 Моделирование случайных событий.....	35
2.1.4 Моделирование случайных величин.....	36
2.2 Имитационное моделирование сложных систем.....	39
2.2.1 Характеристика имитационных моделей.....	39
2.2.2 Принципы изменения модельного времени.....	40
2.2.3 Этапы имитационного моделирования системы.....	43
2.2.4 Способы имитации поведения системы.....	47
Вопросы для самоконтроля.....	48
3 Эксперимент с моделью. Обработка результатов моделирования.....	49
3.1 Планирование эксперимента.....	49
3.1.1 Методы теории планирования.....	49
3.1.2 Стратегическое планирование эксперимента.....	50
3.1.3 Обеспечение точности и достоверности результатов моделирования.....	53
3.2 Статистический анализ результатов моделирования.....	60
3.2.1 Оценивание вероятностных распределений и их числовых характеристик.....	60
3.2.2 Проверка адекватности моделей.....	63
3.2.3 Проверка устойчивости и чувствительности моделей.....	67
3.2.4 Критерии согласия.....	69
3.3 Статистическое исследование зависимостей.....	72
анализ.....	3
7.2 Корреляционный анализ.....	
анализ.....	

3.3.1 Дисперсионный анализ.....	72
3.3.2 Корреляционный анализ.....	74
3.3.3 Регрессионный анализ.....	76
Вопросы для самоконтроля.....	77
4 Моделирование динамических систем.....	78
4.1 Идентификация динамических систем.....	78
4.2 Алгоритмизация непрерывно – детерминированных моделей.....	81
4.2.1 Решетчатые функции. Разностные и рекуррентные уравнения.....	81
4.2.2 Алгоритмизация НДМ на основе z преобразований.....	85
4.2.3 Особенности алгоритмизации моделей систем автоматического управления.....	89
Вопросы для самоконтроля.....	92
5 Моделирование систем массового обслуживания.....	93
5.1 Аналитические модели систем массового обслуживания.....	93
5.1.1 Потoki событий.....	93
5.1.2 Марковские случайные процессы.....	95
5.1.3 Непрерывно-вероятностные модели.....	102
5.2 Имитационное моделирование процессов функционирования систем массового обслуживания.....	108
5.2.1 Формирование воздействий.....	108
5.2.2 Способы построения моделирующего алгоритма.....	110
5.2.3 Особенности имитации процесса функционирования систем...	112
5.2.4 Моделирующие алгоритмы процессов функционирования системы.....	113
5.3 Имитационное моделирование систем массового обслуживания в среде «MATLAB».....	121
5.3.1 Имитация потоков заявок и обслуживаний.....	121
5.3.2 Модели накопителей и каналов обслуживания.....	123
5.3.3 Модели типовых схем систем массового обслуживания.....	124
Вопросы для самоконтроля.....	128
Литература.....	130

Введение

Современный этап информатизации общества выдвигает в разряд приоритетных направлений развитие образования. Для практической деятельности существенное значение приобретают фундаментальные знания, необходимые для развития личности. Важна не только конструктивность приобретаемых знаний, но и умение структурировать их в соответствии с поставленными целями. В этом плане особое место занимает модельный подход, основанный на методологии системного анализа. Возможности его крайне многообразны как по используемым формальным моделям, так и по способам реализации методов моделирования. Физическое моделирование ограничено рамками простых систем.

Исследование сложных и больших систем из-за трудностей математического описания, как правило, осуществляют методами имитационного моделирования. Появление новых информационных технологий не только расширяет возможности моделирующих систем, но и позволяет широко применять многообразие уже разработанных моделей и способов их реализации.

По этим причинам машинное моделирование в настоящее время стало средством решения многих сложных задач автоматизации исследований, экспериментов и проектирования систем. Однако научиться моделированию как рабочему инструменту инженера, освоить его возможности можно только при полном овладении приемами и технологией моделирования процессов функционирования систем на ЭВМ. Именно эту цель преследует данное учебное пособие. Учебное пособие обеспечивает изучение студентами общепрофессиональной дисциплины «Компьютерное моделирование» в системе подготовки специалистов по направлению «Информатика и вычислительная техника».

Учебный материал пособия излагается в пяти разделах.

Первые три раздела посвящены основам теории моделирования систем. При этом в первом и втором разделах рассматриваются принципы построения математических и имитационных моделей, а в третьем – вопросы планирования модельных экспериментов, методы анализа и интерпретации результатов моделирования систем.

Четвертый раздел посвящен идентификации динамических систем и алгоритмизации математических моделей этих систем на основе рекуррентных соотношений, получаемых с помощью дискретного преобразования дифференциальных или разностных уравнений.

В пятом разделе рассматриваются: основные положения теории массового обслуживания; принципы построения моделирующих алгоритмов и особенности имитации процессов функционирования систем массового обслуживания; способы построения имитационных моделей систем массового обслуживания в среде «MATLAB» и методы их исследования.

1 Принципы построения математических моделей систем

1.1 Основные понятия теории моделирования систем

1.1.1 Понятие модели. Виды моделирования

Основным методом научных исследований является эксперимент с изучаемым объектом, явлением или процессом. Под экспериментом понимают научную постановку опытов и наблюдение за поведением исследуемого явления в строго учитываемых условиях. Зачастую он может быть трудновыполним, экономически не выгоден или просто невозможен, например, ввиду отсутствия исследуемого объекта. В этом случае применяют особую форму эксперимента, называемую моделированием. По своему содержанию моделирование направлено на выявление свойств изучаемого объекта, построение его модели и прогнозирование поведения исследуемого объекта. В ряде наиболее важных случаев к целям моделирования относятся:

- обоснование достоверности математического описания объекта;
- получение функциональных зависимостей между переменными модели;
- сравнение стратегий поведения сторон в конфликтных ситуациях;
- идентификация исследуемого объекта;
- оптимизация модели и выбор целевой функции;
- применение модели для обучения и тренировок.

Сущность моделирования заключается в следующем.

Построение модели начинается с изучения объекта и выдвижения гипотезы о характере свойств его на основе ограниченных сведений, некоторых догадках и предположениях. Для этого анализируются объекты-аналоги, из них выбирается прототип, наиболее близкий аналог объекта, исследование свойств которого доступно исследователю. В результате анализа прототипа создается некоторая логическая схема, позволяющая провести эксперимент и уточнить свойства объекта. Такую логическую схему называют моделью объекта. При сходстве математического описания модели и объекта, т.е. при условии их подобия, результаты исследования свойств модели можно пересчитать (перенести) на объект. Модель считается адекватной объекту, если результаты моделирования подтверждаются.

Моделированием называют способ, прием познания, позволяющий с помощью одной системы, чаще всего, искусственной воспроизвести в необходимом объеме и с требуемой точностью исследуемые стороны, свойства другой более сложной системы, являющейся объектом исследования.

Модель-это физическая или абстрактная система, воспроизводящая объект исследования и удобная для проведения экспериментов.

Удобство проведения исследований может определяться различными факторами: легкостью и доступностью получения информации, сокращением сроков и уменьшением материальных затрат на исследование и др.

Рассмотрим краткую классификацию видов моделирования систем (рис. 1.1).



Рисунок 1.1 Классификация видов моделирования систем

Различают моделирование физическое и математическое.

Физическое моделирование предполагает, что в качестве модели используется либо сама исследуемая система (например, в случае производственного эксперимента), либо другая система с той же или подобной физической природой. Обычно изготавливается макетный или опытный образец объекта, проводятся испытания, в процессе которых определяются его выходные параметры и характеристики, оцениваются надежность функционирования и степень выполнения технических требований, предъявленных к объекту. Если вариант технической разработки оказался неудачным, все повторяется сначала, то есть осуществляется повторное проектирование, изготовление опытного образца, испытания и т.д. Примером такого физического моделирования является продувка моделей самолетов в аэродинамических трубах. Понятно, что физическое моделирование сопряжено с большими временными и материальными затратами.

Под математическим моделированием понимается процесс установления соответствия данной реальной системы некоторой математической модели и исследование этой модели, позволяющее получить характеристики реальной системы.

Математическое моделирование может быть как аналитическим, так и компьютерным.

Для **аналитического моделирования** характерно то, что процесс функционирования элементов системы записывается в виде некоторых математических соотношений (алгебраических, интегральных, разностных и т.д.) или логических условий. Аналитическая модель может исследоваться:

аналитически, когда стремятся получить явные зависимости для искомым характеристик системы;

численно, когда, не умея решать уравнения, стремятся получить численные результаты, при конкретных исходных и начальных условиях;

качественно, когда, не имея решения в явном виде, можно найти некоторые свойства решения (например, оценить устойчивость решения).

Математическая модель приближенно описывает реальный процесс, явление или объект с помощью математических соотношений. Математические модели могут представлять собой системы дифференциальных уравнений (обыкновенных или в частных производных), системы алгебраических уравнений, разностные уравнения, линейные, нелинейные уравнения и т.д.

Компьютерное моделирование можно разделить на три вида: численное, имитационное, статистическое.

Для компьютерного моделирования характерно, что математическая модель системы представлена в виде программы на ЭВМ или компьютерной модели, позволяющей проводить с ней вычислительные эксперименты. **При численном моделировании** для построения компьютерной модели используются методы вычислительной математики, а вычислительный эксперимент заключается в численном решении некоторых математических уравнений при заданных значениях параметров и начальных условиях. **Имитационное моделирование** – это вид компьютерного моделирования, для которого характерно воспроизведение на ЭВМ (имитация) процесса функционирования исследуемой системы. При этом имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры, последовательности протекания во времени, что позволяет получить информацию о состоянии системы в заданные моменты времени. **Статистическое моделирование** – это вид компьютерного моделирования, позволяющий получить статистические данные о процессах в моделируемой системе.

1.1.2 Этапы компьютерного моделирования

Для компьютерного моделирования можно выделить три ступени развития. Первая ступень связана с началом появления точных наук. Поэтому некоторые методы вычислений носят имена таких корифеев наук, как Ньютон, Эйлер, а слово «алгоритм» происходит от имени средневекового арабского ученого Аль - Хорезми. Вторая ступень начинается с 50-х годов 20 века, когда появились скромные по нынешним меркам ЭВМ (компьютеры), которые избавили ученых от огромной по объему вычислительной работы. Сейчас общество подошло к третьей ступени развития, когда моделирование «встраивается» в структуры информационного общества.

Компьютерное моделирование какого-либо объекта порождает такой план действий, который можно разбить на три этапа: модель – алгоритм – программа (рис. 1.2).

На первом этапе строится «эквивалент» объекта. Этот «эквивалент» отражает в математической форме важные для данного исследования свойства объекта: законы, которым подчиняется объект, связи, присущие его частям и т.д. Затем математическая модель исследуется теоретическими методами, что позволяет получить предварительные знания об объекте.

Второй этап – разработка алгоритма для реализации модели на компьютере. Модель представляется в форме, удобной для применения численных методов, определяется последовательность вычислительных и логических операций, которые нужно произвести, чтобы найти искомые величины с заданной точностью. Вычислительные алгоритмы не должны искажать основные свойства модели и, следовательно, исходного объекта, быть экономичными и адаптирующимися к особенностям решаемых задач и используемых компьютеров.

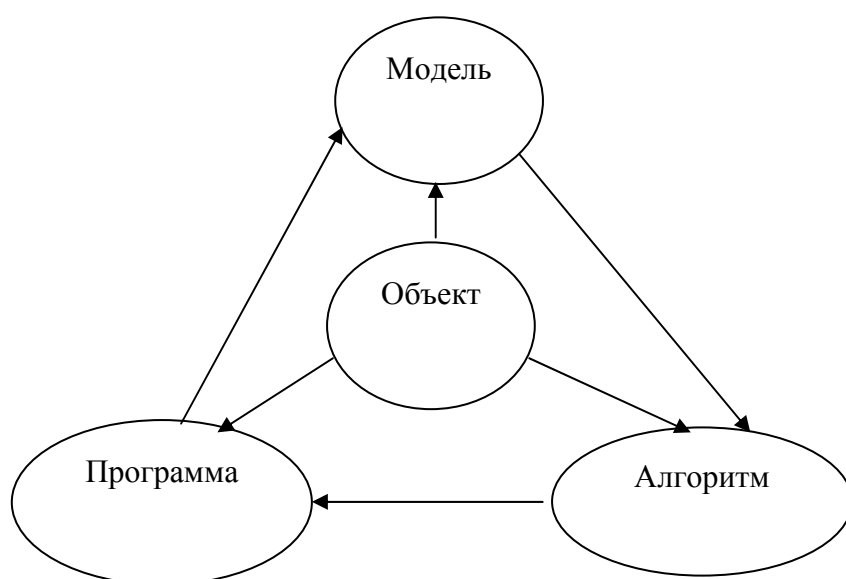


Рисунок 1.2 - Технология компьютерного моделирования

На третьем этапе создаются программы, «переводящие» модель и алгоритм на доступный компьютеру язык. К ним также предъявляются требования экономичности и адаптивности. Их можно назвать «электронным» эквивалентом изучаемого объекта, уже пригодным для испытания на компьютере.

Создав триаду «модель – алгоритм – программа», исследователь получает в руки универсальный, гибкий и недорогой инструмент, который отлаживается и тестируется в «пробных» вычислительных экспериментах. После того, как адекватность (достаточное соответствие) триады исходному объекту имеется, с моделью проводятся разные «опыты», дающие все требуемые свойства и характеристики объекта. Важно, что процесс моделирования сопровождается улучшением и уточнением всех звеньев триады.

Такой метод познания сочетает в себе преимущества и теории, и эксперимента. Действительно, работа не с самим объектом (явлением или процессом), а с компьютерной моделью дает возможность относительно быстро, без существенных затрат исследовать его свойства и поведение в любых мыслимых ситуациях. Это составляет преимущества теории. В то же время вычислительные эксперименты с моделями объектов позволяют, опираясь на мощь вычислительных методов и компьютеров, глубоко и полно изучать объекты, что недоступно чисто теоретическим подходам. Это уже составляет преимущества эксперимента.

Компьютерное моделирование как методология не подменяет собой математику, физику, биологию и другие научные дисциплины, а играет синтезирующую роль. Создание и применение триады невозможно без опоры на самые разные методы и подходы - от качественного анализа нелинейных моделей до современных языков программирования.

Но, решая проблемы информационного общества, нельзя уповать только на мощь компьютеров. Необходимо постоянное совершенствование триады математического моделирования.

1.1.3 Принципы моделирования

Практический опыт, накопленный к настоящему времени в области разработки и использования математических моделей позволяет сформулировать основные принципы моделирования, к которым относят следующие [3].

Принцип информационной достаточности. При полном отсутствии информации об исследуемой системе построение ее модели невозможно. При наличии полной информации о системе ее моделирование лишено смысла. Существует некоторый критический уровень априорных сведений о системе (уровень информационной достаточности), при достижении которого может быть построена ее адекватная модель.

Принцип осуществимости. Создаваемая модель должна обеспечивать достижение поставленной цели исследования с вероятностью, существенно отличающейся от нуля, и за конечное время. Обычно задают некоторое пороговое значение вероятности достижения цели моделирования, а также приемлемую границу времени достижения этой цели.

Принцип множественности моделей. Данный принцип является ключевым. Речь идет о том, что создаваемая модель должна отражать в первую очередь те свойства реальной системы (или явления), которые влияют на выбранный показатель эффективности. Соответственно при использовании любой конкретной модели исследуются лишь некоторые стороны реальной системы. Для более полного ее исследования необходим ряд моделей, позволяющих с разных сторон и с разной степенью детальности отражать рассматриваемую систему.

Принцип агрегирования. В большинстве случаев сложную систему можно представить состоящей из подсистем, для адекватного описания

которых оказываются пригодными некоторые стандартные математические схемы. Принцип агрегирования позволяет, кроме того, достаточно гибко перестраивать модель в зависимости от задач исследования.

Принцип параметризации. В ряде случаев моделируемая система имеет в своем составе некоторые относительно изолированные подсистемы, характеризующиеся определенным параметром, в том числе векторным. Такие подсистемы можно заменять в модели соответствующими числовыми величинами, а не описывать процесс их функционирования. При необходимости зависимость значений этих величин от ситуации может задаваться в виде таблицы, графика или аналитического выражения (формулы). Принцип параметризации позволяет сократить объем и продолжительность моделирования. Однако надо иметь в виду, что параметризация снижает адекватность модели.

1.1.4 Математическая модель

Понятие математической модели не имеет строгого формализованного определения, однако в него вкладывают конкретное содержание, связанное с применением математики. Большинство научных дисциплин по существу являются упорядоченным множеством ММ, построение которых сопровождается обоснованием адекватности отображения ими свойств исследуемых явлений и процессов.

Адекватность ММ является, как правило, большим научным достижением. Она позволяет провести детальное исследование изучаемого объекта и дать надежный прогноз его поведения в различных условиях. Но за адекватность ММ нередко приходится расплачиваться ее усложнением, что вызывает трудности при ее использовании. В этом случае на помощь математике и приходит современная вычислительная техника, существенно расширившая класс ММ, допускающих исчерпывающий количественный анализ.

Такую общность и универсальность ММ можно объяснить тем, что в математике используют абстрактные основополагающие понятия, немногочисленные, но весьма емкие по содержанию. Это позволяет конкретные факты из самых различных областей знаний рассматривать как проявление этих понятий и отношений между ними. Совокупность таких понятий и отношений, выраженных при помощи системы математических символов и обозначений и отражающих некоторые свойства изучаемого объекта, называют **математической моделью** этого объекта.

Структура математической модели

В общем случае исследуемую систему количественно можно охарактеризовать векторами x , β , y внешних, внутренних и выходных параметров соответственно.

Например, для электронного усилителя выходными параметрами являются коэффициент усиления, полоса частот пропускаемых сигналов, входное сопротивление, рассеиваемая мощность, внешними — сопротивление и емкость нагрузки, напряжения источников питания, температура окружающей среды, входные сигналы и внутренними — сопротивления резисторов, емкости конденсаторов, характеристики транзисторов.

При создании системы значения выходных параметров или диапазоны их возможного изменения оговаривают в техническом задании на разработку системы, тогда как внешние параметры характеризуют условия его функционирования.

Для простой системы ММ может представлять собой соотношение

$$y = f(x, \beta), \quad x \in R^k, \beta \in R^m, y \in R^n, \quad (1.1)$$

где f - векторная функция некоторого аргумента. Модель в виде (1.1) позволяет легко вычислять выходные параметры по задаваемым значениям внешних и внутренних параметров, т.е. решать так называемую прямую задачу анализа, так называемый поверочный расчет.

При создании системы возникает необходимость решать более сложную задачу синтеза по обусловленным техническим заданием на проектирование системы значениям внешних и выходных параметров находить его внутренние параметры. В инженерной практике решению такой задачи соответствует проектировочный расчет, часто имеющий целью оптимизацию внутренних параметров по некоторому **критерию оптимальности**. Однако при построении ММ системы функция f в (1.1) обычно заранее не известна и ее требуется установить. Эта наиболее сложная задача решается при **идентификации** ММ.

Задача идентификации может быть решена путем математической обработки информации о ряде таких состояний системы, для каждого из которых известны значения выходных, внутренних и внешних параметров. Один из таких способов связан с применением **регрессионного анализа**. Если информация о внутренних параметрах отсутствует или же внутреннее устройство системы слишком сложное, то ММ строят по принципу **черного ящика**, устанавливая соотношение между внешними и выходными параметрами путем исследования реакции системы на внешние воздействия.

Взаимосвязь выходных, внутренних и внешних параметров, отображаемая соотношением (1.1), определяет **структуру** ММ, которая в общем случае может быть выражена и в неявном виде.

Свойства математических моделей

Из сказанного ранее следует, что при изучении реально существующей или гипотетической системы математические методы применяют к ее математической модели. Это применение будет эффективным, если свойства ММ удовлетворяют определенным требованиям. Рассмотрим эти свойства.

Полнота ММ позволяет отразить в достаточной мере именно те характеристики и особенности системы, которые интересуют нас с точки зрения поставленной цели проведения вычислительного эксперимента. Например, модель может достаточно полно описывать протекающие в системе процессы, но не отражать его габаритные, массовые или стоимостные показатели.

Точность ММ дает возможность обеспечить приемлемое совпадение реальных и найденных при помощи ММ значений выходных параметров системы, составляющих вектор

$$y = (y_1, y_2, \dots, y_i, \dots, y_n)^T \in R^n$$

Пусть y_i^m и y_i^p — найденное при помощи ММ и реальное значения i -го выходного параметра. Тогда относительная погрешность ММ по отношению к этому параметру будет равна

$$\varepsilon_i = \frac{y_i^m - y_i^p}{y_i^p}, \quad i = \overline{1, n}$$

В качестве скалярной оценки вектора

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_i, \dots, \varepsilon_n)^T \in R^n$$

можно принять какую-либо его норму, например

$$\varepsilon = \sqrt{\sum_{i=1}^n \varepsilon_i^2} \quad \text{или} \quad \varepsilon = \max_{i=1, n} |\varepsilon_i|$$

Поскольку выходные параметры системы при помощи ММ связаны с его внешними и внутренними параметрами, то ε , как количественная характеристика точности модели этой системы, будет зависеть от их значений. **Адекватность ММ** — это способность ММ отражать свойства системы с относительной погрешностью не хуже заданной.

В общем смысле под адекватностью ММ понимают правильное качественное и достаточно точное количественное описание именно тех характеристик системы, которые важны в данном конкретном случае. Модель, адекватная при выборе одних характеристик, может быть неадекватной при выборе других характеристик системы. В ряде прикладных областей, еще недостаточно подготовленных к применению количественных математических методов, ММ имеют главным образом качественный характер. Эта ситуация типична, например, для биологической и социальной сфер, в которых количественные закономерности не всегда поддаются строгой математической формализации. В таких случаях под адекватностью ММ естественно понимать лишь правильное качественное описание поведения изучаемых систем.

Экономичность ММ оценивают затратами на вычислительные ресурсы (машинное время и память), необходимые для реализации ММ на ЭВМ. Эти затраты зависят от числа арифметических операций при использовании модели, от размерности пространства фазовых переменных, от особенностей применяемой ЭВМ и других факторов. Очевидно, что требования экономичности, высокой точности и достаточно широкой области адекватности ММ противоречивы и на практике могут быть удовлетворены лишь на основе разумного компромисса. Свойство экономичности ММ часто связывают с ее

простотой. Более того, количественный анализ некоторых упрощенных вариантов ММ может быть осуществлен и без привлечения современной вычислительной техники. Однако его результаты могут иметь лишь ограниченную ценность на стадии отладки алгоритма или программы, если упрощение ММ не согласовано с концептуальной моделью системы.

Робастность ММ характеризует ее устойчивость по отношению к погрешностям исходных данных, способность нивелировать эти погрешности и не допускать их чрезмерного влияния на результат вычислительного эксперимента. Причинами низкой робастности ММ могут быть необходимость при ее количественном анализе вычитания близких друг к другу приближенных значений величин или деления на малую по модулю величину, а также использование в ММ функций, быстро изменяющихся в промежутке, где значение аргумента известно с невысокой точностью.

Продуктивность ММ связана с достоверностью исходных данных. Если они являются результатом измерений, то точность их измерения должна быть выше, чем для тех параметров, которые получаются при использовании ММ. В противном случае ММ будет непродуктивной и ее применение для анализа конкретной системы потеряет смысл. Ее можно будет использовать лишь для оценки и характеристик некоторого класса систем с гипотетическими исходными данными.

Наглядность ММ является желательным, но необязательным свойством. Однако использование ММ и ее модификация упрощаются, если ее составляющие (например, отдельные члены уравнений) имеют ясный содержательный смысл. Это позволяет предвидеть результаты вычислительного эксперимента и облегчить контроль их правильности.

1.1.5 Классификация математических моделей

Математические модели различают в основном по характеру отображаемых свойств системы, степени их детализации, способам получения и формального представления.

Структурные и функциональные модели. Если ММ отображает элементы и их связи в системе, то ее называют **структурной математической моделью**. Если же ММ отражает происходящие в системе какие-либо процессы, то ее относят к **функциональным математическим моделям**. Ясно, что могут существовать и смешанные ММ, которые описывают как функциональные, так и структурные свойства системы. Структурные ММ делят на топологические и геометрические, составляющие два уровня иерархии ММ этого типа. Первые отображают состав системы и связи между его элементами. **Топологические ММ** целесообразно применять на начальной стадии исследования сложной системы. Такая ММ имеет форму графов, таблиц, матриц, списков и т.п., и ее построению обычно предшествует разработка структурной схемы системы.

Геометрическая ММ дополнительно к информации, представленной в топологической ММ, содержит сведения о форме и размерах системы и ее элементов, об их взаимном расположении. В геометрическую ММ обычно входят совокупность уравнений линий и поверхностей и алгебраические соотношения, определяющие принадлежность областей пространства системе или ее элементам. Геометрические ММ находят применение при проектировании элементов технических систем, разработке технической документации и технологических процессов изготовления изделий.

Функциональные ММ состоят из соотношений, связывающих между собой фазовые переменные, т.е. внутренние, внешние и выходные параметры системы. Функционирование сложных систем нередко удается описать лишь при помощи совокупности ее реакций на некоторые известные (или заданные) входные воздействия. Такую разновидность функциональной ММ относят к типу **черного ящика** и обычно называют **имитационной** математической моделью, имея в виду, что она лишь имитирует внешние проявления функционирования, не раскрывая и не описывая существа протекающих в системе процессов. Имитационные ММ находят широкое применение в исследовании сложных систем.

По форме представления имитационная ММ является примером **алгоритмической ММ**, поскольку связь в ней между входными и выходными параметрами системы удается описать лишь в форме алгоритма, пригодного для реализации в виде программы. К типу алгоритмических ММ относят широкий класс как функциональных, так и структурных ММ. Если связи между параметрами системы можно выразить в аналитической форме, то говорят об **аналитических** математических моделях. При создании иерархии ММ одной и той же системы обычно стремятся к тому, чтобы упрощенный вариант ММ был представлен в аналитической форме, допускающей точное решение, которое можно было бы использовать для сравнения при тестировании результатов, полученных при помощи более полных и поэтому более сложных вариантов ММ.

Ясно, что ММ конкретной системы по форме представления может включать признаки как аналитической, так и алгоритмической ММ. Более того, в процессе моделирования аналитическую ММ преобразуют в алгоритмическую.

По способу получения математические модели могут быть **теоретическими** или **эмпирическими**. Первые получают в результате изучения свойств системы, протекающих в ней процессов на основе использования известных фундаментальных законов сохранения, а также уравнений равновесия, а вторые являются итогом обработки результатов внешних наблюдений за проявлением этих свойств и процессов. Один из способов построения эмпирических ММ заключается в проведении экспериментальных исследований, связанных с измерением фазовых переменных системы, и в последующем обобщении результатов этих измерений в алгоритмической форме или в виде аналитических зависимостей. Поэтому по форме представле-

ния эмпирическая ММ может содержать признаки как алгоритмической, так и аналитической ММ. Таким образом, построение эмпирической ММ сводится к решению задачи идентификации.

Особенности функциональных моделей. Одной из характерных особенностей функциональной ММ является наличие или отсутствие среди ее параметров случайных величин. При наличии таких величин ММ называют **стохастической** (или вероятностной), а при их отсутствии - **детерминированной**.

Далеко не все параметры реальных систем можно характеризовать вполне определенными значениями. Поэтому ММ таких систем, строго говоря, следует отнести к стохастическим, поскольку выходные параметры системы будут случайными величинами. Случайными могут быть и значения внешних параметров.

Для анализа стохастических ММ необходимо использовать выводы теории вероятностей, случайных процессов и математической статистики. Однако основная трудность в их применении обычно связана с тем, что вероятностные характеристики случайных величин (математические ожидания, дисперсии, законы распределения) часто не известны или известны с не высокой точностью, т.е. ММ не удовлетворяет требованию продуктивности. В таких случаях эффективнее использовать ММ, более грубую по сравнению со стохастической, но и более устойчивую по отношению к недостоверности исходных данных.

Существенным признаком классификации ММ является их возможность описывать изменение параметров системы во времени. Если при этом в ММ отражено влияние инерционных свойств системы, то ее обычно называют **динамической**. В противоположность этому ММ, которая не учитывает изменение во времени параметров системы, называют **статической**.

Стационарные ММ описывают системы, в которых протекают так называемые установившиеся процессы, т.е. процессы, в которых интересующие нас выходные параметры постоянны во времени. К установившимся относят и периодические процессы, в которых некоторые выходные параметры остаются неизменными, а остальные претерпевают колебания.

Если выходные параметры системы изменяются медленно и в рассматриваемый фиксированный момент времени этими изменениями можно пренебречь, то считают ММ **нестационарной**.

Важным с точки зрения последующего анализа свойством ММ является ее линейность, в смысле связи параметров системы линейными соотношениями. Это означает, что при изменении какого-либо внешнего (или внутреннего) параметра системы линейная ММ предсказывает линейное изменение зависящего от него выходного параметра, а при изменении двух или более параметров — сложение их влияний, т.е. такая ММ обладает свойством **суперпозиции**. Если ММ не обладает свойством суперпозиции, то ее называют **нелинейной**.

Для количественного анализа линейных ММ разработано большое число математических методов, тогда как возможности анализа нелинейных ММ связаны в основном с методами вычислительной математики. Чтобы для исследования нелинейной ММ системы можно было использовать аналитические методы, ее обычно линеаризуют, т.е. нелинейные соотношения между параметрами заменяют приближенными линейными и получают так называемую **линеаризованную** ММ системы. Так как линеаризация связана с внесением дополнительных погрешностей, то к результатам анализа линеаризованной модели следует относиться с определенной осторожностью. Дело в том, что линеаризация ММ может привести к утрате адекватности ее. Учет в ММ нелинейных эффектов особенно важен, например, при описании смены форм движения или положений равновесия, когда малые изменения входных параметров могут вызвать качественные изменения в состоянии системы.

Каждый параметр системы может быть двух типов - непрерывно изменяющимся в некотором промежутке своих значений или принимающим только некоторые дискретные значения. Возможна и промежуточная ситуация, когда в одной области параметр принимает все возможные значения, а в другой - только дискретные. В связи с этим выделяют **непрерывные дискретные** и **смешанные** математические модели. В процессе анализа ММ этих типов могут быть преобразованы одна в другую, но при таком преобразовании следует контролировать выполнение требования **адекватности** ММ рассматриваемой системе.

Формы представления математических моделей. При математическом моделировании сложной системы описать ее поведение одной ММ, как правило, не удастся, а если такая ММ и была бы построена, то она оказалась бы слишком сложной для количественного анализа. Поэтому к таким системам обычно применяют **принцип декомпозиции**. Он состоит в условном разбиении системы на подсистемы, допускающие их независимое исследование с последующим учетом их взаимного влияния друг на друга. В свою очередь, принцип декомпозиции можно применить и к каждой выделенной подсистеме вплоть до уровня достаточно простых элементов. В таком случае возникает **иерархия** ММ связанных между собой подсистем. Иерархические уровни выделяют и для отдельных типов ММ. Например, среди структурных ММ систем к более высокому уровню иерархии относят топологические ММ, а к более низкому уровню, характеризующемуся большей детализацией, - геометрические ММ. Среди функциональных ММ иерархические уровни отражают степень детализации описания процессов, протекающих в системе и ее элементах. С этой точки зрения обычно выделяют три основных уровня: микро - макро - и мета-уровень.

Математические модели микроуровня описывают процессы в системах с распределенными параметрами, а **математические модели макроуровня** - в системах с сосредоточенными параметрами. В первых из

них фазовые переменные могут зависеть как от времени, так и от пространственных координат, а во вторых - только от времени.

Если в ММ макроуровня число фазовых переменных имеет порядок 10^4 - 10^5 , то количественный анализ такой ММ становится громоздким и требует значительных затрат вычислительных ресурсов. Кроме того, при столь большом числе фазовых переменных трудно выделить существенные характеристики системы и особенности ее поведения. В таком случае путем объединения и укрупнения элементов сложной системы стремятся уменьшить число фазовых переменных за счет исключения из рассмотрения внутренних параметров элементов, ограничиваясь, лишь описанием взаимных связей между укрупненными элементами. Такой подход характерен для ММ **метауровня**.

Наиболее распространенной формой представления **динамической (эволюционной)** ММ микроуровня является формулировка краевой задачи для дифференциальных уравнений математической физики. Такая формулировка включает дифференциальные уравнения с частными производными и краевые условия. В свою очередь краевые условия содержат начальные и граничные условия. К начальным условиям относят распределения искомым фазовых переменных в некоторый момент времени. Границы же пространственной области, конфигурация которой соответствует рассматриваемому элементу или системе в целом являются граничными условиями. При представлении ММ целесообразно использовать безразмерные переменные и коэффициенты уравнений.

ММ микроуровня называют одномерной, двумерной или трехмерной, если искомые фазовые переменные зависят от одной, двух или трех пространственных координат соответственно. Два последних типа ММ объединяют в многомерные математические модели микроуровня.

1.2 Моделирование сложных систем

1.2.1 Основные понятия и определения

В разных областях производства, научных исследований, социально-экономического прогнозирования приходится оперировать с объектами, которые называют сложными системами.

Систему называют **сложной**, если в силу свойств самой системы и по характеру решаемых задач необходимо принимать во внимание наличие в системе структурно-взаимосвязанных и взаимодействующих между собой и со средой систем минимального уровня иерархии, обеспечивающих выполнение системой сложной целевой функции.

К основным характеристикам таких объектов относят:

наличие большого числа взаимосвязанных и взаимодействующих между собой элементов системы;

целевая функция, на оптимизацию которой направлено функционирование системы, не совпадает с целевыми функциями элементов, составляющих систему;

иерархичность структуры;

наличие управления и разветвленной информационной сети, осуществляющей многочисленные связи системы как внутри ее элементной базы, так и по ее взаимодействию с внешней средой.

Примерами сложных систем являются: предприятие, энергосистема, автоматизированные системы обработки информации и управления, вычислительные системы, комплексы и сети, САПР и др.

Системой обычно считают целенаправленную совокупность взаимосвязанных и взаимодействующих между собой элементов. Под **элементом** при этом понимают некоторый объект, обладающий рядом свойств, обеспечивающих выполнение некоторых функций, а **связью** - процесс их взаимодействия, важный для целей исследования.

Части сложной системы (подсистемы) можно расчленить (часто лишь условно) на более мелкие подсистемы и т.д. вплоть до выделения элементов сложной системы, которые либо объективно не подлежат дальнейшему расчленению либо относительно их неделимости имеется договоренность. Свойства сложной системы в целом определяются как свойствами составляющих ее элементов, так и характером их взаимодействия, т.е. структурой системы и ее функционированием.

Структура системы определяется как фиксированная совокупность элементов и связей и характеризуется иерархичностью, числом уровней (подсистем) и характером взаимосвязей. Различают структуру с равноправно входящими в нее элементами и иерархическую структуру. **Иерархией** называется структура с наличием подчиненности одних элементов другим, когда воздействия в одном из направлений оказывают гораздо большее влияние на элемент, чем в другом.

Функциональные свойства системы описываются с одной стороны, характеристиками состояния и поведения ее, а с другой – эффективностью и целевой функцией.

Состоянием системы называется множество характеристик элементов системы, изменяющихся во времени и важных для целей ее функционирования.

Процессом функционирования системы назовем множество значений вектора состояний системы, изменяющихся во времени.

Целью функционирования системы называется задача получения желаемого состояния системы.

В свою очередь, определение цели влечет:

необходимость постановки локальных целей для ее элементов;

целенаправленное вмешательство в процесс функционирования системы, называемое **управлением**.

Эффективность системы – это численный показатель, характеризующий качество работы системы в заданных условиях применения. В большин-

стве случаев характеризуется средними показателями производительности, экономической эффективности, величины предотвращенного ущерба, пропускной способности и т. п.

В качестве обобщенного функционального критерия оценки качества системы через ее параметры β_i используется **целевая функция**, которая является математическим представлением цели системы. **Целевая функция** – это функция, максимум которой ищется в задачах математического программирования с учетом имеющихся ограничений.

Количественную сторону целевой функции выражают через показатель качества $\Phi(\beta_i)$, который представляет собой отношение количества удовлетворенных определенных свойств (потребностей) $U(\beta_i)$ к затратам на их удовлетворение $C(\beta_i)$ т.е.

$$\Phi(\beta_i) = U(\beta_i) / C(\beta_i).$$

Характерной особенностью исследования сложных систем является наличие трех специфических этапов, к которым относят:

построение математической модели системы;
математическое формулирование проблемы исследования (формирование целевой функции);
оптимальный анализ и синтез системы.

При реализации этих этапов попутно решаются: проблемы построения общей теории систем; проблема многокритериальности, многомерности.

Среди задач исследования сложных систем можно выделить два класса: **задачи анализа**, направленные на изучение свойств функционирования системы в зависимости от ее структуры;

задачи синтеза, связанные с выбором структуры и значений параметров по заданным свойствам системы.

Во многих практических исследованиях единственно возможным способом их решения оказывается имитационное моделирование процессов функционирования системы на ЭВМ.

1.2.2 Модель сложной системы

Для рассмотрения понятия математической модели необходимо ввести обозначения: временной интервал моделирования системы (интервал модельного времени) – $T_x = [t_0, T]$,

где t_0 - время начала моделирования, обычно полагают $t_0 = 0$, T – время окончания моделирования; $t \in T_x$ – текущее значение модельного времени.

Параметры системы $\beta_1, \beta_2, \beta_3, \dots, \beta_m$ – характеристики системы, остающиеся постоянными на всем интервале моделирования T_x .

Множество переменных системы разбивают на два подмножества – зависимых и независимых переменных.

К **независимым переменным** отнесем следующие характеристики.

Входные воздействия на систему (сигналы): X_1, X_2, \dots, X_l . Входные воздействия в момент $t \in T_X$ характеризуются вектором $X = X(t) = (X_1(t), \dots, X_l(t)) \in X \subset R^l$.

Среди $\{X_i\}$ могут быть управляющие воздействия, например, X_1, X_2, \dots, X_{l_1} ($l_1 \leq l$), а остальные $l - l_1$ воздействий – не управляющие.

Воздействия внешней среды. Среди них могут быть контролируемые (наблюдаемые) и неконтролируемые (ненаблюдаемые), детерминированные и случайные воздействия. В момент $t \in T_X$ они характеризуются вектором

$$\mathcal{G} = \mathcal{G}(t) = (\mathcal{G}_1(t), \dots, \mathcal{G}_k(t)) \in V \subset R^k.$$

В качестве примера приведем наличие дефектов у заготовок для деталей: внешние дефекты – контролируемые, а внутренние (скрытые) – неконтролируемые воздействия; случайные интервалы времени между поступлением деталей на обработку.

Переменные, характеризующие состояние системы: z_1, z_2, \dots, z_n . В отличие от $\{\beta_i\}$ состояния $\{z_i\}$ характеризуют свойства системы, изменяющиеся во времени. Состояние системы в момент $t \in T_X$ описывается вектором $z = z(t) = (z_1(t), \dots, z_n(t)) \in Z \subset R^n$, где Z – пространство состояний или фазовое пространство системы (множество возможных значений вектора z).

Если моменты изменения состояния системы $t_1 < t_2 < \dots$, то последовательность $z(t_1), z(t_2), \dots$ называется фазовой траекторией системы.

Начальное состояние системы характеризуют вектором $z^0 = z^0(t) = (z_1^0(t), \dots, z_n^0(t)) \in Z^0$, Z^0 – пространство начальных состояний.

К **зависимым переменным** относят выходные характеристики системы y_1, y_2, \dots, y_s , определяемые в момент $t \in T_X$ вектором

$$y = y(t) = (y_1(t), \dots, y_s(t)) \in Y \subset R^s.$$

Изобразим связи между зависимыми и независимыми переменными (рис. 2.3).

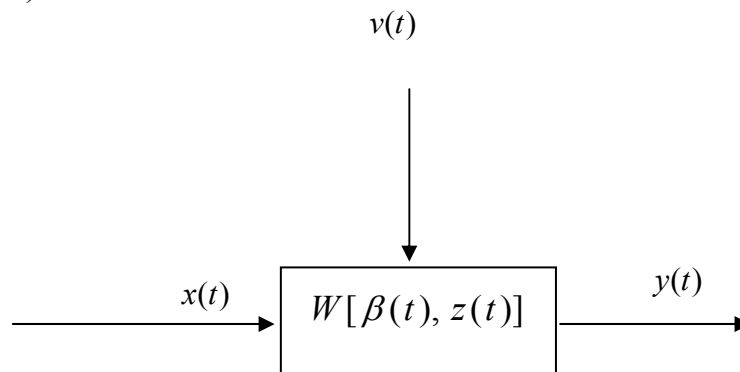


Рис. 2.3 Связь между зависимыми и независимыми переменными системы

свойства. Ограничения можно записать в виде системы неравенств (2.3), (2.4).

$$\beta_j^- \leq \beta_j \leq \beta_j^+, \quad (2.3)$$

$$z_i^- \leq z_i \leq z_i^+. \quad (2.4)$$

Формирование системы ограничений производится в результате параметрической идентификации.

1.2.4 Общий подход к формированию математических моделей

Этапы создания математической модели отражают разные степени формализации. Каждый этап представляется соответствующим описанием.

По степени формализации различают:

содержательное описание системы;

формализованное описание системы;

информационно-программное описание системы.

Содержательное описание (концептуальная модель) в словесной форме концентрирует сведения:

о физической природе и количественных характеристиках элементарных явлений исследуемой системы;

о степени и характере взаимодействия между ними;

о месте и значении каждого из элементарных явлений в общем функционировании системы.

Формализованное описание представляет материал в формальном виде. Для построения формализованной схемы необходимо:

установить систему параметров, определяющих процесс и дать их количественную характеристику;

установить взаимосвязь между характеристиками и параметрами с учетом факторов, принимаемых во внимание при формализации;

подготовить исходный материал для окончательной формализации;

определить частные целевые функции подсистем и завершить подготовку данных для окончательного определения целевой функции системы.

Информационно-программное описание включает математическое моделирование в полном объеме и реализацию ее на ЭВМ с использованием аппаратных и инструментальных средств.

Этапы математического моделирования

Создание математических моделей осуществляют в три этапа, которыми являются:

определение исходных множеств;

структурная идентификация;

параметрическая идентификация.

При **определении исходных множеств** осуществляется выбор параметров и состояний системы, которые оказывают существенное влияние на значение первоначально выбранной целевой функции.

В общем случае процесс выделения параметров и состояний задача не формализуемая. В каждом конкретном случае выделение этих множеств является процессом творческой деятельности исследователя. Может использоваться аппарат экспертных оценок.

Чтобы определить эти множества необходимо иметь исходный материал в виде характеристики жизненного цикла системы, который отражает ее период существования (проектирование, производство и эксплуатация).

Жизненный цикл обычно рассматривается на этапах создания и функционирования.

Структурная идентификация. На этом этапе обеспечивается структурная адекватность модели исследуемой системы к задачам системного исследования. Структурная идентификация включает:

определение входов и выходов модели;

декомпозицию модели;

оптимальное выделение в ней отдельных структурных элементов путем декомпозиции и агрегирования системы и ее элементов;

выбор структуры элементов модели, то есть математическая формализация структурных элементов и их взаимосвязей.

Каждая из этих задач структурной идентификации выполняется по определенным правилам. В результате структурной идентификации создается математическая модель (2.1), (2.2).

Параметрическая идентификация позволяет определить реальные численные значения параметров системы и пределы их изменений, то есть дополнить модель зависимостями (2.3), (2.4).

Вследствие сложности параметрической идентификации базой для ее проведения часто выбирают теорию планирования эксперимента, при этом определяются параметры, которые нельзя идентифицировать без эксперимента и осуществляется корректировка структуры модели системы.

Основные правила построения математических моделей

Корректное и обоснованное представление требований к целевой функции системы на основании жесткого отбора интересующих исследователя характеристик системы. Любое отклонение и не выполнение этого требования приводит к ухудшению качества системы и к необоснованному завышению стоимости.

Представление системы в виде совокупности элементарных подсистем и описание их математическими зависимостями, причем наиболее полно представляется описание интересующей исследователя подсистемы.

Математический аппарат, выбранный для описания системы в целом и ее подсистем должен обеспечить требуемую адекватность модели системе и достаточную простоту модели в целом.

Предельное упрощение системы за счет исключения из нее второстепенных, мало влияющих параметров.

Результат функционирования модели не должен выходить за пределы, определяемыми предоставленными системе ресурсами.

Эффективность математических методов исследования значительно возрастает при переходе от детерминированных моделей к вероятностным.

1.3 Типовые математические схемы моделирования

Модель сложной системы, рассмотренная ранее, представляет собой математическую схему моделирования общего вида. На практике для формализации концептуальных моделей ряда систем выгоднее применять типовые математические схемы моделирования, учитывающие с одной стороны способ представления времени в модели (непрерывная переменная или дискретная), а с другой стороны степень случайности моделируемых процессов. По этим признакам различают следующие математические схемы моделирования (классы ММ).

Непрерывно – детерминированные модели (D – схемы).

Дискретно – детерминированные модели (F – схемы).

Дискретно – вероятностные модели (P – схемы).

Непрерывно - вероятностные модели (Q – схемы).

Сетевые модели (N – схемы).

Агрегатные модели (A – схемы).

Непрерывно-детерминированные модели. В этих моделях время t полагается непрерывной переменной, а случайными факторами в системе пренебрегают. Математический аппарат моделей – теория дифференциальных и интегральных уравнений, с помощью которой достигается адекватное описание динамических систем. Наиболее глубоко разработан операторный метод описания и исследования процессов функционирования динамических систем и их структур.

Примером непрерывно – детерминированной модели одноканальной системы автоматического управления является неоднородное дифференциальное уравнение с постоянными коэффициентами.

$$a_n y^{(n)}(t) + a_{n-1} y^{(n-1)}(t) + \dots + a_0 y(t) = b_m x^{(m)}(t) + b_{m-1} x^{(m-1)}(t) + \dots + b_0 x(t). \quad (2.5)$$

В этом уравнении $x(t)$ - входное воздействие; $y(t)$ – выходная величина, характеризующая положение объекта управления; a_i, b_i - внутренние параметры системы.

Если динамическая система описывается нелинейным дифференциальным уравнением, то его линеаризуют и решают как линейное.

Применение непрерывно – детерминированных моделей позволяет количественно осуществлять не только анализ динамических систем, но и оптимальный синтез их.

Дискретно-детерминированные модели. В дискретно-детерминированных (ДД) моделях время t является дискретной переменной

$t = \tau * \Delta t$, где Δt – шаг дискретизации, а $\tau = 0, 1, 2, \dots$ – дискретные моменты времени.

Основной математический аппарат, используемый при построении ДД – моделей – это теория разностных уравнений и аппарат дискретной математики, в частности, теория конечных автоматов.

Разностное уравнение – это уравнение, содержащее конечные разности искомой функции

$$\Phi(x_\tau, x_{\tau+1}, \dots, x_{\tau+n}, u_\tau, u_{\tau+1}, \dots, u_{\tau+m}, \Theta, \tau) = 0, \quad (2.6)$$

где $x_\tau = x(\tau * \Delta t), u_\tau = u(\tau * \Delta t)$ – соответственно состояние системы и внешнее воздействие в дискретные моменты времени τ .

В прикладных задачах ДД – модели в виде (2.6) часто возникают как промежуточные при исследовании НД – моделей на ЭВМ, когда аналитическое решение дифференциального уравнения получить не удается и приходится применять разностные схемы.

Кратко рассмотрим теорию конечных автоматов, которая используется для построения ДД – моделей.

Конечный автомат – это математическая модель дискретной системы, которая под действием входных сигналов $u \in U$ вырабатывает выходные сигналы $y \in Y$, и которая может иметь некоторые изменяемые внутренние состояния $x \in X$; здесь U, Y, X – конечные множества.

Конечный автомат характеризуется: входным алфавитом U ; выходным алфавитом Y ; внутренним алфавитом состояний X ; начальным состоянием $x_0 \in X$; функцией переходов $x' = \Phi(x, u) : XU \rightarrow X$; функцией выходов $y = \Psi(x, u) : XU \rightarrow Y$.

Процесс функционирования конечного автомата таков. В τ -м такте ($\tau = 0, 1, \dots$) на вход автомата, находящегося в состоянии $x_\tau \in X$, поступает входной сигнал $u_\tau \in U$, на который автомат реагирует переходом на $\tau+1$ -м такте в состояние $x_{\tau+1} \in X$ и выдачей выходного сигнала $y_\tau \in Y$. Например, конечный автомат Мили описывается следующими рекуррентными соотношениями:

$$\begin{aligned} x_{\tau+1} &= \Phi(x_\tau, u_\tau), x(0) = x_0, \\ y_\tau &= \Psi(x_\tau, u_\tau), \tau = 0, 1, \dots \end{aligned} \quad (2.7)$$

Дискретно-вероятностные модели. В дискретно-вероятностной модели учитываются случайные элементы исследуемой сложной системы. Основным математическим аппарат, используемый при построении и исследовании ДВ – моделей, – это теория разностных стохастических уравнений и теория вероятностных автоматов.

Разностное стохастическое уравнение – это такое уравнение, которое содержит случайные параметры Θ или случайные входные воздействия $\{u_\tau\}$.

Пусть на вероятностном пространстве (Ω, F, P) определен случайный M – вектор параметров $\Theta = (\Theta_0, \dots, \Theta_{M-1}) \in R^M$ и случайная последовательность входных воздействий $u_0, u_1, u_2, \dots \in U$.

Нелинейное разностное стохастическое уравнение порядка (n, m) имеет вид $\Phi(x_\tau, x_{\tau-1}, \dots, x_{\tau-n}, u_\tau, u_{\tau-1}, \dots, u_{\tau-m}, \Theta, \tau) = 0$, (2.8)

где $\tau = n, n+1, \dots; x_0, x_1, \dots, x_{n-1}$ – заданные начальные состояния системы; $\Phi(\cdot)$ – заданная функция $n + m + M + 3$ переменных.

Решением этого уравнения является определенная на множестве (Ω, F, P) случайная последовательность состояний моделируемой системы:

$$x_{n+2}, x_{n+3}, \dots \in X.$$

Если функция $\Phi(\cdot)$ линейная по $\{x_\tau, u_\tau\}$, то (2.8) примет вид:

$$\sum_{i=0}^n \Theta_i x_{\tau-i} + \sum_{j=0}^m \Theta_{j+n+1} x_{\tau-j} = u_\tau, \quad \tau = n+1, n+2, \dots, \quad (2.9)$$

где $\Theta = (\Theta_0, \dots, \Theta_{n+m+1})$ – вектор $M = n + m + 2$ параметров.

Другой математический аппарат построения ДВ – моделей сложных систем представляет теория вероятностных автоматов.

Вероятностный автомат, определенный на множестве (Ω, F, P) , есть конечный автомат, в котором функция переходов $x' = \Phi(x, u, w)$, $w \in \Omega$ и функция выходов $y = \Psi(x, u, w)$ являются случайными функциями, имеющими некоторые вероятностные распределения.

Примем обозначения для вероятностных распределений $\Phi(\cdot), \Psi(\cdot)$: $\Pi_{(i,k)} = P\{x_0 = k | u_0 = i\}$, $\Pi = (\Pi_{(i,k)})$ – начальное распределение вероятностей, $i \in S(I) = \{0, 1, \dots, I\}$, $k \in S(K)$; $P = (p_{(j,l)|(i,k)})$, $p_{(j,l)|(i,k)} = P\{x_{\tau+1} = l, y_\tau = j | x_\tau = k, u_\tau = i\}$ ($j \in S(\tau), l \in S(k)$) – вероятность события, состоящего в том, что находящийся в τ –м такте в состоянии $x_\tau = k$ автомат под воздействием входного сигнала $u_\tau = i$ выдаст выходной сигнал $y_\tau = j$ и перейдет на $\tau + 1$ –м такте в состояние $x_{\tau+1} = l$;

$$a_{l|(i,k)} = P\{x_{\tau+1} = l | x_\tau = k, u_\tau = i\},$$

$$b_{j|(i,k)} = P\{y_\tau = j | x_\tau = k, u_\tau = i\}.$$

Математическая модель вероятностного автомата полностью определяется пятью элементами: U, Y, X, P, Π .

Непрерывно – вероятностные модели. При построении и исследовании НВ – моделей используется теория стохастических дифференциальных уравнений и теория массового обслуживания.

Стохастическое дифференциальное уравнение (в форме Ито) имеет вид:

$$dx(t) = a(t, x(t))dt + b(t, x(t))dw(t),$$

где $x(t)$ – случайный процесс, определяющий состояние системы в момент времени t ; $w(t)$ – стандартный винеровский случайный процесс; $b(\cdot), a(\cdot)$ – коэффициенты диффузии и переноса. НВ – модель часто используется при моделировании стохастических систем управления, процессов обмена.

Теория массового обслуживания разрабатывает и исследует математические модели различных по своей природе процессов функционирования систем, например: поставок сырья и комплектующих изделий некоторому предприятию; заданий, поступающих на ЭВМ от удаленных терминалов; вызов на телефонных станциях и т.д. Для функционирования таких систем характерна стохастичность: случайность моментов времени появления заявок на обслуживание и т.д.

Система, описываемая как система массового обслуживания (СМО), состоит из $L \geq 1$ приборов обслуживания Π_1, \dots, Π_L . Прибор обслуживания Π_i состоит из накопителя заявок H_i , в котором могут одновременно находиться l_i заявок ($0 \leq l_i \leq m_i$), и канала K_i обслуживания заявок; m_i ($0 \leq m_i \leq \infty$) – емкость накопителя H_i , то есть число мест в очереди на обслуживание заявок в канале K_i ($i=1, L$).

На каждый элемент прибора Π_i поступают потоки событий; в накопитель H_i – поток заявок $\{g_i\}$, на канал K_i – поток «обслуживаний» $\{u_i\}$. Поток заявок $\{g_i\}$ представляет последовательность интервалов времени между моментами появления заявок на входе СМО и образует подмножество неуправляемых переменных СМО. А поток $\{u_i\}$ представляет собой последовательность интервалов времени между моментами начала и окончания обслуживания заявок и образует подмножество управляемых переменных.

Заявки, обслуженные СМО, образуют выходной поток $\{y_i\}$ – последовательность интервалов времени между моментами выхода заявок. Не обслуженные заявки, но покинувшие СМО по различным причинам, образуют выходной поток потерянных заявок.

Сетевые модели используют для формализации причинно – следственных связей в сложных системах с параллельными процессами. В основе этих моделей лежит сеть Петри. При графической интерпретации сеть Петри представляет собой граф особого вида, состоящий из вершин двух типов – *позиций* и *переходов*, соединенных ориентированными дугами, причем каждая дуга может связывать лишь разнотипные вершины (позицию с переходом или переход с позицией). Вершины-позиции обозначаются кружками, вершины-переходы – черточками. С содержательной точки зрения переходы соответствуют событиям, присущим исследуемой системе, а позиции – условиям их возникновения.

Таким образом, совокупность переходов, позиций и дуг позволяет описать причинно-следственные связи, присущие системе, но в статике. Чтобы сеть Петри «оживила», вводят еще один вид объектов сети – так называемые *фишки* или *метки* позиций, которые перемещаются по переходам сети при условии наличия метки во входной позиции и отсутствии метки в выходной позиции. Расположение фишек в позициях сети называется *разметкой сети*.

Агрегатные модели. Анализ существующих задач приводит к выводу о том, что комплексное решение проблем возможно лишь в том случае, если моделирующие системы имеют в своей основе единую математическую схе-

му моделирования. Такой подход к формализации процесса функционирования сложной системы предложен Бусленко Н.П. [1] и базируется на понятии «агрегата».

При агрегатном описании сложная система разбивается по подсистемы, сохраняя при этом связи обеспечивающие взаимодействие их. Если подсистема оказывается сложной, то процесс расчленения продолжается до тех пор, пока не образуются подсистемы, которые в условиях рассматриваемой задачи могут считаться удобными для математического описания.

В результате этого получается многоуровневая конструкция из взаимосвязанных элементов объединенных в подсистемы различных уровней. Элементами агрегатной модели являются агрегаты. Связи между агрегатами и внешней средой осуществляются с помощью операторов сопряжения. Сам агрегат тоже может рассматриваться как агрегатная модель, то есть разбиваться на элементы следующего уровня.

Любой агрегат характеризуется множествами: моментов времени T , входных X и выходных Y сигналов, состояний агрегата Z в каждый момент времени t . Процесс функционирования агрегата состоит из скачков состояний δz в моменты поступлений входных сигналов x и изменений состояний между этими моментами t_n и t_{n+1} .

Моменты скачков δz , не являющиеся моментами поступления входных сигналов называют особыми моментами времени t_δ , а состояния $z(t_\delta)$ – особыми состояниями агрегатной схемы. В множестве состояний Z выделяют подмножество $Z^{(y)}$, что если $z(t_\delta)$ достигает $Z^{(y)}$, то это состояние является моментом выдачи выходного сигнала y .

Вопросы для самопроверки

1. Что понимают под моделированием системы?
2. Какая из целей моделирования достигается в результате построения модели?
3. Определите основные отличия имитационного моделирования от аналитического.
4. Назовите виды математических моделей, выделяемые по характеру отображаемых свойств объекта, и дайте им характеристику.
5. Сформулируйте определение сложной системы и укажите ее отличия от обычной системы.
6. Какой вид получает математическая модель, если осуществлена структурная идентификация системы?
7. Объясните, почему отыскание ограничений на параметры и характеристики состояния системы осуществляют в ходе эксперимента?
8. Назовите операции, выполняемые при структурной идентификации.
9. Какую роль играет целевая функция в моделирование сложных систем?
10. Какие математические схемы моделирования часто применяют при исследовании вычислительных систем?

2 Принципы имитационного моделирования систем

2.1 Статистическое моделирование систем

2.1.1. Характеристика методов моделирования вероятностных объектов

Вероятностным объектом называют систему, находящуюся под воздействием случайных внешних или внутренних факторов: возмущений и помех. Они могут быть как непрерывными величинами или функциями, так и дискретными. При анализе функционирования таких систем возникают различные вероятностные задачи, например, определение математического ожидания, дисперсии, корреляционной функции, спектральной плотности реакции системы по математическому ожиданию, дисперсии, корреляционной функции, спектральной плотности воздействия на систему.

Аналитическое решение такой задачи возможно лишь в отдельных случаях, чаще приходится прибегать к имитационному моделированию на ЭВМ. При этом возможны два метода решения этих задач: детерминированный и статистический. **Детерминированный метод** предполагает моделирование аналитических соотношений между вероятностными характеристиками входного воздействия и реакции системы. Задача решается путем детерминированного преобразования вероятностной характеристики воздействия в одноименную вероятностную характеристику реакции. Метод имеет ограниченное применение.

Статистический метод предполагает моделирование системы, находящейся под влиянием случайных факторов. Задача решается путем статистической обработки результатов исследования модели. Метод является универсальным, поскольку применим и к детерминированным задачам. В этом случае производится замена детерминированной задачи эквивалентной схемой некоторой вероятностной системы, выходные характеристики которой совпадают с решением детерминированной задачи. В результате точное решение задачи заменяется приближенным. Однако с ростом числа испытаний погрешности оценок уменьшаются. При достаточно большом числе испытаний полученные результаты приобретают статистическую устойчивость и с определенной точностью могут быть приняты в качестве оценок неизвестных характеристик системы.

Накопление значительного количества данных о выходных характеристиках вероятностной системы осуществляют в ходе статистического эксперимента с моделью. Под статистическим экспериментом обычно понимают постановку опытов с моделью системы и наблюдение за поведением модели при случайных воздействиях. Для его осуществления имитационная модель должна отражать логику функционирования системы во времени и обеспечивать возможность многократного повторения опытов с одновременным накоплением и статистической обработкой данных.

В основе статистического эксперимента лежит метод статистических испытаний (метод Монте-Карло). Суть его состоит в том, что результат испытания зависит от значения некоторой случайной величины, распределенной по заданному закону. Поэтому результат каждого отдельного испытания также носит случайный характер. Проведя серию испытаний, получают множество частных значений наблюдаемой характеристики (выборку). Полученные статистические данные обрабатываются и представляются в виде численных оценок интересующих исследователя величин (характеристик системы).

Теоретической основой метода статистических испытаний являются предельные теоремы теории вероятностей (теорема Чебышева, теорема Бернулли, теорема Пуассона). Принципиальное значение предельных теорем состоит в том, что они гарантируют высокое качество статистических оценок при весьма большом числе испытаний.

В качестве примера рассмотрим применение метода статистических испытаний для вычисления площади круга заданного радиуса[3]. Данная задача явно относится к классу детерминированных задач. Пусть круг имеет радиус $r=5$, и его центр находится в точке с координатами $(1,2)$. Уравнение соответствующей окружности имеет вид:

$$(x - 1)^2 + (y - 2)^2 = 25 .$$

Для решения задачи методом Монте-Карло впишем круг в квадрат. Его вершины будут иметь координаты $(-4,-3)$, $(-4,7)$, $(6,7)$. Любая точка внутри квадрата или на его границе должна удовлетворять неравенствам $-4 < x < 6$ и $-3 < y < 7$. При решении задачи естественно исходить из того, что все точки в этом квадрате могут появляться с одинаковой вероятностью, то есть x и y распределены равномерно со следующими плотностями вероятности:

$$f(x) = 1/10 \text{ для } -4 \leq x \leq 6 \text{ и } f(x) = 0 \text{ – в противном случае;}$$

$$f(y) = 1/10 \text{ для } -3 \leq y \leq 7 \text{ и } f(y) = 0 \text{ – в противном случае.}$$

Проведя некоторое количество испытаний (то есть, получив множество случайных точек, принадлежащих квадрату), подсчитаем число точек, попавших внутрь круга или на окружность. Если выборка состоит из n наблюдений и m точек попали внутрь круга, то оценку площади круга можно получить из соотношения:

$$S_{KP} = S_{KB} * m / n.$$

В табл.2.1. приведены оценки $S_{кр}$, полученные для разных значений n , причем для каждого n выполнялось 5 прогонов (точное значение $S_{кр} = 78,54$ см). Прогоны отличаются друг от друга последовательностями случайных чисел, из которых формировались координаты точек (рис.3.1).

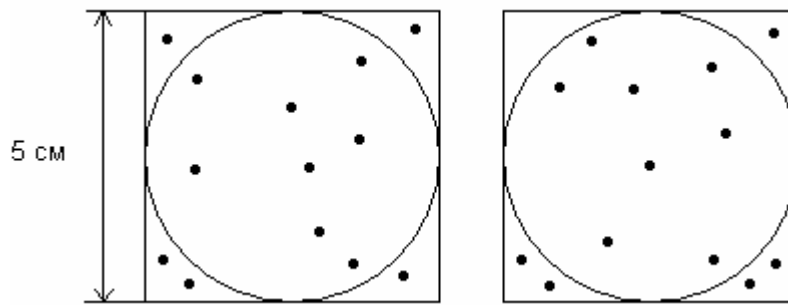


Рисунок 2.1 - Различные результаты двух прогонов при неизменном количестве точек ($n=13$). Слева показан результат первого прогона, справа – второго.

На основании полученных результатов могут быть сделаны выводы, которые справедливы для любого имитационного эксперимента независимо от физической природы и типа моделируемой системы:

каждый прогон модели можно рассматривать как одно наблюдение в проводимом эксперименте на модели;

Таблица 2.1 - Результаты статистических испытаний

Номер прогона	Оценки площади круга $S(\text{кр})$				
	Объем испытаний (n)				
	100	200	1000	5000	10000
1	78	79,5	78,8	78,22	78,77
2	70	77,0	77,3	78,6	78,23
3	81	79,5	80,2	77,72	78,88
4	70	77,0	79,5	77,76	78,63
5	79	77,0	79,8	79,0	78,21
Среднее	75,6	78,0	79,12	78,26	78,544
Дисперсия	27,3	1,88	1,29	0,3	0,1

с увеличением продолжительности прогона (то есть наблюдения или объема испытаний) отклонение измеряемой величины от ее точного значения уменьшается, поскольку наблюдаемая система переходит в стационарное состояние;

влияние переходных условий можно уменьшить, если увеличить количество прогонов модели (то есть количество экспериментов);

существует предел, за которым увеличение продолжительности прогона модели уже не дает существенного повышения точности результата, измеряемой дисперсией.

Из рассмотренного примера следует, что имитационное моделирование не ограничивается разработкой модели и написанием соответствующей программы, а требует подготовки и проведения статистического эксперимента. В связи с этим результаты имитационного моделирования следует рассматривать как экспериментальные данные, требующие специальной обработки и анализа. В частности, для любого модельного эксперимента необходимо ответить на перечисленные ниже вопросы.

Какова должна быть продолжительность прогона для достижения стационарных условий?

Как получить статистически независимые наблюдения?

Сколько наблюдений необходимо для обеспечения требуемой точности?

Ответы на эти вопросы будут даны в последующих разделах.

2.1.2. Формирование базовой последовательности случайных чисел

Поскольку имитационная модель позволяет исследовать поведение различных систем с учетом влияния случайных факторов, то эти факторы, в зависимости от их природы, могут быть отражены в модели как случайные события, случайные величины (дискретные или непрерывные) или как случайные функции (процессы).

Например, если с помощью создаваемой имитационной модели предполагается исследовать надежность вычислительной системы, то возникновение отказа будет представлено в модели как случайное событие. Если же модель предназначена для оценки временных параметров процесса обслуживания клиентов в банке, то интервал времени до появления очередного клиента удобнее всего описать как случайную величину, распределенную по некоторому закону.

Методы генерации случайных чисел. В основе всех методов и приемов моделирования случайных факторов лежит использование случайных чисел имеющих равномерное распределение на интервале (0,1).

«Истинно» случайные числа формируются с помощью аналого-цифровых преобразователей на основе сигналов физических генераторов, использующих естественные источники случайных шумов (радиоактивный распад, шумы электронных и полупроводниковых устройств и т.п.).

Случайные числа, генерируемые аппаратно или программно на ЭВМ,

3 Эксперимент с моделью. Обработка результатов моделирования

3.1 Планирование экспериментов

3.1.1 Методы теории планирования

После разработки и отладки ИМ переходят к экспериментальным исследованиям свойств системы. Основная проблема, возникающая при этом, заключается в том, как за конечное время получить достоверную и точную информацию о свойствах системы. Решение этой проблемы осуществляют планированием эксперимента. Попутно решается ряд частных задач, важнейшими из которых являются:

- снижение затрат машинного времени;
- обеспечение точности и достоверности получаемых результатов;
- проверка адекватности модели исследуемой системе.

В ходе планирования разрабатывают план эксперимента, который определяет объем и порядок проведения вычислений на ЭВМ, а также приемы накопления и статистической обработки результатов.

Рассмотрим процесс планирования эксперимента применительно к такой цели моделирования как отыскание функциональной зависимости между переменными модели. В зависимости от своей роли каждая переменная может быть либо фактором, либо реакцией. Если изучается влияние переменной x на переменную y , то x – фактор, а y – реакция.

Каждый фактор $x_i, i = 1, 2, \dots, k$ может принимать в эксперименте несколько значений – уровней. Каждый из факторов имеет верхний и нижний уровни, расположенные симметрично относительно некоторого нулевого уровня. Точка в факторном пространстве, соответствующая нулевым уровням всех факторов, называется центром плана. Интервалом варьирования фактора называется некоторое число J , прибавление которого к нулевому уровню дает верхний уровень, а вычитание – нижний.

Фиксированный набор уровней факторов в одном из опытов эксперимента представляет собой лишь одну комбинацию уровней факторов, которая определяет одно из возможных состояний системы. Каждому фиксированному набору уровней факторов соответствует точка в многомерном пространстве, называемом **факторным пространством**.

Из-за временных ограничений эксперимент выполняют не во всех точках факторного пространства, а лишь в точках допустимой области. Для этого находят минимальное число испытаний (число точек факторного пространства), которое обеспечивает выявление функциональной зависимости $y = f(x_1, x_2, \dots, x_k)$ с требуемой точностью и достоверностью.

В первую очередь при планировании эксперимента определяют основные свойства факторов, поскольку они могут быть управляемые и неуправляемые, наблюдаемые и ненаблюдаемые, количественные и

качественные, фиксированные и случайные. Из них отбирают те, которые удовлетворяют требованию управляемости и оказывают непосредственное воздействие на объект исследования. При изменении нескольких факторов по уровням важна их совместимость (осуществимость их комбинаций) и независимость (возможность установления факторов на любом уровне независимо от уровней других факторов).

Кроме того, факторы могут влиять друг на друга, поэтому при проведении машинного эксперимента необходимо создать условия, способствующие выявлению такого влияния. Таким образом, для планирования эксперимента необходимо:

1. Выделить факторы, влияющие на искомые характеристики и на их основе описать исследуемую функциональную зависимость;
2. Установить диапазон изменения выделенных факторов ($x_{i\min}, \dots, x_{i\max}$);
3. Определить координаты точек факторного пространства, в которых необходимо провести эксперимент;
4. Оценить необходимое число реализаций и их порядок в эксперименте.

Решение первых трех задач, позволяющее отыскать фиксированное число опытов для получения наиболее достоверного значения функции $y = f(x_1, x_2, \dots, x_k)$, называют **стратегическим планированием**. Решение четвертой задачи называют **тактическим планированием**, поскольку требуется определить минимальное число испытаний, при котором статистическая оценка искомой функции может быть получена с заданной точностью.

3.1.2 Стратегическое планирование эксперимента

Отыскание комбинаций уровней внешних и внутренних факторов, при которых может быть получена наиболее полная и достоверная информация о поведении системы составляет ядро данного вида планирования.

Для этого должны быть решены две основные задачи:

- идентификация факторов;
- выбор уровней факторов.

Из опыта известно, что для большинства систем 20% факторов определяют 80% свойств системы, а остальные 80% факторов определяют лишь 20% свойств системы [1].

Поэтому при идентификации факторов производится их ранжирование по степени влияния на значение наблюдаемой переменной (показателя эффективности).

По итогам идентификации все факторы делят на две группы – первичные и вторичные. *Первичные* – это те факторы, в исследовании влияния которых экспериментатор заинтересован непосредственно. *Вторичные* факторы не являются предметом исследования, но влиянием их нельзя пренебречь.

Выбор уровней факторов главным образом связан с определением интервалов между уровнями каждого из факторов [7]. Критерием выбора величины интервалов выступает и относительная точность данных на различных участках области исследуемых значений, и характер экспериментальной функции. Кроме того, при определении числа уровней факторов учитывают возможный диапазон изменения факторов и объем моделирования, который не должен быть чрезмерным.

Способы построения стратегического плана. Эксперимент, в котором реализуются все возможные комбинации уровней факторов, называется *полным факторным экспериментом* (ПФЭ).

Общее число N различных комбинаций уровней в ПФЭ для k факторов можно вычислить как:

$$N = l_1 \cdot l_2 \cdot \dots \cdot l_i \cdot \dots \cdot l_k,$$

где l_i – число уровней i -го фактора.

Если число уровней для всех факторов одинаково, то $N = L^k$ (L – число уровней).

Недостаток ПФЭ – большие временные затраты на подготовку и проведение эксперимента. Например, если в модели отражены 3 фактора, влияющие на значение выбранного показателя эффективности, каждый из которых имеет 4 возможных уровня (значения), то план проведения ПФЭ будет включать 64 эксперимента. Если при этом каждый из них длится хотя бы одну минуту (с учетом времени на изменение значений факторов), то на однократную реализацию ПФЭ потребуется более часа.

Поэтому использование ПФЭ целесообразно только в том случае, если в ходе имитационного эксперимента исследуется взаимное влияние всех факторов, фигурирующих в модели.

Если взаимное влияние факторов отсутствует или их эффектом можно пренебречь, то проводят частичный факторный эксперимент (ЧФЭ). К ним относятся [3]:

- рандомизированный план;
- латинский план;
- классический план;
- дробный факторный эксперимент.

Рандомизированный план – предполагает выбор сочетаний уровней для каждого прогона случайным образом. При использовании этого метода отправной точкой в формировании плана является число экспериментов, которые считает возможным (или необходимым) провести исследователь.

Латинский план (или «латинский квадрат») – используется в том случае, когда проводится эксперимент с одним первичным фактором и несколькими вторичными. Суть такого планирования состоит в следующем. Если первичный фактор A имеет I уровней, то для каждого вторичного фактора также выбирается I уровней. Выбор комбинации уровней факторов

выполняется на основе специальной процедуры, которую мы рассмотрим на примере.

Пусть в эксперименте используется первичный фактор А и два вторичных фактора – В и С, число уровней факторов 1 равно 4. Соответствующий план можно представить в виде квадратной матрицы размером $l \times l$ (4×4) относительно уровней фактора А. при этом матрица строится таким образом, чтобы в каждой строке и в каждом столбце данный уровень фактора А встречался только один раз:

Значение фактора В	Значение фактора С			
	С1	С2	С3	С4
В1	А1	А2	А3	А4
В2	А2	А3	А4	А1
В3	А3	А4	А1	А2
В4	А4	А1	А2	А3

В результате имеет место план, требующий $4 \times 4 = 16$ прогонов, в отличие от ПФЭ, для которого нужно $4^3 = 64$ прогона.

Классический план – это эксперимент с изменением факторов по одному. Суть его состоит в том, что один из факторов «пробегают» все l уровней, а остальные $k-1$ факторов поддерживаются постоянными. Такой план обеспечивает исследование эффектов каждого фактора в отдельности. Он требует всего $N = l_1 + l_2 + \dots + l_k$ прогонов (l_i – число уровней i -го фактора).

Для рассмотренного выше примера (3 фактора, по 4 уровня каждый) $N = 4 + 4 + 4 = 12$. Еще раз подчеркнем, что такой план применим (как и любой ЧФЭ) только при отсутствии взаимодействия между факторами.

Дробный факторный эксперимент. Каждый фактор имеет два уровня – нижний и верхний, поэтому общее число вариантов эксперимента $N = 2^k$, где k – число факторов. Матрица плана для $k=2$ приведена ниже.

Матрица плана для $k=2$

Номер эксперимента	Значение факторов	
	X_1	X_2
1	0	0
2	0	1
3	1	0
4	1	1

Планы, построенные по такому принципу, обладают свойствами: симметричности, нормированности и ортогональности, что обеспечивает повышение качества проводимых экспериментов.

3.1.3 Обеспечение точности и достоверности результатов моделирования

Тактическое планирование эксперимента с машинной моделью системы связано с вопросами эффективного использования выделенных для эксперимента машинных ресурсов и определением конкретных способов проведения испытаний модели, намеченных планом эксперимента, построенным при стратегическом планировании. Тактическое планирование машинного эксперимента связано, прежде всего, с решением следующих проблем [2]:

- определение начальных условий и их влияния на достижение установившегося результата при моделировании;
- обеспечение точности и достоверности результатов моделирования;
- уменьшение дисперсии оценок искомых характеристик;
- выбор правил автоматической остановки имитационного эксперимента с моделями систем.

Первая проблема при проведении машинного эксперимента возникает вследствие искусственного характера процесса функционирования модели, которая в отличие от реальной системы работает эпизодически, т. е. только когда экспериментатор запускает машинную модель и проводит наблюдения. Поэтому всякий раз, когда начинается очередной прогон модели процесса функционирования системы, требуется определенное время для достижения условий равновесия, которые соответствуют условиям функционирования реальной системы.

Таким образом, начальный период работы машинной модели искажается из-за влияния начальных условий запуска модели. Для решения этой проблемы либо исключается из рассмотрения информация о модели, полученная в начальной части периода моделирования, либо начальные условия выбираются так, чтобы сократить время достижения установившегося режима. Все эти приемы позволяют только уменьшить, но не свести к нулю время переходного процесса при проведении машинного эксперимента с моделью.

Обеспечение точности и достоверности. Решение второй проблемы тактического планирования машинного эксперимента связано с оценкой точности и достоверности результатов моделирования при заданном числе реализаций или с необходимостью оценки необходимого числа реализаций при заданных значениях точности и достоверности результатов моделирования системы.

Как уже отмечалось, статистическое моделирование системы это эксперимент с машинной моделью. Обработка результатов подобного имитационного эксперимента принципиально не может дать точных значений показателя эффективности E системы в лучшем случае можно получить только некоторую оценку \tilde{E} такого показателя. При этом экономические вопросы затрат людских и машинных ресурсов, обосновывающие целесообразность статистического моделирования вообще, оказываются тесно связанными с вопросами точности и достоверности оценки показателя эффективности E системы на ее модели.

Таким образом, количество реализаций N при статистическом моделировании системы должно выбираться исходя из двух основных соображений: определения затрат ресурсов на машинный эксперимент с моделью (включая построение модели и ее машинную реализацию) и оценки точности и достоверности результатов эксперимента с моделью системы (при заданных ограничениях на ресурсы). Очевидно, что требования получения более хороших оценок и сокращения затрат ресурсов являются противоречивыми и при планировании машинных экспериментов на базе статистического моделирования необходимо решить задачу нахождения разумного компромисса между ними.

Из-за наличия стохастичности и ограниченности числа реализаций N в общем случае $\tilde{E} \neq E$. При этом величину \tilde{E} называют абсолютной **точностью оценки**. Вероятность того, что неравенство

$$|E - \tilde{E}| < \varepsilon, \quad (3.1)$$

выполняется, называется достоверностью оценки, т.е.

$$Q = P\{|E - \tilde{E}| < \varepsilon\}. \quad (3.2)$$

Величина $\varepsilon_0 = \varepsilon/E$ называется относительной точностью оценки, а достоверность оценки соответственно будет иметь вид

$$Q = P\{|(E - \tilde{E})/E| < \varepsilon_0\}.$$

Для того, чтобы при статистическом моделировании системы по заданным E (или E_0) и Q определить количество реализаций N или, наоборот, при ограниченных ресурсах (известном N) найти необходимые E и Q , следует детально изучить соотношение (3.2). Сделать это можно не во всех случаях, так как закон распределения вероятностей величины $|E - \tilde{E}|$ для многих практических случаев исследования систем установить не удастся либо в силу ограниченности априорных сведений о системе, либо из-за сложности вероятностных расчетов. Основным путем преодоления подобных трудностей является выдвижение предположений о характере законов распределения случайной величины \tilde{E} , т. е. оценки показателя эффективности системы.

Рассмотрим взаимосвязь точности и достоверности результатов с количеством реализаций при машинном эксперименте, когда в качестве показателей эффективности E выступают **вероятность p , математическое ожидание a и дисперсия σ^2** .

Пусть цель машинного эксперимента с моделью некоторой системы — получение оценки \tilde{p} вероятности появления $p=P(A)$ некоторого события A , определяемого состояниями процесса функционирования исследуемой системы. В качестве оценки вероятности p в данном случае выступает частота $\tilde{p}=m/N$, где m — число положительных исходов.

Тогда соотношение (3.2), связывающее точность и достоверность оценок с количеством реализаций, будет иметь вид

$$P\{|p - m/N| < \varepsilon\} = Q, P\{p - \varepsilon < m/N < p + \varepsilon\} = Q. \quad (3.3)$$

Для ответа на вопрос о законе распределения величины $\tilde{p}=m/N$ представим эту частоту в виде

$$\tilde{p} = m/N = (1/N) \sum_{i=1}^N x_i,$$

так как количество наступлений события A в данной реализации из N реализаций является случайной величиной ξ , принимающей значения $x_1 = 1$ с вероятностью p и $x_2 = 0$ с вероятностью $1 - p$.

Математическое ожидание и дисперсия случайной величины ξ будут таковы:

$$M[\xi] = x_1 p + x_2 (1 - p) = 1p + 0(1 - p) = p;$$

$$D[\xi] = (x_1 - M[\xi])^2 p + (x_2 - M[\xi])^2 (1 - p) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p).$$

Тогда

$$M[\tilde{p}] = M[m/N] = (1/N) M\left[\sum_{i=1}^N x_i\right] = (1/N) N M[\xi] = p.$$

Это соотношение говорит о несмещенности оценки \tilde{p} для вероятности p . С учетом независимости значений величин x_i получим

$$D[\tilde{p}] = D[m/N] = (1/N^2) D\left[\sum_{i=1}^N x_i\right] = (1/N^2) N D[\xi] = p(1 - p)/N.$$

В силу центральной предельной теоремы теории вероятностей частоту m/N при достаточно больших N можно рассматривать как случайную величину, описываемую нормальным законом распределения вероятностей с математическим ожиданием p и дисперсией $p(1-p)/N$. Поэтому соотношение (5.3) можно переписать так:

$$P\left\{p - \varepsilon < \frac{m}{N} < p + \varepsilon\right\} = \Phi_0\left(\frac{p + \varepsilon - p}{\sqrt{p(1-p)}}\sqrt{N}\right) - \Phi_0\left(\frac{p - \varepsilon - p}{\sqrt{p(1-p)}}\sqrt{N}\right) = Q.$$

Учитывая, что $\Phi_0(-z) = 1 - \Phi_0(z)$, получим

$$2\Phi_0(\varepsilon\sqrt{N}/\sqrt{p(1-p)}) = 1 + Q; \Phi_0(\varepsilon\sqrt{N}/\sqrt{p(1-p)}) = (1 + Q)/2 = \varphi.$$

Тогда $\varepsilon\sqrt{N}/\sqrt{p(1-p)} = t_\varphi$,

где t_φ — квантиль нормального распределения вероятностей порядка $\varphi = (1 + Q)/2$; находится из специальных таблиц.

В результате точность оценки \tilde{p} вероятности p можно определить как

$$\varepsilon = t_\varphi \sqrt{p(1-p)/N},$$

т. е. точность оценки вероятностей обратно пропорциональна \sqrt{N} .

Из соотношения для точности оценки ε можно вычислить количество реализаций

$$N = t_\varphi^2 p(1-p)/\varepsilon^2, \quad (3.4)$$

для получения оценки \tilde{p} с точностью ε и достоверностью Q .

Пример. Необходимо рассчитать количество реализаций N при статистическом моделировании системы, когда в качестве показателя эффективности используется вероятность p при достоверности $Q = 0,95 (t_\varphi = 1,96)$ и точности $\varepsilon = 0,01; 0,02; 0,05$. Так как значения p до проведения статистического моделирования системы неизвестно, то вычислим множество оценок N для диапазона возможных значений p , т. е. от 0 до 1 с дискретом в 0,1. Результаты расчетов с использованием выражения (3.4) представлены в табл.1. Из таблицы видно, что при переходе от $p = 0,1(0,9)$ к $p = 0,5$ количество реализаций N возрастает примерно в три раза, а при переходе от $\varepsilon = 0,05$ к $\varepsilon = 0,01$ количество реализаций N возрастает примерно в 25 раз [2].

Таблица 1

Вероятность p	Точность ε		
	0,05	0,02	0,01
0,1 (0,9)	140	900	3600
0,2 (0,8)	250	1500	6200
0,3 (0,7)	330	2100	8400
0,4 (0,6)	380	2300	9400
0,5 (0,5)	390	2400	9800

При тактическом планировании машинного эксперимента, когда решается вопрос о выборе количества реализаций N , значение p неизвестно. Поэтому на практике проводят предварительное моделирование для произвольно выбранного значения N_0 , определяют $p_0 = m/N_0$, а затем по

(3.4) вычисляют, используя вместо p значение p_0 , необходимое количество реализаций N . Такая процедура оценки N может выполняться несколько раз в ходе машинного эксперимента с системой.

При отсутствии возможности получения каких-либо априорных сведений о вероятности p использование понятия абсолютной точности теряет смысл. Действительно, можно, например, предварительно задать точность результатов моделирования $\varepsilon=0,01$, а искомая p в результате окажется хотя бы на порядок ниже, т. е. $p \leq 0.001$. В таких случаях целесообразно задавать относительную точность результатов моделирования ε_0 . Тогда соотношение (3.4) примет вид

$$N=t_{\varphi}^2 (1-p)/(\varepsilon_0 p). \quad (3.5)$$

Соотношение (3.5) наглядно иллюстрирует специфику статистического моделирования систем, выражающуюся в том, что для оценивания малых вероятностей p с высокой точностью необходимо очень большое число реализаций N . На практике для оценки вероятностей с порядком в 10^{-k} количество реализаций целесообразно выбирать равным 10^{+k} . Очевидно, что даже для сравнительно простых систем метод статистического моделирования приводит к большим затратам машинного времени.

Другим распространенным случаем в практике машинных экспериментов с моделью является необходимость оценки показателей эффективности E системы по результатам определения среднего значения некоторой случайной величины. Пусть случайная величина ξ имеет математическое ожидание a и дисперсию σ^2 . В реализации с номером i она принимает значение x_i . В качестве оценки математического ожидания a используется среднее – арифметическое значение выборки

$$\bar{x} = (1/N) \sum_{i=1}^N x_i.$$

В силу центральной предельной теоремы теории вероятностей при больших значениях N среднее – арифметическое \bar{x} будет иметь распределение, близкое к нормальному с математическим ожиданием a и дисперсией σ^2/N . Для математического ожидания a точность оценки равна $\varepsilon = t_{\varphi} \sigma / \sqrt{N}$, а количество реализаций

$$N = t_{\varphi}^2 \sigma^2 / \varepsilon^2, \text{ или } N = t_{\varphi}^2 \varepsilon^2 / (\varepsilon_0^2 a^2). \quad (3.6)$$

Аналогично, если в качестве показателя эффективности E системы выступает дисперсия σ^2 , а в качестве ее оценки используется величина S^2 , то математическое ожидание и дисперсия соответственно будут

$$M[S^2] = (N-1)\sigma^2 / N; \quad D[S^2] = (\mu_4 - \sigma^4) / N,$$

где μ_4 - центральный момент четвертого порядка случайной величины.

Для дисперсии σ^2 точность оценки $\varepsilon = t_{\varphi} \sqrt{(\mu_4 - \sigma^4) / N}$.

Отсюда количество реализаций будет

$$N = t_{\varphi}^2 (\mu_4 - \sigma^4) / \varepsilon^2, \text{ или } N = t_{\varphi}^2 [(\mu_4 / \sigma^4) - 1] / \varepsilon_0^2. \quad (3.7)$$

Для частного случая, когда случайная величина имеет нормальное распределение при $\mu_4 = 3\sigma^4$, получим $N = t_\varphi^2 2\sigma^4 / \varepsilon^2 = 2t_\varphi^2 / \varepsilon_0^2$.

Таким образом, на основании соотношений (3.4) — (3.7) можно сделать вывод, что количество реализаций при статистическом моделировании существенно зависит от дисперсии оцениваемой случайной величины. Поэтому выгодно выбирать такие оцениваемые показатели эффективности E системы, которые имеют малые дисперсии.

Уменьшение дисперсии оценок. Основным недостатком методов планирования экспериментов, на основе использовании простой случайной выборки, является медленная сходимость выборочных средних к истинным средним с ростом объема выборки N . Это приводит к необходимости использования методов уменьшения ошибок, не требующих увеличения N . Такие методы называются *методами понижения дисперсии* и делятся на три группы [3]:

активные (предусматривают формирование выборки специальным образом;

пассивные (применяются после того, как выборка уже сформирована);

косвенные (в них для получения оценок наблюдаемой переменной используются значения некоторых вспомогательных величин).

Активных методов понижения дисперсии известно достаточно много. Выбор конкретного метода определяется, как правило, спецификой модели и целями эксперимента. Рассмотрим те из них, которые направлены на снижение влияния переходного периода. Выбор объясняется тем, что наличие и длительность переходного режима оказывают существенное влияние на точность результатов моделирования. Вместе с тем, большинство ИМ используется для изучения функционирования системы в установившемся режиме.

Существуют три основных метода уменьшения ошибок, обусловленных наличием переходного периода:

значительное увеличение длительности прогона;

исключение из рассмотрения переходного периода;

инициализация модели при некоторых специально выбранных начальных условиях.

На практике снижения влияния переходного периода обычно добиваются одни из следующих способов: методом повторения; методом подинтервалов; методом циклов.

Метод повторения. При использовании этого метода каждое наблюдение получается при помощи отдельного прогона модели, причем все прогоны начинаются при одних и тех же начальных условиях, но используются различные последовательности случайных чисел. Преимуществом метода является статистическая независимость получаемых наблюдений. Недостаток состоит в том, что наблюдения могут оказаться сильно смещенными под влиянием начальных условий.

Метод подинтервалов. Данный метод основан на разбиении каждого прогона модели на равные промежутки времени. Начало каждого интервала совпадает с началом очередного этапа наблюдений (на рис.3.1 в качестве наблюдаемой переменной используется длина очереди заявок – Q).

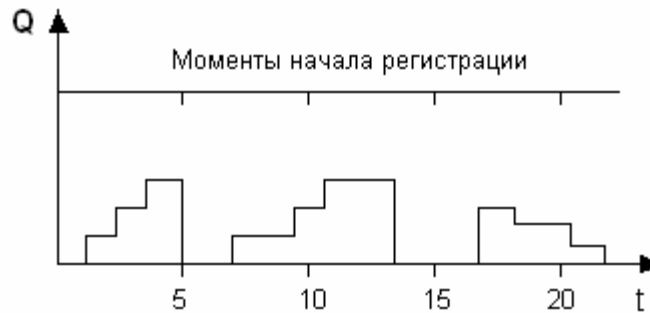


Рисунок 3.1 - Пример использования метода подинтервалов

Достоинство метода состоит в том, что влияние переходных условий со временем уменьшается, и наблюдения точнее отражают поведение системы в стационарном режиме. Недостаток – значения наблюдаемых переменных, полученных в начале очередного интервала, зависят от конечных условий предыдущего интервала (т.е. между интервалами имеется автокорреляция).

Метод циклов. При использовании метода циклов влияние автокорреляции уменьшается за счет выбора интервалов таким образом, чтобы в их начальных точках условия были одинаковыми. Например, в качестве таких условий можно рассматривать длину очереди заявок на обслуживание. В этом случае удобно совмещать начало очередного интервала с моментом, когда длина очереди становится равной нулю. Недостатком метода является меньшее по сравнению с методом подинтервалов число получаемых наблюдений.

Метод стратифицированной выборки. Данный метод относится к группе пассивных методов понижения дисперсии. Пассивные методы влияют на подготовку и проведение эксперимента, но реализуются на этапе обработки и анализа результатов моделирования. Суть метода стратифицированной выборки состоит в следующем.

Выборка разделяется на части, называемые слоями (стратами). При этом необходимо, чтобы значения элементов выборки как можно меньше различались внутри одного слоя и как можно больше – между различными слоями. Внутри каждого слоя производят случайный отбор элементов и вычисляют среднее значение слоя y_i . Полученные оценки используют для вычисления МОЖ по выборке в целом:

$$\tilde{Y} = \frac{1}{N} \sum_{i=1}^k n_i y_i,$$

где N , n_i – объем выборки и i -го слоя соответственно; k – число слоев.

Если считать, что оценки y_i независимы, то дисперсия по выборке в целом равна:

$$D_y = \frac{1}{N} \sum_{i=1}^k n_i D_{yi}.$$

Здесь D_{yi} – дисперсия для i -го слоя.

При удачном выборе слоев величины D_{yi} будут малы, а значит, и выборочная дисперсия D_y будет предпочтительнее, чем для оценки, полученной методами простой случайной выборки.

Косвенные методы понижения дисперсии основаны на том, что зачастую некоторые из выходных характеристик модели получить (вычислить) легче, чем другие. Их использование предполагает не только весьма глубокое знание сущности процессов, протекающих в системе, но и наличие формального описания взаимной зависимости параметров модели.

3.2 Статистический анализ результатов моделирования

3.2.1 Оценивание вероятностных распределений и их числовых характеристик

Важнейшим этапом в исследовании сложных систем является этап статистического анализа результатов экспериментов, к которым относятся выборки значений: входных воздействий; воздействий внешней среды; фазовой и выходной траекторий; показателей эффективности системы. Имитационная модель, описывая динамическое поведение исследуемой стохастической системы во времени, позволяет получить результаты, подобные тем, которые регистрируются в ходе натуральных экспериментов с реальной системой. Поэтому для статистического анализа результатов имитационных и натуральных экспериментов применимы одни и те же методы. Отличие состоит в том, что имитационное моделирование позволяет получать именно те данные о системе и в том объеме, которые необходимы для решения задач исследования.

В общем случае результаты моделирования образуют некоторый набор выборок, каждая из которых представляет собой ряд случайных величин (данных). Основными **задачами статистического анализа** таких данных являются:

- определение эмпирического закона распределения случайных величин;

- проверка однородности распределений;

- сравнение средних значений и дисперсий переменных полученных в результате моделирования.

Пусть $\xi \in R^1$ – случайная величина (СВ) с функцией распределения $F(x) = P\{\xi < x\}$, являющаяся математической моделью единичного наблюдения одной из компонент или одного из показателей, используемых в ходе имитационного моделирования.

На практике наибольшее распространение имеют два класса функций распределения $F(\cdot)$: 1) *непрерывные* и 2) *дискретные*. В первом случае существует плотность распределения вероятностей СВ ξ :

$$f(x) = \frac{dF(x)}{dx}, \quad x \in R^1. \quad (3.8)$$

Во втором случае СВ ξ принимает значения из дискретного множества $A = \{a_1, a_2, \dots, a_k\}$, $a_1 < a_2 < \dots < a_k$, $k \leq \infty$ и имеет дискретное распределение вероятностей

$$p_j = P\{\xi = a_j\} \quad (j = \overline{1, k}), \quad \sum_{j=1}^k p_j = 1. \quad (3.9)$$

Вероятностная модель наблюдения ξ полностью описывается *функциональными характеристиками*: функцией распределения $F(\cdot)$, плотностью распределения (3.8) или дискретным распределением вероятностей (3.9). Когда получить полное вероятностное описание не представляется возможным, то на практике используются числовые характеристики. Любая числовая характеристика $\lambda \in R$ является некоторым функционалом от $F(\cdot)$: $\lambda = \lambda(F(\cdot))$. Множество числовых характеристик состоит из двух подмножеств.

Числовые характеристики положения (сдвига):

- математическое ожидание (среднее) $\mu = M\{\xi\}$;
- медиана $M_e : F(M_e) = 1/2$;
- мода $M_0 = \arg \max_x f(x)$;
- наибольшее a_+ и наименьшее a_- значения, $P\{a_- \leq \xi \leq a_+\} = 1$,
 $P\{a_- + \varepsilon \leq \xi \leq a_+ - \varepsilon\} < 1$, $\varepsilon > 0$.

Числовые характеристики рассеяния (масштаба):

- дисперсия $D = D\{\xi\} = M\{(\xi - \mu)^2\}$;
- среднеквадратичное (стандартное) отклонение $\sigma = \sqrt{D\{\xi\}} \geq 0$;
- коэффициент вариации (если $\mu \neq 0$) $K = \frac{\sigma}{|\mu|} \geq 0$;
- размах $\alpha = a_+ - a_-$;
- коэффициент асимметрии $\beta_1 = M\{(\xi - \mu)^3\} / \sigma^3$;
- коэффициент эксцесса (островершинности) $\beta_2 = M\{(\xi - \mu)^4\} / \sigma^4 - 3$.

Если ξ имеет гауссовский закон распределения вероятностей $Z(\xi) = N_1(a, B)$, то $\mu = M_e = M_0 = a$, $a_- = -\infty$, $a_+ = +\infty$, $D = B$, $\sigma = \sqrt{B}$, $K = \sqrt{B} / |a|$, $\alpha = +\infty$, $\beta_1 = \beta_2 = 0$.

В процессе моделирования функциональные и числовые характеристики вероятностного распределения часто неизвестны и возникает задача их статистического оценивания по случайной выборке $Z = (Z_1, Z_2, \dots, Z_n) \in R^n$ объема n из распределения $F(\cdot)$, полученной в результате n «независимых прогонов» имитационной модели. Сначала рассмотрим задачу оценивания $F(\cdot)$, $f(\cdot)$, $\{p_j\}$ по Z . *Эмпирическая функция распределения*

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x)}(Z_i), \quad x \in R \quad (3.10)$$

является состоятельной, несмещенной оценкой с наименьшей вариацией

$$M\{(\hat{F}(x) - F(x))^2\} = F(x)(1 - F(x))/n, \quad x \in R.$$

В случае дискретной вероятностной модели (3.9) относительная частота

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^n \delta_{Z_i, a_j} \quad (j = \overline{1, k})$$

является несмещенной, состоятельной эффективной и асимптотически нормально распределенной оценкой элементарной вероятности p_j ; при этом

$$\text{вариация } M\{(\hat{p}_j - p_j)^2\} = p_j(1 - p_j)/n.$$

В случае непрерывной вероятностной модели (3.8) для оценивания плотности $f(\cdot)$ применяется гистограмма. Для её построения разобьем промежуток $[a_-, a_+]$ концентрации плотности $f(\cdot)$ на L ячеек точками деления $b_0 < b_1 < \dots < b_L$:

$$b_0 = a_-, \quad b_L = a_+, \quad [a_-, a_+] = \bigcup_{l=1}^L [b_{l-1}, b_l].$$

$$\text{Обозначим } \nu_l = \sum_{i=1}^n I_{[b_{l-1}, b_l]}(Z_i) \quad (l = \overline{1, L}) \quad (3.11)$$

число выборочных значений из Z , попавших в l -ю ячейку гистограммы. *Гистограммой* является статистика

$$\hat{f}(x) = \sum_{l=1}^L \frac{\nu_l}{n(b_l - b_{l-1})} \cdot I_{[b_{l-1}, b_l]}(x), \quad x \in R.$$

Эта оценка является смещенной и несостоятельной в общем случае.

Для оценивания числовых характеристик по выборке Z широко используется *подстановочный принцип*:

$$\hat{\lambda} = \lambda(\hat{F}(\cdot)),$$

где $\hat{F}(\cdot)$ – выборочная функция распределения (3.10).

Статистические оценки числовых характеристик положения имеют вид:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n Z_i; \quad \hat{M}_l : \hat{F}(\hat{M}_l) = \frac{1}{2}; \quad (3.12)$$

$$\hat{M}_0 = \arg \max_x \hat{f}(x); \quad \hat{a}_+ = \max_i Z_i; \quad \hat{a}_- = \min_i Z_i.$$

Выборочное среднее $\hat{\mu} = \bar{x}$ является состоятельной несмещенной оценкой с вариацией $M\{(\hat{\mu} - \mu)^2\} = \sigma^2/n$. Статистические точечные оценки числовых характеристик рассеяния принимают вид:

$$\hat{D} = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \hat{\mu})^2; \quad \hat{\sigma} = \sqrt{\hat{D}}; \quad \hat{K} = \hat{\sigma} / |\hat{\mu}|; \quad \hat{a} = \hat{a}_+ - \hat{a}_-;$$

$$\hat{\beta}_1 = \frac{\sum (Z_i - \hat{\mu})^3 / n}{\hat{\sigma}^3} \quad (3.13)$$

$$\hat{\beta}_2 = \frac{\sum (Z_i - \hat{\mu})^4 / n}{\hat{\sigma}^4} - 3.$$

3.2.2 Проверка адекватности моделей

Решения, принимаемые исследователем по результатам имитационного моделирования, могут быть конструктивными только при выполнении двух основных условий:

- полученные результаты обладают требуемой точностью и достоверностью;
- исследователь способен правильно интерпретировать полученные результаты, и знает, каким образом они могут быть использованы.

Возможность выполнения первого условия закладывается, в основном, еще на этапе разработки модели и частично – на этапе планирования эксперимента. Достоверность результатов моделирования предполагает, что модель, с помощью которой они получены, не только является «правильной», но отвечает и некоторым дополнительным требованиям, предъявляемым к имитационным моделям.

Оценка качества имитационной модели

Оценка качества модели является завершающим этапом ее разработки и преследует две цели:

- проверить соответствие модели ее предназначению (целям исследования);
- оценить достоверность и статистические характеристики результатов, получаемых при проведении модельных экспериментов.

При аналитическом моделировании достоверность результатов определяется двумя основными факторами:

- корректным выбором математического аппарата, используемого для описания исследуемой системы;
- методической ошибкой, присущей данному математическому методу.

При имитационном моделировании на достоверность результатов влияет целый ряд дополнительных факторов, основными из которых являются:

моделирование случайных факторов, основанное на использовании датчиков случайных чисел, которые могут вносить «искажения» в поведение модели;

наличие нестационарного режима работы модели;

использование нескольких разнотипных математических методов в рамках одной модели;

зависимость результатов моделирования от плана эксперимента;

необходимость синхронизации работы отдельных компонент модели;

наличие модели рабочей нагрузки, качество которой зависит, в свою очередь, от тех же факторов.

Пригодность имитационной модели для решения задач исследования характеризуется тем, в какой степени она обладает так называемыми *целевыми свойствами*. Основными из них являются: **адекватность**; **устойчивость**; **чувствительность**. Рассмотрим их подробнее.

Оценка адекватности

В общем случае под адекватностью понимают степень соответствия модели тому реальному явлению или объекту, для описания которого она строится.

Вместе с тем, создаваемая модель ориентирована, как правило, на исследование определенного подмножества свойств этого объекта. Поэтому можно считать, что адекватность модели определяется степенью ее соответствия не столько реальному объекту, сколько целям исследования. В наибольшей степени это утверждение справедливо относительно моделей проектируемых систем (то есть в ситуациях, когда реальная система вообще не существует).

Итак, каким же образом можно оценить адекватность разработанной модели реально существующей системе?

Процедура оценки основана на сравнении измерений на реальной системе и результатов экспериментов на модели и может проводиться различными способами. Наиболее распространенные из них:

по средним значениям откликов модели и системы;

по дисперсиям отклонений откликов модели от среднего значения откликов системы;

по максимальному значению относительных отклонений откликов модели от откликов системы.

Названные способы оценки достаточно близки между собой, поэтому ограничимся рассмотрением первого из них. При этом способе проверяется гипотеза о близости среднего значения наблюдаемой переменной Y среднему значению отклика реальной системы Y_p . В результате N опытов на реальной системе получают множество значений (выборку) Y_p . Выполнив N экспериментов на модели, также получают множества значений наблюдаемой переменной Y .

Затем вычисляются оценки математического ожидания и дисперсии откликов модели и системы, после чего выдвигается гипотеза о близости средних значений величин Y_p и Y (в статистическом смысле). Основой для проверки гипотезы является t -статистика (распределение Стьюдента). Ее значение, вычисленное по результатам испытаний, сравнивается с критическим значением $t_{кр}$, взятым из справочной таблицы. Если выполняется неравенство $t_n < t_{кр}$, то гипотеза принимается.

Необходимо еще раз подчеркнуть, что статистические методы применимы только в том случае, если оценивается адекватность модели существующей системе. На проектируемой системе провести измерения, естественно, не представляется возможным. Единственный способ преодолеть это препятствие заключается в том, чтобы принять в качестве эталонного объекта концептуальную модель проектируемой системы. Тогда оценка адекватности программно реализованной модели заключается в проверке того, насколько корректно она отражает концептуальную модель.

В этом случае проверку адекватности модели целесообразно производить методом оценки расхождения эмпирического закона распределения, определенного по результатам моделирования, с некоторым предполагаемым теоретическим законом распределения характеристик исследуемой системы. Пусть результаты моделирования $Z = (z_1, \dots, z_n)$ представляют собой случайную выборку объема n из некоторого распределения вероятностей с неизвестной функцией распределения. Пусть $F_0(x)$ – некоторая фиксированная предполагаемая функция распределения, например, задаваемая требованиями имитационной модели. Определяются простая гипотеза $H_0 : F(x) = F_0(x), x \in R$ и сложная альтернатива общего вида $H_1 = \bar{H}_0$.

Задача проверки адекватности модели $F_0(\cdot)$ заключается в построении критерия для проверки H_0, H_1 по выборке Z с заданным уровнем значимости $\varepsilon \in (0,1)$. Гипотеза H_0 означает, что результаты моделирования Z согласуются с распределением $F_0(\cdot)$, и поэтому она называется *гипотезой согласия*. Критерий для проверки H_0, H_1 называется *критерием согласия*. Далее рассмотрим два основных критерия согласия: χ^2 – критерий Пирсона и критерий Колмогорова.

χ^2 критерий согласия Пирсона. Пусть, как и при построении гистограммы, согласно (6.4) вычислены частоты $\{v_l\}$ попадания выборочных значений Z в ячейки гистограммы и соответствующие гипотетические вероятности попадания в ячейки, если верна H_0 :

$$p_l^0 = P_{H_0} \{Z_i \in [b_{l-1}, b_l)\} = F_0(b_l) - F_0(b_{l-1}), \quad l = \overline{1, L}.$$

Статистикой χ^2 для задачи проверки адекватности модели называется статистика

$$\chi^2 = n \sum_{l=1}^L \frac{(v_l/n - p_l^0)^2}{p_l^0} = \sum_{l=1}^n \frac{(v_l - np_l^0)^2}{np_l^0} \geq 0. \quad (3.14)$$

Статистика (6.7) характеризует взвешенную сумму квадратов отклонений относительно частот $\frac{v_l}{n}$ от гипотетических значений $p_l^0 (l = \overline{1, L})$. Чем больше χ^2 , тем сильнее выборка Z не согласуется с гипотезой H_0 .

χ^2 – критерий согласия Пирсона основан на (6.7) и имеет вид:

$$\text{принимается гипотеза} \begin{cases} H_0, \text{ если } 1 - G_{l-1}(\chi^2) > \varepsilon, \\ H_1, \text{ если } 1 - G_{l-1}(\chi^2) \leq \varepsilon, \end{cases} \quad (3.15)$$

где $G_{l-1}(\cdot)$ – функция χ^2 – распределения с $L-1$ степенью свободы, ε – уровень значимости. Статистика $P = 1 - G_{l-1}(\chi^2)$ называется P – значением. Основное свойство критерия (3.15) выражается следующей теоремой.

Теорема. Если $0 < p_l^0 < 1$, $l = \overline{1, L}$, то при $n \rightarrow \infty$ асимптотический размер (вероятность ошибки I рода) критерия (3.15) совпадает с наперёд заданным уровнем значимости ε .

Таким образом, при увеличении числа прогонов n гарантируется заданный уровень надёжности решений. Чтобы повысить скорость сходимости вероятности ошибки первого рода $P_1 \rightarrow \varepsilon$, необходимо воспользоваться произволом в выборе числа L ячеек и границ $\{b_l\}$ таким образом, чтобы $n \cdot p_l^0 \geq 5$ ($l = \overline{1, L}$).

Критерий согласия Колмогорова. Пусть согласно (3.10) по результатам моделирования Z вычислена выборочная функция распределения $\hat{F}(\cdot)$, а гипотетическая функция распределения $F_0(\cdot)$ – непрерывна. Статистика

$$D_n = \sup_{x \in R} |\hat{F}(x) - F_0(x)| \in [0, 1] \quad (3.16)$$

называется *расстоянием Колмогорова* между $\hat{F}(\cdot)$ и $F_0(\cdot)$.

Отметим важное свойство статистики (3.16).

Теорема. Если верна H_0 и $u = F_0(x)$ – непрерывная функция, то распределение вероятностей статистики D_n не зависит от конкретного вида $F_0(\cdot)$:

$$Z_{H_0} \{D_n\} = Z \{D_n^*\}, \quad D_n^* = \sup_{0 \leq u \leq 1} |\Psi_n(u) - u|, \quad (3.17)$$

где

$$\Psi_n(u) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, u)}(u_i) \quad (3.18)$$

выборочная функция распределения, вычисленная по случайной выборке $U = (u_1, u_2, \dots, u_n)$ объема n из стандартного равномерного распределения $R[0, 1]$.

Ценность этого свойства состоит в том, что $Z_{H_0}\{D_n\}$ не зависит от $F_0(\cdot)$, является некоторым стандартным вероятностным распределением, по которому составлены таблицы для каждого значения n . Кроме того, это распределение при большом n описывается *распределением Колмогорова*

$$P_{H_0}\{\sqrt{n}D_n < y\} \rightarrow K(y) = 1 - 2 \sum_{j=1}^{+\infty} (-1)^{j-1} e^{-2j^2 y^2},$$

$$y \geq 0. \quad (3.19)$$

Сходимость (3.19) достаточно быстрая и функцию распределения Колмогорова $K(\cdot)$ из (3.19) рекомендуется использовать, начиная с $n \geq 20$. Поэтому критерий согласия Колмогорова, основанный на свойствах статистики D_n , имеет вид:

$$\text{принимается гипотеза } \begin{cases} H_0, \text{ если } 1 - K \cdot (\sqrt{n}D_n) > \varepsilon, \\ H_1, \text{ если } 1 - K \cdot (\sqrt{n}D_n) \leq \varepsilon, \end{cases} \quad (3.20)$$

где ε – уровень значимости.

3.2.3 Проверка устойчивости и чувствительности

При оценке адекватности модели как существующей, так и проектируемой системы реально может быть использовано лишь ограниченное подмножество всех возможных значений входных параметров. В связи с этим для обоснования достоверности получаемых результатов моделирования большое значение имеет проверка устойчивости модели.

Устойчивость модели – это ее способность сохранять адекватность при исследовании эффективности системы на всем возможном диапазоне рабочей нагрузки, а также при внесении изменений в конфигурацию системы.

Каким образом может быть оценена устойчивость модели? Универсальной процедуры проверки устойчивости модели не существует. Разработчик вынужден прибегать к методам «для данного случая», частичным тестам и здравому смыслу. Часто полезна апостериорная проверка. Она состоит в сравнении результатов моделирования и результатов измерений на системе после внесения в нее изменений. Если результаты моделирования приемлемы, уверенность в устойчивости модели возрастает.

В общем случае можно утверждать, что чем ближе структура модели структуре системы и чем выше степень детализации, тем устойчивее модель,

Устойчивость результатов моделирования может быть также оценена методами математической статистики. Она заключается в том, чтобы проверить гипотезу относительно свойств некоторого множества элементов, называемого генеральной совокупностью, оценивая свойства какого-либо подмножества генеральной совокупности (то есть выборки). В генеральной совокупности исследователя обычно интересует некоторый признак, который обусловлен случайностью и может иметь качественный или количественный характер.

В данном случае именно устойчивость результатов моделирования можно рассматривать как признак, подлежащий оценке. Для проверки гипотезы об устойчивости результатов может быть использован как критерий согласия Смирнова, так и критерий Вилкоксона. Эти критерии применяют для проверки того, относятся ли две выборки к одной и той же генеральной совокупности.

При статистической оценке устойчивости модели соответствующая гипотеза может быть сформулирована следующим образом: при изменении входного воздействия или структуры модели закон распределения результатов моделирования остается неизменным.

Проверку указанной гипотезы H проводят при следующих исходных данных.

Пусть имеются две выборки $X = (x_1, x_2, \dots, x_n)$ и $Y = (y_1, y_2, \dots, y_n)$, полученные для различных значений рабочей нагрузки; относительно законов распределения X и Y никаких предположений не делается.

Значения обеих выборок упорядочиваются вместе по возрастанию. Затем анализируется взаимное расположение x_i, y_j . В случае $y_j < x_i$ говорят, что пара значений (x_i, y_j) образует инверсию.

Например, пусть для $n=m=3$ после упорядочивания получилась такая последовательность значений: $y_1, x_1, y_3, x_2, y_2, x_3$, тогда имеем инверсии: $(x_1, y_1), (x_2, y_1), (x_2, y_3), (x_3, y_1), (x_3, y_2), (x_3, y_3)$. Подсчитывают полное число инверсий U . Если гипотеза верна, то U не должно сильно отклоняться от своего математического ожидания M :

$$M = n * m / 2.$$

От гипотезы отказываются, если $|U - M| > U_{кр}$ ($U_{кр}$ определяют по таблице для заданного уровня значимости).

Очевидно, что устойчивость является положительным свойством модели. Однако, если изменение входных воздействий или параметров модели (в некотором заданном диапазоне) не отражается на значениях выходных параметров, то польза от такой модели невелика. В связи с этим возникает задача оценивания **чувствительности** модели к изменению входных и внутренних параметров системы.

Такую оценку проводят по каждому параметру X в отдельности. Основана она на том, что обычно диапазон возможных изменений параметра известен. Одна из наиболее простых и распространенных процедур оценивания состоит в следующем.

Вычисляется величина относительного среднего приращения параметра X :

$$\Delta X = 2(X_{\max} - x_{\min})100\% / (X_{\max} + X_{\min})$$

Проводится пара модельных экспериментов при значениях $X = X_{\max}$, $X = X_{\min}$ и средних фиксированных значениях остальных параметров. Определяются значения отклика модели $Y_1 = f(X_{\max})$, $Y_2 = f(X_{\min})$. Вычисляется относительное приращение наблюдаемой переменной Y :

$$\Delta Y = 2|Y_1 - Y_2|100\% / (Y_1 + Y_2).$$

В результате для k -го параметра модели имеют пару значений $\{\Delta X_k, \Delta Y_k\}$, характеризующую чувствительность модели по этому параметру.

Аналогично формируются пары для остальных параметров модели, которые образуют множество $\{\Delta X_k, \Delta Y_k\}$.

Данные, полученные при оценке чувствительности модели, могут быть использованы, в частности, при планировании экспериментов: большее внимание должно уделяться тем параметрам, по которым модель является более чувствительной.

3.2.4 Критерии согласия

В некоторых случаях имитационная модель сложной системы может быть реализована в виде набора отдельных моделей ее подсистем. При проведении экспериментов с такой моделью в целях сокращения затрат времени бывает необходимо заменять моделирование работы одной из подсистем некоторым числовым параметром (принцип параметризации), либо случайной величиной, распределенной по заданному закону. Чтобы такая замена была выполнена корректно, исследователь должен располагать описанием зависимости данного числового параметра от времени и других факторов, фигурирующих в модели.

При имитационном моделировании подбор законов распределений выполняется на основе статистических данных, полученных в ходе эксперимента. Оценку соответствия эмпирического закона распределения случайной величины некоторому теоретическому закону осуществляют проверкой статистических гипотез.

Статистическая гипотеза – это утверждение относительно значений одного или более параметров распределения некоторой величины или о самой форме распределения. Обычно выбирают две исходные гипотезы: основную – H_0 и альтернативную ей – H_1 .

Статистическая проверка гипотезы – это процедура выяснения, следует ли принять основную гипотезу H_0 или отвергнуть ее. Если в результате проверки гипотеза H_0 ошибочно отвергается, то имеет место ошибка первого рода (характеризующаяся более тяжелыми последствиями); если гипотеза H_0 принимается при истинности H_1 – ошибка второго рода.

Вероятности ошибок I и II рода (α и β) зависят от критерия, на основании которого будет выбираться одна из гипотез. Очевидно, что

вероятности этих двух ошибок взаимосвязаны, то есть чем больше значение α , тем меньше β , и наоборот. Обычное решение этой дилеммы состоит в том, что выбирают некоторое фиксированное значение α (как правило, $0,05$, $0,1$, $0,001$) и надеются, что β будет также мало. Фиксированное значение α называется *уровнем значимости*.

Для выбранного значения α определяется так называемая критическая область B , удовлетворяющая условию:

$$p (Z \in B \mid H_0 \text{ верна}) \leq \alpha .$$

Здесь Z – контрольная величина (критерий), представляющая собой некоторую функцию от выборки (результатов эксперимента).

Проверка гипотезы состоит в следующем. Производится выборка (эксперимент), на основании чего вычисляется z – частное значение критерия Z . Если $z \in B$, то от гипотезы H_0 отказываются. Если z не принадлежит B , то говорят, что полученные наблюдения *не противоречат* принятой гипотезе.

Разумеется, прежде чем выдвигать гипотезу относительно значений параметров распределения, необходимо определить вид самого закона распределения. Наиболее распространенный на практике и достаточно эффективный метод подбора закона распределения основан на использовании графического представления экспериментальных данных. Они отображаются в виде так называемой гистограммы относительных частот, которая может быть построена как вручную, так и с помощью соответствующих инструментальных средств, входящих в состав большинства пакетов моделирования. Внешний вид гистограммы показан на рис.3.2.

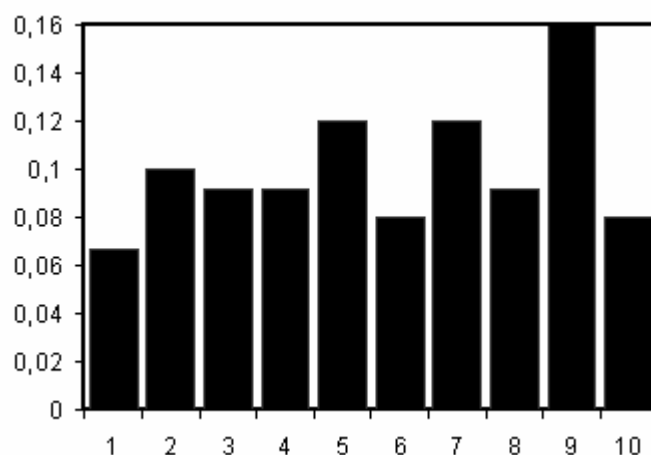


Рисунок 3.2 - Пример гистограммы относительных частот

Методика построения гистограммы относительных частот заключается в следующем.

а) вычисляется величина интервала гистограммы из следующего соотношения: $d = (y_{\max} - y_{\min})/n$, где $(y_{\max} - y_{\min})$ – диапазон изменения наблюдаемой переменной, n – число интервалов, выбранных исследователем.

б) по результатам моделирования определяется число попаданий значений y в i -й интервал.

в) вычисляется относительная частота попаданий наблюдаемой переменной в каждый интервал: $G_i = R_i / N$, где R_i – число попаданий в i -й интервал; N – объем выборки.

г) на каждом i -м интервале строится прямоугольник со сторонами (d, G_i) . Сумма площадей прямоугольников гистограммы равна единице.

Для наиболее часто используемых статистических гипотез разработаны критерии, позволяющие проводить их проверку с наибольшей достоверностью. Рассмотрим некоторые из них.

Критерий Стьюдента

Этот критерий служит для проверки гипотезы о равенстве средних значений двух нормально распределенных случайных величин X и Y в предположении, что дисперсии их равны (хотя и неизвестны). Сравниваемые выборки могут иметь разный объем.

В качестве критерия используют величину t :

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)D_x + (n_2 - 1)D_y}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}.$$

Величина t подчиняется распределению Стьюдента. Критическое значение критерия $t_{кр}$ определяется по таблице для выбранного значения числа α и числа степеней свободы $k = n_1 + n_2 - 2$. Если вычисленное по указанной формуле значение t удовлетворяет неравенству $t \geq t_{кр}$, то гипотезу H_0 отвергают.

Критерий Фишера

Этот критерий служит для проверки гипотезы о равенстве дисперсий D_x и D_y при условии, что X и Y распределены нормально. В качестве контрольной величины используется отношение дисперсий $F = D_x / D_y$ (большая из дисперсий должна быть в числителе).

Величина F подчиняется F -распределению (Фишера) с $m_1 = n_1 - 1, m_2 = n_2 - 1$ степенями свободы. Проверка гипотезы состоит в следующем.

Для величины $a = \alpha / 2$ и величин m_1, m_2 по таблице F -распределения выбирают значения $F(a, m_1, m_2)$. Если наблюдаемое F , вычисленное по выборке, больше этого критического значения, гипотеза должна быть отклонена с вероятностью ошибки α .

Критерий согласия Колмогорова-Смирнова

При использовании этого критерия имеющуюся выборку (y_1, \dots, y_n) упорядочивают по возрастанию и строят следующую эмпирическую функцию распределения:

$$F_y(y) = \begin{cases} 0, & -\infty < y < y_1 \\ i/n, & y_1 \leq y < y_{i+1}, i=1, \dots, n-1 \\ 1, & y_n \leq y < \infty \end{cases}$$

Контрольной является величина:

$$D = \max |F_y(y) - F_0(y)|.$$

Гипотеза $H_0: F_y(y) = F_0(y)$ отвергается, если вероятность попадания соответствующего критерия в критическую область оказывается меньше выбранного исследователем уровня значимости α . Критическое значение критерия, как и предыдущих случаях, находится по таблице.

Разумеется, проведение вручную расчетов, необходимых для проверки статистических гипотез, требует значительных затрат времени и сил. Поэтому многие современные математические пакеты имеют в своем составе средства, позволяющие свести к минимуму число операций, выполняемых пользователем вручную.

3.3 Статистическое исследование зависимостей

Представление результатов моделирования в форме функциональной зависимости показателя эффективности от того или иного фактора является конечной целью моделирования системы. Решение этой задачи может усложняться из-за возникновения взаимовлияния факторов. Поэтому отыскание аналитических зависимостей, связывающих между собой различные параметры, фигурирующие в модели, может быть основано на совместном использовании группы методов математической статистики: дисперсионного, корреляционного и регрессионного анализа.

3.3.1 Дисперсионный анализ

Дисперсионный анализ применяется для решения задачи сравнения средних значений нескольких выборок. Если при проверке окажется, что их математические ожидания отличаются незначительно, то все выборки объединяются в одну, что существенно увеличивает информацию о свойствах исследуемой системы.

Постановка задачи. Пусть генеральные совокупности случайных величин $\{y^{(1)}\}, \{y^{(2)}\}, \dots, \{y^{(n)}\}$ имеют нормальное распределение и одинаковую

дисперсию. Необходимо по выборочным средним значениям при некотором уровне значимости проверить нулевую гипотезу H_0 о равенстве математических ожиданий (то есть о независимости значений y от значений исследуемого фактора x).

Решение задачи для однофакторного дисперсионного анализа состоит в следующем. Пусть интересующий нас фактор x_j имеет k уровней. Для каждого из них получена выборка значений наблюдаемой переменной y : $y_j(1), y_j(2), \dots, y_j(i), \dots, y_j(k)$, где, k – количество уровней фактора x_j .

Оценим влияние фактора x факторной дисперсией D_x

$$D_x = \sum_{i=1}^k (y_i - \bar{y})^2 / k,$$

где \bar{y} – среднее – арифметическое значение y .

Если генеральная дисперсия $D[y]$ известна, то для оценки случайности разброса наблюдений необходимо сравнивать $D[y]$ с S_e^2 , используя F – критерий. При попадании наблюдаемого критерия F_3 в критическую область влияние фактора x считается значимым, а разброс значений x – неслучайным.

Если генеральная дисперсия $D[y]$ неизвестна до проведения машинного эксперимента, то необходимо найти ее оценку.

Пусть серия наблюдений на i – том уровне фактора x имеет вид $y_{i1}, y_{i2}, \dots, y_{in}$, где n – число повторных наблюдений на i – том уровне. Тогда на i -том уровне среднее значение

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij},$$

а среднее значение по всем уровням

$$\bar{y} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} = \frac{1}{k} \sum_{i=1}^k \bar{y}_i.$$

Общая выборочная дисперсия всех наблюдений

$$S_e^2 = \frac{1}{kn-1} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{1}{kn} \left(\sum_{i=1}^k \sum_{j=1}^n y_{ij} \right)^2 \right].$$

При этом разброс значений y определяется суммарным влиянием случайных причин и фактора x .

Задача дисперсионного анализа в том, чтобы разложить общую дисперсию $D[y]$ на составляющие связанные со случайными и неслучайными причинами. Оценка генеральной дисперсии, связанной со случайными причинами

$$\tilde{D}_0[y] = \frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n y_{ij} - \frac{1}{n} \sum_{i=1}^k \left(\sum_{j=1}^n y_{ij} \right)^2 \right],$$

а оценка факторной дисперсии определяется выражением

$$\tilde{D}_x = \tilde{D}[y] - \tilde{D}_0[y]$$

Учитывая, что факторная дисперсия наиболее заметна при анализе средних значений на i -том уровне фактора, а остаточная дисперсия (дисперсия случайности) для средних значений в n раз меньше, чем для отдельных измерений. Тогда более точная оценка выборочной дисперсии

$$\tilde{D}_x + \frac{1}{n} \tilde{D}_0[y] = \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y})^2.$$

Многофакторный дисперсионный анализ позволяет оценивать влияние на наблюдаемую переменную уже не одного, а произвольного числа факторов и выбрать из группы факторов, участвующих в эксперименте, те, которые действительно влияют на его результат.

Необходимо отметить, что дисперсионный анализ может использоваться для оценки влияния факторов, имеющих как количественный, так и качественный характер, поскольку в уравнении дисперсионного анализа фигурируют не сами факторы, а только их «эффекты». В том случае, если все факторы носят количественный характер, взаимосвязь между ними и наблюдаемой переменной может быть описана с помощью уравнения регрессии.

3.3.2 Корреляционный анализ

С помощью корреляционного анализа на основе оценки разброса значений переменных относительно их средних значений устанавливают степень зависимости между переменными x и y . Линейная зависимость между ними при нормальности их совместного распределения измеряется коэффициентом корреляции

$$r_{xy} = M[X - \bar{X}] * [Y - \bar{Y}] / \sigma_x \sigma_y.$$

Независимо от способа получения выборки для переменных необходимо:

построить диаграмму рассеяния, то есть графически отобразить точки (x_i, y_i) на плоскости (x, y) ;

провести анализ диаграммы рассеяния (рис. 3.3) для определения вида рассеяния.

Если $r_{xy} \neq 0$, то линейная зависимость между переменными есть, и она тем сильнее, чем больше r_{xy} . При $|r_{xy}|=1$ имеет место функциональная линейная зависимость между x и y вида $y = b_0 + b_1x$, причем, если $r_{xy} = +1$, то говорят о положительной корреляции, то есть большие значения одной величины соответствуют большим значениям другой; при $r_{xy} = -1$ имеет место отрицательная корреляция; при $0 < r_{xy} < 1$ вероятна линейная корреляция с рассеянием (рис. 7.1в), либо нелинейная корреляция (рис. 7.1г).

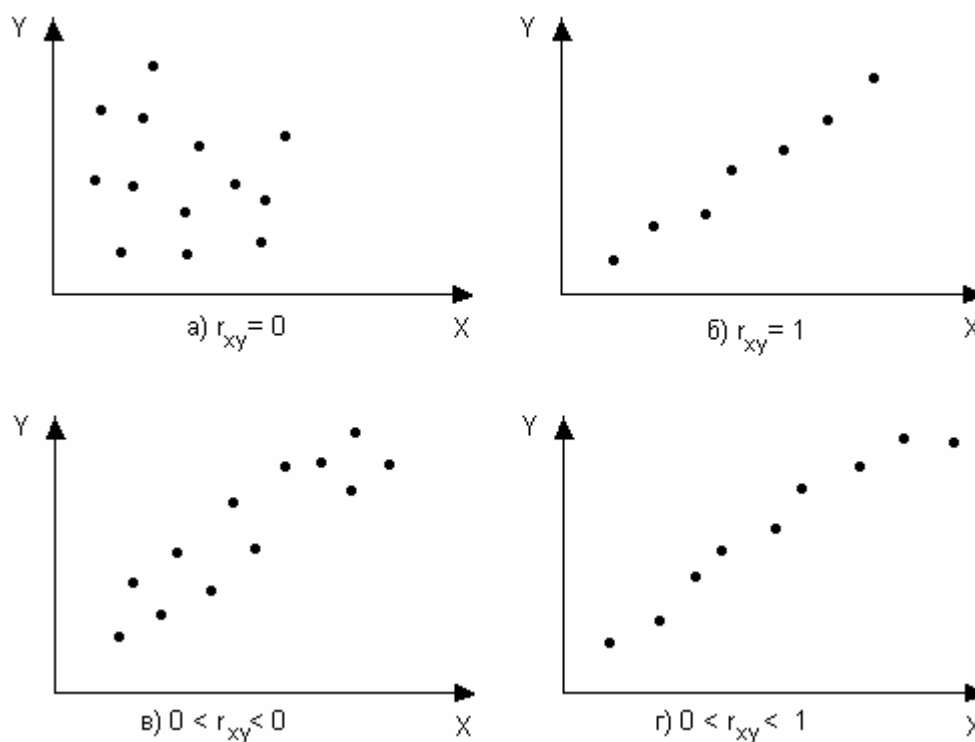


Рисунок 3.3 - Графическое представление корреляции между переменными

Для оценки точности определения r_{xy} может использоваться коэффициент w , подчиняющийся нормальному распределению

$$w = \ln\left[\frac{1+r_{xy}}{1-r_{xy}}\right] / 2$$

с математическим ожиданием и дисперсией

$$\mu_w = \ln\left[\frac{1+r_{xy}}{1-r_{xy}}\right] / 2, \quad \sigma_w^2 = 1/(N-3).$$

После анализа диаграммы рассеяния и получения значения r_{xy} необходимо убедиться в том, что действительно имеется статистически значимая корреляционная зависимость между переменными. Это делают проверкой нулевой гипотезы о равенстве коэффициента корреляции нулю.

Если нулевая гипотеза отвергается, то корреляционная зависимость признается значимой.

Наличие тесной зависимости между переменными еще не означает их причинно-следственной связи, поскольку может иметь место коррелированность последовательности случайных чисел, используемых для имитации случайных воздействий. Поэтому необходимо уточнить зависимость методами регрессионного анализа.

3.3.3 Регрессионный анализ

Регрессионный анализ заключается в построении модели в виде уравнения регрессии, которое связывает зависимое переменное с независимым и содержит неизвестные параметры, являющиеся коэффициентами уравнения. Если уравнение предполагается линейным относительно параметров, то регрессия считается линейной, в противном случае регрессия нелинейная. При этом решаются две задачи:

установление наличия причинно-следственной связи между переменными;

прогнозирование значений зависимой переменной по значениям независимой.

Если зависимость между x и y предполагается линейной, то она может быть представлена уравнением $\hat{y} = b_0 + b_1x$, где b_0, b_1 – параметры уравнения.

В этом случае целью регрессионного анализа является отыскание наилучших в статистическом смысле оценок параметров b_0, b_1 уравнения. Знание их позволяет легко найти прогнозируемое значение y при $x = x_i$.

Для нахождения оценок значений параметров b_0, b_1 применяют метод наименьших квадратов (МНК), согласно которому функция ошибки

$$F_0 = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (b_0 + b_1x_i - y_i)^2, \quad (3.21)$$

где N – объем выборки.

Условием минимума функции (3.21) является равенство нулю первых частных производных функции по параметрам b_0, b_1 , т.е.

$$\frac{\partial F_0}{\partial b_0} = 2(Nb_0 + b_1 \sum x_i - \sum y_i) = 0,$$

$$\frac{\partial F_0}{\partial b_1} = 2b_0 \sum x_i + 2b_1 \sum x_i^2 - 2 \sum x_i y_i = 0.$$

Решая эту систему уравнений, получают выражения для параметров уравнения регрессии b_0, b_1

$$b_0 = \left(\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i \right) / \left[N \sum x_i^2 - \left(\sum x_i \right)^2 \right],$$

$$b_1 = \left(N \sum x_i y_i - \sum x_i \sum y_i \right) / \left[N \sum x_i^2 - \left(\sum x_i \right)^2 \right].$$

Для нормально распределенных процессов примерно 67% точек находится в пределах стандартного отклонения от линии регрессии и 95% - в пределах двух стандартных отклонений.

В случае нескольких независимых переменных имеет место множественная линейная регрессия. В этом случае для отыскания оценок также используется метод наименьших квадратов.

В случае нелинейной регрессии основой для построения регрессионной модели опять-таки является МНК. Однако в этом случае для отыскания оценок параметров строится система нелинейных уравнений (относительно параметров), а для ее решения используются различные итерационные методы.

Вопросы для самопроверки

1. В чем необходимость и сущность стратегического планирования эксперимента?
2. Какой план применяют при идентификации факторов?
3. Как обеспечивается требуемая точность и достоверность результатов моделирования?
4. Перечислите числовые характеристики распределения экспериментальных данных.
5. Сущность оценки адекватности модели.
6. Устойчивость модели и процедура ее проверки.
7. В каких случаях применяется дисперсионный анализ?
8. Почему после корреляционного анализа необходимо производить и регрессионный анализ?
9. Раскройте процедуру регрессионного анализа.
10. Чем отличается критерий согласия Смирнова от критерия согласия Пирсона?

4 Моделирование динамических систем

4.1 Идентификация динамических систем

Под динамическим объектом понимают объект, выход которого зависит не только от текущих значений входного воздействия, но и от его значений в предыдущие моменты времени. Среди них выделяют линейные и нелинейные, стационарные и нестационарные, непрерывные и дискретные объекты.

Построение модели динамического объекта, чаще всего осуществляют по реализациям входных и выходных сигналов. Такая процедура носит название идентификации объекта, она обычно включает решение следующих основных задач:

- выбор класса математической модели;
- выбор класса входных сигналов;
- выбор критерия соответствия модели и объекта;
- выбор метода и алгоритма решения.

Наименее формализовано решение задачи выбора класса моделей, в качестве которых в теории управления используют дифференциальные, разностные и интегральные уравнения, а также частотные характеристики, интерполяционные ряды и др. модели.

Если модель объекта представить в виде схемы (рис.4.1) с входным сигналом $x(t)$, выходным теоретическим сигналом $\tilde{y}(t)$ и наблюдаемым выходным сигналом $y(t)$, то их связи выражаются равенствами (4.1)

$$\tilde{y}(t) = \psi[x(t)], \text{ а } y(t) = \psi[x(t)] + e(t), \quad (4.1)$$

где ψ - оператор связи входного и выходного сигналов,
 $e(t)$ - аддитивный шум наблюдения.

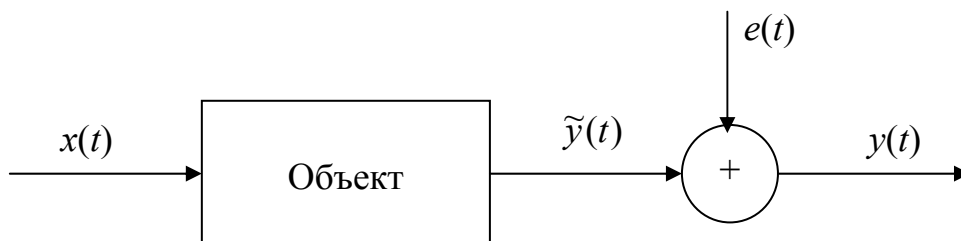


Рисунок 4.1 – схема модели динамического объекта

В этом случае целью идентификации объекта становится определение вида оператора связывающего входной сигнал $x(t)$ с выходным теоретическим сигналом $\tilde{y}(t)$ на основе наблюдений за выходным сигналом $x(t)$ и реакцией $y(t)$ на интервале времени Δt . Задача может решаться как аналитически, так и экспериментально.

При экспериментальном подходе осуществляют подбор адекватной структуры модели и выбор входного воздействия $x(t)$ так, чтобы по результатам эксперимента можно было найти оценки всех параметров модели.

Отсюда, под идентификацией динамического объекта в узком смысле необходимо понимать процедуру определения структуры и параметров его модели, которые при одинаковых входных сигналах для модели и объекта обеспечивают близость выхода модели к выходу объекта при наличии определенного критерия качества. При этом критериями близости выходных сигналов модели и объекта могут быть: среднеквадратичная, абсолютная, относительная погрешности, максимум правдоподобия и др.

Параметрическая идентификация модели по этим критериям осуществляется с помощью методов поиска экстремума, основными из которых являются: метод Ньютона-Канторовича; градиентный метод; метод случайного поиска и др. Но особая роль отводится рекуррентному методу наименьших квадратов.

Таким образом, к этапам моделирования динамического объекта относят:

- структурную идентификацию, суть которой сводится к определению структуры модели либо на основе теоретических соображений, либо в результате эксперимента;
- параметрическую идентификацию (проведение идентифицирующего эксперимента и определение оценок параметров по экспериментальным данным);
- проверку адекватности (оценка качества модели по значению выбранного критерия близости выходов модели и объекта).

В зависимости от типа объекта, вида модели, способа обработки информации (статистического или нестатистического) и типа эксперимента (пассивного или активного) различают множество задач параметрической идентификации. В этих задачах основную роль играют модели динамических объектов в форме дифференциальных уравнений, передаточных функций, частотных (амплитудная и фазовая) и временных (импульсная и переходная) характеристик.

Если модель линейного стационарного объекта описана в виде неоднородного линейного дифференциального уравнения (ДУ) с постоянными коэффициентами, например, выражением

$$\begin{aligned} a_n y^{(n)}(t) + a_{n-1} y^{(n-1)} + \dots + a_1 y^{(1)}(t) + a_0 y(t) = \\ b_m x^{(m)}(t) + b_{m-1} x^{(m-1)}(t) + \dots + b_1 x^{(1)}(t) + b_0 x(t) \end{aligned} \quad (4.2)$$

то из него, используя преобразование Лапласа, можно получить выражения для остальных видов моделей.

Для этого преобразуют уравнение (4.2) в операторное при нулевых начальных условиях, т.е.

$$(a_n p^n + \dots + a_1 p + a_0) \cdot y(p) = (b_m p^m + \dots + b_1 p + b_0) \cdot x(p). \quad (4.3)$$

Откуда передаточная функция $W(p)$ будет равна

$$W(p) = \frac{y(p)}{x(p)} = \frac{b_m p^m + \dots + b_1 p + b_0}{a_n p^n + \dots + a_1 p + a_0}. \quad (4.4)$$

Условием физической реализуемости объекта с передаточной функцией (4.4) является выполнение неравенства $m \leq n$.

Затем путем несложных математических операций получают: амплитудно-фазовую характеристику $W(j\omega)$

$$W(j\omega) = W(p) \Big|_{p=j\omega}; \quad (4.5)$$

амплитудную и фазовую частотные характеристики

$$W(\omega) = |W(j\omega)|, \quad \varphi(\omega) = \arg W(j\omega) = \arctg \frac{\text{Im}(\omega)}{\text{Re}(\omega)}.$$

Используя обратное преобразование Лапласа, получают выражения для импульсной $w(t)$ и переходной $h(t)$ характеристик

$$w(t) = L^{-1} \{W(p)\}, \quad h(t) = L^{-1} \left\{ \frac{W(p)}{p} \right\}. \quad (4.6)$$

Изображение и оригинал выходного сигнала можно определить соотношениями

$$Y(p) = W(p) \cdot X(p); \quad y(t) = \int_{-\infty}^{\infty} W(t - \tau) \cdot x(\tau) d\tau. \quad (4.7)$$

Передаточная функция (4.4) отражает структуру объекта, и ее получение означает осуществление структурной идентификации объекта. Если выражение (4.4) представить в виде произведения передаточных функций типовых динамических звеньев, то структуру объекта можно отобразить графом или схемой.

Проведение такой декомпозиции объекта на типовые звенья с известными характеристиками позволяет в дальнейшем осуществить параметрическую идентификацию в процессе эксперимента, т.е. установить ограничения на параметры и характеристики объекта.

4.2 Алгоритмизация непрерывно – детерминированных моделей

Для моделирования объекта на ЦВМ необходима алгоритмизация математического описания, которая заключается в получении соответствующего ему алгоритма. Алгоритмизация, например, ДУ сводится к составлению алгоритма численного решения этого ДУ.

При этом стремятся получить аналитическую форму алгоритма в виде явной функции соответствующих аргументов или в виде рекуррентной формулы, которая основана на использовании решетчатых функций и находит самое широкое применение при алгоритмизации функций, ДУ, таблиц, математических моделей САУ и др. на ЦВМ.

4.2.1 Решетчатая функция. Разностные и рекуррентные уравнения

Решетчатая функция (РФ) - это вещественная функция целочисленного аргумента k , принимающего значения $0, \pm 1, \pm 2, \dots$. Наиболее часто приходится иметь дело с РФ $f[k]$, соответствующими непрерывным функциям $f(t)$ непрерывного аргумента. Непрерывная функция, совпадающая с дискретами решетчатой функции, называется огибающей этой РФ. Каждая непрерывная функция $f(t)$ может служить огибающей различных РФ $f_i[k] = f(T_i k)$, отличающихся параметром T_i - периодом дискретизации функции $f(t)$.

Различным математическим формам характеризующим, или представляющим функцию $f(t)$ можно поставить в соответствие аналоги, определяющие РФ $f[k]$.

Аналогами производных $f'(t)$, $f''(t)$, и т.д. являются первая, вторая и т.д. правые и левые разности РФ $f[k]$:

$$\begin{aligned} \Delta f_k &= f_{k+1} - f_k; & \nabla f_k &= f_k - f_{k-1}; \\ \Delta^2 f_k &= \Delta f_{k+1} - \Delta f_k; & \nabla^2 f_k &= \nabla f_k - \nabla f_{k-1} \end{aligned} \quad (4.8) \quad (4.9)$$

Поскольку аналогично (4.8,4.9) получаем:

$$\begin{aligned} \Delta f_{k+1} &= f_{k+2} - f_{k+1} & \text{и} & & \nabla f_{k-1} &= f_{k-1} - f_{k-2}, & \text{то} \\ \Delta^2 f_k &= f_{k+2} - 2f_{k+1} + f_k & (4.10) & & \nabla f_{k-1} &= f_{k-1} - f_{k-2} & \nabla^2 f_k &= f_k - 2f_{k-1} + f_{k-2} & (4.11) \end{aligned}$$

Для квадратичной РФ $f_k = k^2$ правые разности

$$\begin{aligned} \Delta f_k &= (k+1)^2 - k^2; & \Delta^2 f_k &= (k+2)^2 - 2(k+1)^2 + k^2 = 2; & \Delta^3 f_k &= 2 - 2 = 0, \\ \Delta^4 f_k &= 0 - 0 = 0 & \text{и} & & \text{левые} & & \text{разности} & & \nabla f_k &= k^2 - (k-1)^2 = 2k - 1; \\ \nabla^2 f_k &= k^2 - 2(k-1)^2 + (k-2)^2 = 2; & \Delta^3 f_k &= 2 - 2 = 0, & \nabla^4 f_k &= 0. \end{aligned}$$

Из примера следует, что если разность n -го порядка является постоянной величиной, то разности более высоких порядков равны нулю.

Для правых разностей (4.8) аналогично (4.10) получаем:

$$\Delta^3 f_k = f_{k+3} - 3f_{k+2} + 3f_{k+1} - f_k \quad (4.12)$$

$$\Delta^4 f_k = f_{k+4} - 4f_{k+3} + 6f_{k+2} - 4f_{k+1} + f_k \quad (4.13)$$

$$\Delta^n f_k = f_{k+n} - C_1 f_{k+n-1} + C_2 f_{k+n-2} - \dots + (-1)^2 C_r f_{k+n-r} + \dots + (-1)^n C_n f_k,$$

где C_r - биномиальные коэффициенты.

Таким образом, правая разность n -го порядка является линейной функцией настоящего f_k и n будущих $f_{k+1}, f_{k+2}, \dots, f_{k+n}$ значений РФ.

Для левых разностей (4.9) аналогично (4.11) получаем

$$\nabla^3 f_k = f_k - 3f_{k-1} + 3f_{k-2} - f_{k-3}, \quad (4.14)$$

$$\nabla^4 f_k = f_k - 4f_{k-1} + 6f_{k-2} - 4f_{k-3} + f_{k-4}, \quad (4.15)$$

$$\nabla^n f_k = f_k - C_1 f_{k-1} + C_2 f_{k-2} - \dots + (-1)^2 C_r f_{k-r} + \dots + (-1)^n C_n f_{k-n}.$$

Откуда левая разность n -го порядка является линейной функцией настоящего f_k и n прошлых $f_{k-1}, f_{k-2}, \dots, f_{k-n}$ значений РФ.

Из (4.15) следует, что если $f[k < 0] = 0$, то при любом n

$$\nabla^n f[0] = f[0]. \quad (4.16)$$

Дифференциальное уравнение представляет собой неявную зависимость между двумя непрерывными функциями $y(t)$ и $x(t)$, которую можно с помощью РФ $y[k], x[k]$ описать разностным уравнением, как правым, так и левым.

Разностные и рекуррентные уравнения. Простейшее линейное ДУ первого порядка имеет вид:

$$y'(t) + ay(t) = bx(t) \quad (4.17)$$

Его аналогом являются линейные разностные уравнения первого порядка правое и левое соответственно.

$$\Delta y_k + ay_k = bx_k, \quad \nabla y_k + ay_k = bx_k. \quad (4.18)$$

Для конкретизации решений этих уравнений задаются начальные условия $y(0)$ и $y[0]$.

Если в уравнении (4.18) раскрыть первые разности, т.е. подставить $\Delta y_k = y_{k+1} - y_k$, $\nabla y_k = y_k - y_{k-1}$, то получим рекуррентные уравнения

$$\begin{aligned} y_{k+1} + (a-1)y_k &= bx_k; \\ (a+1)y_k - y_{k-1} &= bx_k. \end{aligned} \quad (4.19)$$

Из уравнений (4.19) находят рекуррентные соотношения

$$y_{k+1} = bx_k + (1-a)y_k, \quad y_k = \frac{bx_k - y_{k-1}}{a-1},$$

с помощью которого можно определить вид зависимости, задаваясь начальными условиями и входным воздействием.

Уравнения (4.19) называют правыми и левыми рекуррентными уравнениями.

Правому разностному уравнению общего вида

$$\Delta^n y_k + a_{n-1} \Delta^{n-1} y_k + \dots + a_0 y_k = b_m \Delta^m x_k + \dots + b_0 x_k \quad (4.20)$$

с правыми начальными условиями

$$y[0], \Delta y[0], \dots, \Delta^{n-1} y[0] \quad (4.21)$$

соответствует правое рекуррентное уравнение

$$y_{k+n} + \alpha_{n-1} y_{k+n-1} + \dots + \alpha_0 y_k = \beta_m x_{k+m} + \beta_{m-1} x_{k+m-1} + \dots + \beta_0 x_k \quad (4.22)$$

с правыми начальными условиями

$$y[0], y[1], \dots, y[n-1]. \quad (4.23)$$

Начальные условия (4.21) легко пересчитываются в (4.23) и наоборот.

Например, для правого разностного и рекуррентного уравнений 4-го порядка имеем начальные условия

$$y_0, \Delta y_0, \Delta^2 y_0, \Delta^3 y_0, y_0, y_1, y_2, y_3.$$

Согласно (8.8), (8.10), (8.12) при $k=0$ имеем

$$\Delta y_0 = y_1 - y_0; \quad \Delta^2 y_0 = y_2 - 2y_1 + y_0; \quad \Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0.$$

Откуда легко находятся

$$y_1 = \Delta y_0 + y_0; \quad y_2 = \Delta^2 y_0 + 2\Delta y_0 + y_0; \quad y_3 = \Delta^3 y_0 + 3\Delta^2 y_0 + 3\Delta y_0 + y_0.$$

Левому разностному уравнению общего вида

$$\nabla^n y_k + a_{n-1} \nabla^{n-1} y_k + \dots + a_0 y_k = b_m \nabla^m x_k + b_{m-1} \nabla^{m-1} x_k + \dots + b_0 x_k \quad (4.24)$$

с левыми начальными условиями

$$y_0, \nabla y_0, \nabla^2 y_0, \dots, \nabla^{n-1} y_0 \quad (4.25)$$

соответствует левое рекуррентное уравнение

$$y_k - \alpha_{n-1} y_{k-1} + \dots + \alpha_0 y_{k-n} = \beta_m x_k + \beta_{m-1} x_{k-1} + \dots + \beta_0 x_{k-m} \quad (4.26)$$

с левыми начальными условиями

$$y[-1], y[-2], \dots, y[-n]. \quad (4.27)$$

Например, для левых уравнений 4-го порядка

$$\begin{aligned} \nabla^4 y_k + a_3 \nabla^3 y_k + \dots + a_0 y_k &= x_k, \\ y_k + \alpha_3 y_{k-1} + \alpha_2 y_{k-2} + \alpha_1 y_{k-3} + \alpha_0 y_{k-4} &= x_k \end{aligned}$$

левыми начальными условиями будут следующие

$$y_0, \nabla y_0, \nabla^2 y_0, \nabla^3 y_0; y_{-1}, y_{-2}, y_{-3}, y_{-4}.$$

Согласно (4.9), (4.11), (4.14) при $k=0$

$$\nabla y_0 = y_0 - y_{-1}, \quad \nabla^2 y_0 = y_0 - 2y_{-1} + 2y_{-2}, \quad \nabla^3 y_0 = y_0 - 3y_{-1} + 3y_{-2} - y_{-3}.$$

Значения y_0 находим из рекуррентного уравнения при $k=0$, а остальные получаем из последней системы

$$y_{-1} = y_0 - \nabla y_0; \quad y_{-2} = y_0 - 2\nabla y_0 + \nabla^2 y_0;$$

$$y_{-3} = y_0 - 3\nabla y_0 + 3\nabla^2 y_0 - \nabla^3 y_0;$$

$$y_{-4} = y_0 - 4\nabla y_0 + 6\nabla^2 y_0 - 4\nabla^3 y_0 + \nabla^4 y_0.$$

Значения $\nabla^4 y_0$ определяется из разностного уравнения при $k=0$.

Полученные результаты по аналогии распространяются на уравнения более высоких порядков.

Из (4.26) легко получить явную рекуррентную формулу, не прибегая к классической методике решения разностных уравнений.

Пример: Задано левое разностное уравнение, неоднородное второго порядка $3\nabla^2 y_k + 19\nabla y_k + 28y_k = 3 \cdot 0,5^k$, причем $y[k < 0] = 0$. Преобразуем это уравнение в рекуррентное $y_k - 0,5y_{k-1} + 0,06y_{k-2} = 0,06 \cdot 0,5^k$.

Так как $y[k < 0] = 0$, то $y[-1] = y[-2] = 0$ и согласно (8.16) $y_0 = \nabla y_0 = \nabla^2 y_0$. С учетом этого и разностного и рекуррентного уравнений находим $y_0 = \nabla y_0 = \nabla^2 y_0 = 0,06$.

Из рекуррентного уравнения находим рекуррентную формулу

$$y_k = 0,5y_{k-1} - 0,06y_{k-2} + 0,06 \cdot 0,5^k$$

и при $k=0; 1; 2; 3; \dots$ получаем значения $y_0 = 0,06$, $y_1 = 0,06$; $y_2 = 0,0414$, $y_3 = 0,0246 \dots$

Таким образом, алгоритмизация ДУ сводится к преобразованию его сначала в разностное уравнение, а затем в рекуррентное уравнение и рекуррентную формулу, которую можно легко превратить в алгоритмическую модель, реализуемую в виде программы на ЦВМ.

4.2.2 Алгоритмизация НДМ на основе z - преобразований

Понятие о z – преобразовании Лапласа. Процесс получения явных рекуррентных соотношений существенно облегчается, если для этого используется дискретное преобразование РФ и, в частности его аналог z – преобразование.

Под z – преобразованием РФ $f[k]$ понимают непрерывную функцию

$$f(z) = z\{f[k]\} = \sum_{k=0}^{\infty} f[k]z^{-k} \quad (4.28)$$

комплексной переменной z при условии, что определяющий ее бесконечный ряд сходится. Соответствие оригинала $f[k]$ и его z -изображения $f(z)$ символически записывают в виде $f[k] \div f(z)$.

К основным свойствам z -изображений РФ относятся следующие:

- а) $cf[k] \leftrightarrow cf(z)$, где $c = const$;
- б) $f_1[k] + f_2[k] \leftrightarrow f_1(z) + f_2(z)$;
- в) запаздывание оригинала $f[k-n] \div f(z)$ изображается

$$z\{f[k-n]\} = z^{-n}f(z), \text{ при } f[k < 0] = 0;$$
- г) $\exp(\pm\lambda k)f[k] \div f(z \cdot \exp(\pm\lambda))$.
- д) изображение правых разностей

$$z\{\Delta^n f[k]\} = (z-1)^n f(z), \text{ при нулевых НУ};$$
- е) изображение левых разностей

$$z\{\nabla^n f[k]\} = \left(\frac{z-1}{z}\right)^n f(z), \text{ при нулевых НУ};$$
- ж) изображение сумм

$$s(z) = z\left\{\sum_{\phi=0}^k f[\phi]\right\} = \frac{f(z)}{z-1};$$

$$S(z) = z\left\{\sum_{\phi=0}^k f[\phi]\right\} = \frac{zf[z]}{z-1}.$$

Конечное и начальное значения оригинала

$$f[\infty] = \lim_{k \rightarrow \infty} f[k] = \lim_{z \rightarrow 1} (z-1)f(z);$$

$$f[0] = \lim_{k \rightarrow 0} f[k] = \lim_{z \rightarrow \infty} f(z).$$

Примерами соответствий $f(p) \div f(t) \div f[k] \div f(z)$ являются следующие:

$$\frac{1}{p} \div 1(t) \div 1[k] \div \frac{z}{z-1}; \quad \frac{1}{p^2} \div t \div T \cdot k \div \frac{Tz}{(z-1)^2};$$

$$\frac{1}{p+\alpha} \div e^{-\alpha t} \div e^{-\alpha T k} \div \frac{z}{z-a}, a = e^{-\alpha T};$$

$$\frac{1}{(p+\alpha)^2} \div t \cdot e^{-\alpha t} \div T k e^{-\alpha T k} \div \frac{Taz}{(z-a)^2} \text{ и др.}$$

Они сведены в таблицу соответствий, которую обычно включают в учебники и учебные пособия по теории автоматического управления.

Обратное z -преобразование, т.е. определение оригинала по его изображению, выполняется четырьмя способами. Два первых из них аналогичны обратному преобразованию Лапласа для непрерывных функций, два последних аналогов не имеют.

Способы нахождения оригинала по его z -изображению.

Первый способ заключается в непосредственном применении таблицы соответствий оригиналов и z -изображений. Если в таблице имеется выражение аналогичное $f(z)$, то соответствующая РФ $f[k]$ определяется немедленно.

Пример 1. Дано $f(z) = z(z-0.1)$. Определить $f[k]$. По таблице находим $\frac{z}{z-a} \div a^k$, откуда $f(z) = \frac{z}{z-0.1} \div 0.1^k = f[k]$.

Второй способ. Применение таблицы соответствий $f[k] \div f(z)$ после разложения $f(z)$ на простейшие (табличные) составляющие

$$f(z) = \sum_{i=1}^v f_i(z). \quad (4.29)$$

После определения $f_i[k] \div f_i(z)$ находят

$$f[k] = \sum_{i=1}^v i_i[k]. \quad (4.30)$$

Обычно z -изображения РФ имеют вид

$$f(z) = z \cdot \frac{b_0 + b_1 z + \dots + b_m z^m}{a_0 + a_1 z + \dots + a_n z^n}, \text{ где } m < n. \quad (4.31)$$

Поэтому практически раскладывают на составляющие функцию

$$f(z) = z \cdot \frac{b_0 + b_1 z + \dots + b_m z^m}{a_0 + a_1 z + \dots + a_n z^n} = \frac{b(z)}{a(z)}. \quad (4.32)$$

Эта операция реализуется после решения алгебраического уравнения n -й степени $a(z) = 0$ и определения его корней z_1, z_2, \dots, z_n . После чего выполняется разложение

$$\varphi(z) = \sum_{i=1}^v \varphi_i(z) \quad (4.33)$$

Каждой составляющей $\varphi_i(z)$ соответствует составляющая $f_i(z) = z\varphi_i(z)$, которая должна быть приведена к табличному виду.

Пример 2. Дано $f(z) = \frac{0.3z}{z^2 - 1.8z + 0.8}$.

$$\varphi(z) = \frac{b(z)}{a(z)} = \frac{a}{z-1} + \frac{b}{z-0.8} = \varphi_1(z) + \varphi_2(z).$$

Так как $\frac{0.3}{(z-1)(z-0.8)} = \frac{(a+b)z - (0.8a+b)}{(z-1)(z-0.8)}$, то значения коэффициентов опре-

деляют из системы уравнений $a+b=0$, $-(0.8a+b)=0.3$ и равны $a=1.5$, $b=-1.5$.

Таким образом $\varphi_1(z) = \frac{1.5}{(z-1)}$; $\varphi_2(z) = \frac{-1.5}{(z-0.8)}$. Следовательно

$$f_1(z) = z\varphi_1(z) = \frac{1.5z}{z-1} \div 1.5 = f_1[k]; \quad f_2(z) = z\varphi_2(z) = \frac{-1.5z}{z-0.8} \div 1.5 \cdot 0.8^k = f_2[k]. \quad \text{Откуда}$$

$$f[k] = f_1[k] + f_2[k] = 1.5(1 - 0.8^k). \quad \text{При } k = 0, 1, 2, 3, \dots \quad f_0 = 0; \quad f_1 = 0.3; \quad f_2 = 0.54; \\ f_3 = 0.732; \dots$$

Третий способ. Разложение z -изображения по степеням z^{-1} . По определению

$$f(z) = f[0]z^0 + f[1]z^{-1} + \dots + f[k]z^{-k} \quad (4.34)$$

Поэтому при разложении (4.31) в ряд (4.34) коэффициенты при z^{-k} равны $f[k]$, т.е. являются значениями РФ оригинала при $k = 0, 1, 2, 3, \dots$. Выражение (4.31) представляют виде $f(z) = z^r B(z) / a(z) = z^r \varphi(z)$, так чтобы степени полиномов $B(z)$ и $a(z)$ были равны ($m = n$). Тогда делением $B(z)$ на $a(z)$ получают ряд (4.34) для $\varphi(z)$.

$$\varphi(z) = \varphi[0]z^0 + \varphi[1]z^{-1} + \varphi[2]z^{-2} + \dots \quad (4.35)$$

Переход от $\varphi(z) \div \varphi[k]$ к $f(z) \div f[k]$ осуществляют на основании соотношений

$$f(z) = z^r \varphi(z); \quad f[k] = \varphi[k+r]. \quad (4.36)$$

Пример 3. Дано как в примере 2

$$f(z) = z^{-1} \frac{0.3z^2}{z^2 - 1.8z + 0.8} = z^{-1} \frac{B(z)}{a(z)}.$$

Делением $B(z)$ на $a(z)$ получают $\varphi(z) = 0.3 + 0.54z^{-1} + 0.732z^{-2} + \dots$. Согласно (4.36) $f(z)z^{-1}\varphi(z) = 0z^0 + 0.3z^{-1} + 0.54z^{-2} + 0.732z^{-3} + \dots$. Откуда $f[0] = 0$, $f[1] = 0.3$, $f[2] = 0.54$, $f[3] = 0.732$,

Четвертый способ. Выражение (4.31) представляют виде

$$f(z) = \frac{\beta_0 + \beta_1 z^{-1} + \dots + \beta_n z^{-n}}{1 + \alpha_1 z^{-1} + \dots + \alpha_n z^{-n}}. \quad (4.37)$$

Если принять для простоты $n = 3$, то

$$f(z) = \frac{\beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \beta_3 z^{-3}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \alpha_3 z^{-3}}.$$

Отсюда с учетом (8.34) имеем

$$\beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \beta_3 z^{-3} = (1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \alpha_3 z^{-3})(f[0] + f[1]z^{-1} + f[2]z^{-2} + f[3]z^{-3} + \dots).$$

Откуда, раскрыв скобки и приравняв коэффициенты при одинаковых степенях z в правой и левой частях, находят рекуррентные соотношения для $f_0 = \beta_0$, $f_1 = \beta_1 - \alpha_1 f_0$, $f_2 = \beta_2 - \alpha_1 f_1 + \alpha_2 f_0$, и т.д. Эти соотношения легко распространить на общий случай (4.37).

Эффективным способом получения рекуррентных соотношений является использование дискретной передаточной функции (ДПФ).

При нулевых НУ операторная форма разностных и рекуррентных уравнений имеет вид

$$A(z)y(z) = B(z)x(z), \quad (4.38)$$

где $A(z)$, $B(z)$ - полиномы z . В этом случае возможно представление этих уравнений в форме ДПФ

$$\frac{y(z)}{x(z)} = \frac{B(z)}{A(z)} = D(z). \quad (4.39)$$

Если $y[k \leq 0] = 0$, то $D[z]$ соответствует правым уравнениям, а если $y[k < 0] = 0$, - левым. Располагая ДПФ, можно легко получить z -изображение неизвестной РФ

$$y(z) = D(z)x(z) \quad (4.40)$$

по z - изображению известной $x(z)$ и затем соответствующее рекуррентное уравнение.

Пример 4. Дана ДПФ $D(z) = \frac{y(z)}{x(z)} = \frac{10z^2 + z}{z^2 - 0.4z + 0.03}$, при чем $y[k < 0] = 0$.

Требуется найти рекуррентное уравнение.

Для решения преобразуем ДПФ в операторное уравнение:

$$z^2 y(z) - 0.4zy(z) + 0.03y(z) = 10z^2 x(z) + zx(z).$$

После деления на z^2 оно имеет вид:

$$y(z) - 0.4z^{-1}y(z) + 0.03z^{-2}y(z) = 10x(z) + z^{-1}x(z).$$

Откуда левое рекуррентное уравнение будет иметь вид:

$$y[k] - 0.4y[k-1] + 0.03y[k-2] = 10x[k] + x[k-1].$$

Таким образом, для алгоритмизации математического описания объекта, полученного в форме передаточной функции, необходимо по ней определить дискретную передаточную функцию, преобразовать ее в операторное уравнение, разделив последнее на z^n , где n - старшая степень полинома знаменателя ДПФ, и, используя свойства z -преобразования, найти оригинал операторного уравнения, т.е. рекуррентное уравнение.

4.2.3 Особенности алгоритмизации моделей систем автоматического управления

Алгоритмизация дискретных систем. Моделирование САУ на ЦВМ заключается в алгоритмизации и численном решении уравнений, описывающих поведение этих систем при различных воздействиях. Наиболее просто осуществляется моделирование дискретных систем. Дискретная линейная стационарная система с одним входом и одним выходом описывается левым разностным уравнением с постоянными коэффициентами и нулевыми НУ. Алгоритмизация его сводится к приданию ему рекуррентной формы, из которой следует рекуррентное соотношение реакции системы $y[k]$ на входное воздействие $x[k]$

$$y[k] = \beta_m x_k + \beta_{m-1} x_{k-1} + \dots + \beta_0 x_{k-m} - \alpha_{n-1} y_{k-1} - \dots - \alpha_0 y_{k-n}. \quad (4.41)$$

Это выражение справедливо и для нестационарных систем, когда α и β зависят от $k, k-1, k-2, \dots$.

Пусть $x[k], y[k]$ - реальные значения входного воздействия и реакции дискретной системы, соответствующие моментам реального времени $t = 0, T, 2T, \dots, kT, \dots$; $x_m[k], y_m[k]$ - машинные значения РФ $x[k], y[k]$, реализованные ЦВМ; T_m - машинное время квантования. Пренебрегая эффектом квантования по уровню, имеем при $x_m[k] = x[k]$ реакцию $y_m[k] = y[k]$.

При этом масштаб времени

$$m_t = \frac{t}{t_m} = \frac{Tk}{T_m k} = \frac{T}{T_m}. \quad (4.42)$$

Если T_m является функцией k , то m_t - переменная величина. Моделирование САУ на ЦВМ выполняется, как правило, при $m_t > 1$.

Алгоритмизация непрерывных систем. Моделирование непрерывной системы требует получения на ЭВМ реакции некоторой эквивалентной дискретной системы $y[k]$ на входное воздействие $x[k]$ при условии, что реакция $y(t)$ и входное воздействие $x(t)$ непрерывной системы являются огибающими РФ $y[k], x[k]$.

Моделирование непрерывной линейной стационарной САУ при заданном входном воздействии сводится к решению неоднородного линейного ДУ с постоянными коэффициентами, решение которого можно найти ранее изложенными методами.

При различных входных воздействиях решение выполняется либо на основе аппроксимации операции интегрирования, либо с применением метода Рунге-Кутты.

Один из простейших методов аппроксимации операции интегрирования заключается в замене кривой непрерывной функции $x(t)$ ломаной, проведенной через дискреты РФ $x[k] = x[Tk]$ (рисунок 4.2).

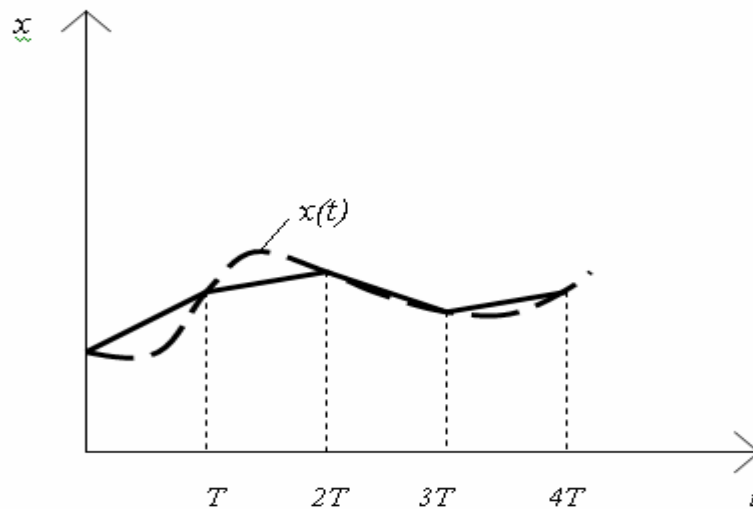


Рисунок 4.2 – Кусочно – линейная аппроксимация непрерывной функции $x(t)$

С учетом этого, приближенные равенства

$$y(T) = \int_0^T x(t)dt \approx \tilde{y}[0] + T(x[1] + x[0]) / 2,$$

$$y(2T) = \int_0^{2T} x(t)dt \approx \tilde{y}[1] + T(x[2] + x[1]) / 2,$$

.....
Тогда непрерывную функцию

$$y[t] = \int_0^t x(t)dt \tag{4.43}$$

приближенно можно выразить РФ $\tilde{y}[k] \approx y[k] = y[Tk]$, значения которой определяются рекуррентной формулой

$$\tilde{y}[k] = \tilde{y}[k-1] + (T/2)(x[k] + x[k-1]) \tag{4.44}$$

Изображение по Лапласу выражения (4.43) при нулевых НУ

$$y(p) = (1/p)x(p). \tag{4.45}$$

Для определения соотношения между z -изображениями $x(z)$, $\tilde{y}(z)$ функций $x[k]$, $\tilde{y}[k]$ подвергается z -преобразованию выражение (4.44) и получаем

$$\begin{aligned}\tilde{y}(z) &= z^{-1}\tilde{y}(z) + (T/2)[x(z) + z^{-1}x(z)], \text{ или} \\ \tilde{y}(z) &= \frac{T(z+1)}{2(z-1)}x(z).\end{aligned}\quad (4.46)$$

Следовательно, имеет место соответствие

$$y(p) = \frac{1}{p}x(p) \div \frac{T(z+1)}{2(z+1)}x(z) = \tilde{y}(z), \quad (4.47)$$

причем $x(z) \div x[k] = x[Tk]$, $\tilde{y}(z) \div \tilde{y}[k] = y[Tk]$.

В случае n -кратного интегрирования аналогично имеем

$$y(p) = \frac{1}{p^n}x(p) \div \left[\frac{T(z+1)}{2(z-1)} \right]^n x(z) = \tilde{y}(z). \quad (4.48)$$

Система 1-го порядка описывается передаточной функцией

$$W(p) = \frac{y(p)}{x(p)} = \frac{b_0}{p + a_0} \quad (4.49)$$

или уравнением

$$y(p) + a_0 \frac{1}{p}y(p) = b_0 \frac{1}{p}x(p). \quad (4.50)$$

На основании соответствия (4.47) получаем уравнение

$$\tilde{y}(z) + a_0 \frac{T(z+1)}{2(z-1)}\tilde{y}(z) = b_0 \frac{T(z+1)}{2(z-1)}x(z)$$

и соответствующую ему дискретную передаточную функцию

$$D(z) = \frac{\tilde{y}(z)}{x(z)} = \frac{\beta_1 z + \beta_0}{z + \alpha_0}, \quad (4.51)$$

где $\alpha_0 = \frac{a_0 T - 2}{a_0 T + 2}$; $\beta_1 = \beta_2 = \frac{b_0 T}{a_0 T + 2}$.

При любой форме $x(t) \div x(p)$ и нулевых НУ из (4.51) получают уравнение

$$\tilde{y}(z) + \alpha_0 z^{-1}\tilde{y}(z) = \beta_1 x(z) + \beta_0 z^{-1}x(z)$$

и затем рекуррентное соотношение для воспроизведения реакции системы

$$y(Tk) \approx \tilde{y}[k] = \beta_1 x[k] + \beta_0 x[k-1] - \alpha_0 \tilde{y}[k-1].$$

Вопросы для самоконтроля

- 1 Что называют идентификацией динамической системы, и какие задачи для ее осуществления необходимо решить?
- 2 Какова цель структурной идентификации динамического объекта?
- 3 К чему сводится параметрическая идентификация динамической системы;
- 4 Каким образом производится проверка адекватности модели исследуемой системе?
- 5 В чем сущность алгоритмизации математического описания динамической системы?
- 6 Какую функцию называют решетчатой, и для описания каких процессов ее применяют?
- 7 Как осуществляется преобразование дифференциального уравнения в рекуррентное соотношение?
- 8 Объясните, почему с помощью z – преобразования легче отыскать рекуррентное соотношение?
- 9 Перечислите способы нахождения решетчатой функции по ее z - изображению?
- 10 Как получить рекуррентное соотношение по математической модели объекта в форме передаточной функции?

5 Моделирование систем массового обслуживания

5.1 Аналитические модели систем массового обслуживания

5.1.1 Поток событий

Потоком событий называется последовательность однородных событий, появляющихся одно за другим в случайные моменты времени. Примеры: поток вызовов на телефонной станции; поток сбоев ЭВМ; поток заявок на проведение расчетов в вычислительном центре и т.п.

Поток событий наглядно изображается рядом точек с абсциссами $Q_1, Q_2, \dots, Q_n, \dots$ (рис. 5.1) с интервалами между ними: $T_1 = Q_2 - Q_1, T_2 = Q_3 - Q_2, \dots, T_n = Q_{n+1} - Q_n$. При его вероятностном описании поток событий может быть представлен как последовательность случайных величин: $Q_1; Q_2 = Q_1 + T_1; Q_3 = Q_1 + T_1 + T_2$; и т.д. На рисунке в виде ряда точек изображен не сам поток событий (он случаен), а только одна его конкретная реализация.

Ранее упоминалось о потоках событий и некоторых их свойствах; здесь рассмотрим их более подробно. Поток событий называется *стационарным*, если его вероятностные характеристики не зависят от выбора начала отсчета или, более конкретно, если вероятность попадания того или другого числа событий на любой интервал времени зависит только от длины этого интервала и не зависит от того, где именно на оси $\theta-t$ он расположен.

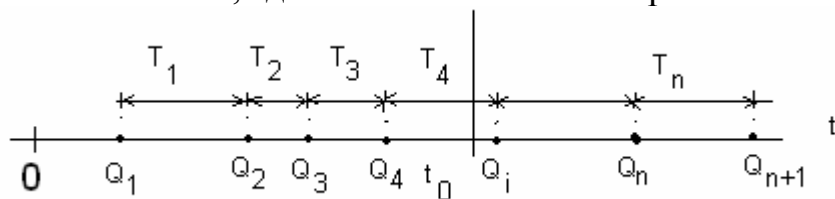


Рисунок 5.1 – Реализация потока событий

Поток событий называется *ординарным*, если вероятность попадания на элементарный интервал времени Δt двух или более событий пренебрежимо мала по сравнению с вероятностью попадания одного события.

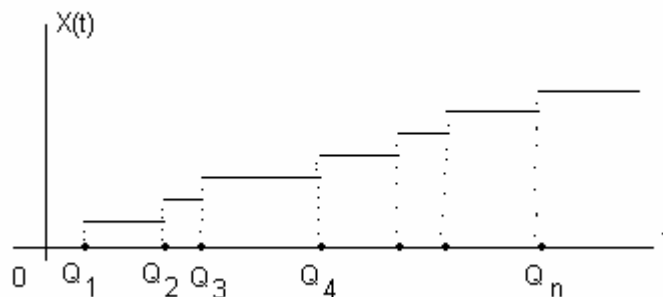


Рисунок 5.2 – Поток событий как случайный процесс

Ординарный поток событий можно интерпретировать как случайный процесс $X(t)$ - число событий, появившихся до момента t (рис. 5.2). Случайный процесс $X(t)$ скачкообразно возрастает на одну единицу в точках Q_1, Q_2, \dots, Q_n .

Поток событий называется *потоком без последствия*, если число событий, попадающих на любой интервал времени τ , не зависит от того, сколько событий попало на любой другой не пересекающийся с ним интервал. Практически отсутствие последствия в потоке означает, что события, образующие поток, появляются в те или другие моменты времени независимо друг от друга.

Поток событий называется *простейшим*, если он стационарен, ординарен и не имеет последствия. Интервал времени T между двумя соседними событиями простейшего потока имеет показательное распределение

$$f(t) = \lambda e^{-\lambda t} \quad (\text{при } t > 0); \quad (5.1)$$

где $\lambda = 1/M [T]$ - величина, обратная среднему значению интервала T .

Ординарный поток событий без последствия называется *пуассоновским*. Простейший поток является частным случаем стационарного пуассоновского потока. *Интенсивностью* λ потока событий называется среднее число событий, приходящееся на единицу времени. Для стационарного потока $\lambda = const$; для нестационарного потока она в общем случае зависит от времени: $\lambda = \lambda(t)$.

Мгновенная интенсивность потока $\lambda(t)$ определяется как предел отношения среднего числа событий, которые произошли за элементарный интервал времени $(t, t + \Delta t)$, к длине Δt этого интервала, когда она стремится к нулю. Среднее число событий, наступающих на интервале времени τ , следующем непосредственно за моментом t_0 (см. рис. 5.1), равно

$$a(t_0, \tau) = \int_{t_0}^{t_0 + \tau} \lambda(t) dt.$$

Если поток событий стационарный, то $a(t_0, \tau) = a(\tau) = \lambda \tau$.

Ординарный поток событий называется *потоком Пальма* (*рекуррентным* потоком, или потоком с ограниченным последствием), если интервалы времени T_1, T_2, \dots между событиями (см. рис. 5.1) представляют собой независимые, одинаково распределенные случайные величины. В связи с одинаковостью распределений T_1, T_2, \dots поток Пальма всегда стационарен. Простейший поток является частным случаем потока Пальма; в нем интервалы между событиями распределены по показательному закону (5.1), где λ - интенсивность потока.

Потоком Эрланга k -го порядка называется поток событий, получающийся «прореживанием» простейшего потока, когда сохраняется каждая k -я точка (событие) в потоке, а все промежуточные выбрасываются (на рис. 5.3 показано получение потока Эрланга 4-го порядка из простейшего потока).

Интервал времени между двумя соседними событиями в потоке Эрланга k -го порядка представляет собой сумму k независимых случайных величин T_1, T_2, \dots, T_k имеющих показательное распределение с параметром λ :

$$T = \sum_{i=1}^k T_i. \quad (5.2)$$

Закон распределения случайной величины T называется *законом Эрланга k -го порядка* и имеет плотность

$$f_k(t) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} \quad (\text{при } t > 0). \quad (5.3)$$

Математическое ожидание, дисперсия и среднеквадратичное отклонение случайной величины T (5.2) соответственно равны:

$$m_t = k / \lambda; \quad D_t = k / \lambda^2; \quad \sigma_t = \sqrt{k / \lambda}. \quad (5.4)$$

Коэффициент вариации случайной величины (5.2) равен

$$U_t = \sigma_t / m_t = 1 / \sqrt{k}. \quad (5.5)$$

При увеличении порядка потока Эрланга «степень случайности» интервала между событиями стремится к нулю.

Если одновременно с «прореживанием» простейшего потока изменять масштаб по оси θ - t (делением на k), получится *нормированный* поток Эрланга k -го порядка, интенсивность которого не зависит от k .

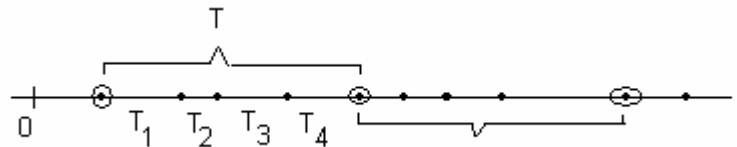


Рис.5.3

Числовые характеристики случайной величины в нормированном потоке Эрланга k -го порядка равны:

$$M|\bar{T}| = 1 / \lambda; \quad D|\bar{T}| = 1 / k\lambda^2; \quad \bar{\sigma}_t = 1 / (\lambda\sqrt{k}); \quad u_t = 1 / \sqrt{k}.$$

При увеличении k нормированный поток Эрланга неограниченно приближается к *регулярному потоку* с постоянным интервалом $I = 1 / \lambda$ между событиями.

5.1.2 Марковские случайные процессы

Случайный процесс называют *марковским*, если он обладает следующим свойством: для любого момента времени t_0 вероятность любого состояния системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, каким образом система пришла в это состояние.

В данной главе будем рассматривать только марковские процессы с дискретными состояниями S_1, S_2, \dots, S_n . Такие процессы удобно иллюстрировать с помощью графа состояний (рис. 5.4), где прямоугольниками (или кружками) обозначены состояния S_1, S_2, \dots системы S , а стрелками —

возможные переходы из состояния в состояние (на графе отмечаются только непосредственные переходы, а не переходы через другие состояния).

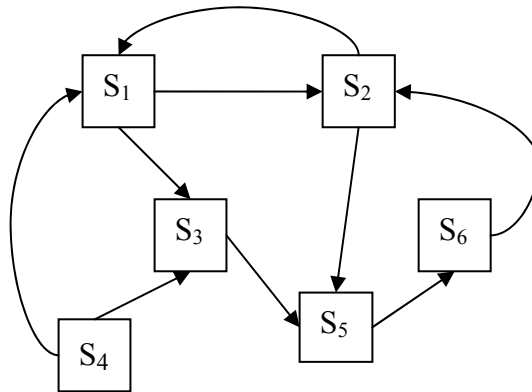


Рисунок 5.4 – Граф состояний случайного процесса

Иногда на графе состояний отмечают не только возможные переходы из состояния в состояние, но и возможные задержки в прежнем состоянии; это изображается стрелкой («петлей»), направленной из данного состояния в него же, но можно обходиться и без этого. Число состояний системы может быть как конечным, так и бесконечным (но счетным).

Марковский случайный процесс с дискретными состояниями и дискретным временем обычно называют *марковской цепью*. Для такого процесса моменты t_1, t_2, \dots , когда система S может менять свое состояние, удобно рассматривать как последовательные шаги процесса, а в качестве аргумента, от которого зависит процесс, рассматривать не время t , а номер шага: $1, 2, \dots, k, \dots$. Случайный процесс в этом случае характеризуется последовательностью состояний

$$S(0), S(1), S(2), \dots, S(k), \quad (5.6)$$

если $S(0)$ — начальное состояние системы (перед первым шагом); $S(1)$ — состояние системы непосредственно после первого шага; ...; $S(k)$ — состояние системы непосредственно после k -го шага

Событие $S_i, (i=1, 2, \dots)$ является случайным событием, поэтому последовательность состояний (5.6) можно рассматривать как последовательность случайных событий. Начальное состояние $S(0)$ может быть как заданным заранее, так и случайным. О событиях последовательности (5.6) говорят, что они образуют марковскую цепь.

Рассмотрим процесс с n возможными состояниями S_1, S_2, \dots, S_n . Если обозначить через $X(t)$ номер состояния, в котором находится система S в момент t , то процесс описывается целочисленной случайной функцией $X(t) > 0$, возможные значения которой равны $1, 2, \dots, n$. Эта функция совершает скачки от одного целочисленного значения к другому в заданные моменты t_1, t_2, \dots (рис. 5.5) и является непрерывной слева, что отмечено точками на рис. 5.5.

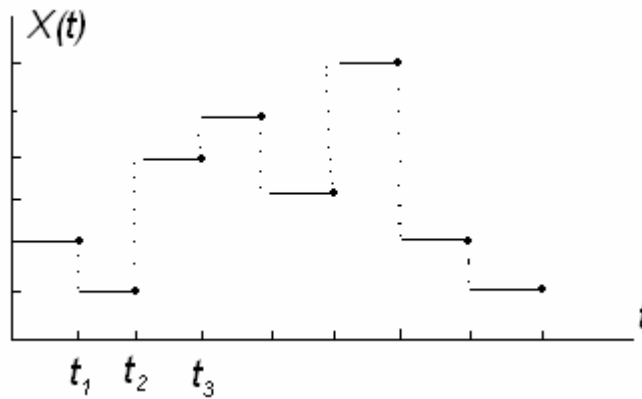


Рисунок 5.5 – График случайного процесса

Рассмотрим одномерный закон распределения случайной функции $X(t)$. Обозначим через $P_i(k)$ вероятность того, что после k -го шага [и до $(k+1)$ -го] система S будет в состоянии S_i ($i=1,2,\dots,n$). Вероятности $p_i(k)$ называются *вероятностями состояний* цепи Маркова. Очевидно, для любого k

$$\sum_{i=1}^n p_i(k) = 1. \quad (5.7)$$

Распределение вероятностей состояний в начале процесса

$$p_1(0), p_2(0), \dots, p_i(0), \dots, p_n(0) \quad (5.8)$$

называется *начальным распределением вероятностей* марковской цепи. В частности, если начальное состояние $S(0)$ системы S в точности известно, например $S(0)=S_i$, то начальная вероятность $P_i(0) = 1$, а все остальные равны нулю.

Вероятностью перехода на k -м шаге из состояния S_i в состояние S_j называется условная вероятность того, что система после k -го шага окажется в состоянии S_j при условии, что непосредственно перед этим (после $k-1$ шагов) она находилась в состоянии S_i . Вероятности перехода иногда называются также «переходными вероятностями».

Марковская цепь называется *однородной*, если переходные вероятности не зависят от номера шага, а зависят только от того, из какого состояния и в какое осуществляется переход:

$$P\{S(k) = S_j \mid S(k-1) = S_i\} = P_{ij} \quad (5.9)$$

Переходные вероятности однородной марковской цепи P_{ij} образуют квадратную таблицу (матрицу) размером $n \times n$:

$$\| P_{ij} \| = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1j} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2j} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{i1} & P_{i2} & \dots & P_{ij} & \dots & P_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nj} & \dots & P_{nn} \end{pmatrix} \quad (5.10)$$

$$\sum_{j=1}^n P_{ij} = 1 (i = 1, \dots, n) . \quad (5.11)$$

Матрицу, обладающую таким свойством, называют *стохастической*. Вероятность P_{ij} есть не что иное, как вероятность того, что система, пришедшая к данному шагу в состояние S_j , в нем же и задержится на очередном шаге.

Если для однородной цепи Маркова заданы начальное распределение вероятностей (5.8) и матрица переходных вероятностей (5.10), то вероятности состояний системы $p_i(k) (i = 1, 2, \dots, n)$ могут быть определены по рекуррентной формуле

$$p_i(k) = \sum_{j=1}^n p_j(k-1) P_{ij} (i = 1, 2, \dots, n) \quad (5.12)$$

Для неоднородной цепи Маркова вероятности перехода в матрице (5.10) и формуле (5.12) зависят от номера шага k .

Для однородной цепи Маркова, если все состояния являются существенными, а число состояний конечно, существует предел

$$\lim_{u \rightarrow \infty} P_i(u) = P_i, \text{ определяемый из системы уравнений } P_i = \sum_{j=1}^n P_j P_{ji} \text{ и } \sum_{i=1}^n P_i = 1.$$

Сумма переходных вероятностей в любой строке матрицы равна единице.

При фактических вычислениях по формуле (5.12) надо в ней учитывать не все состояния S_j , а только те, для которых переходные вероятности отличны от нуля, т.е. те, из которых на графе состояний ведут стрелки в состояние S_i .

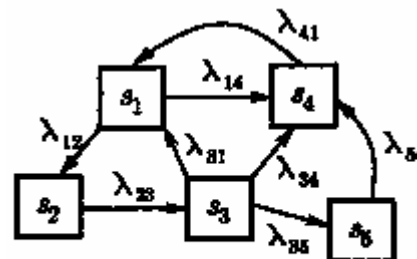


Рисунок 5.6

Марковский случайный процесс с дискретными состояниями и непрерывным временем иногда называют «непрерывной цепью Маркова». Для такого процесса вероятность перехода из состояния S_i в S_j для любого момента времени равна нулю. Вместо вероятности перехода p_{ij} рассматривают *плотность вероятности перехода* λ_{ij} , которая определяется как предел отношения вероятности перехода из состояния S_i в состояние S_j за малый промежуток времени Δt , примыкающий к моменту t , к длине этого промежутка, когда она стремится к нулю. Плотность вероятности перехода может быть как постоянной ($\lambda_{ij} = const$), так и зависящей от времени [$\lambda_{ij} = \lambda_{ij}(t)$]. В первом случае марковский случайный процесс с дискретными состояниями и непрерывным временем называется *однородным*. Типичный пример такого процесса - случайный процесс $X(t)$, представляющий собой число появившихся до момента t событий в простейшем потоке (рис. 5.2).

При рассмотрении случайных процессов с дискретными состояниями и непрерывным временем удобно представлять переходы системы S из состояния в состояние как происходящие под влиянием некоторых потоков

событий. При этом плотности вероятностей перехода получают смысл интенсивностей λ_{ij} соответствующих потоков событий (как только происходит первое событие в потоке с интенсивностью λ_{ij} , система из состояния S_i скачком переходит в S_j). Если все эти потоки пуассоновские, то процесс, протекающий в системе S , будет марковским.

Рассматривая марковские случайные процессы с дискретными состояниями и непрерывным временем, удобно пользоваться графом состояний, на котором против каждой стрелки, ведущей из состояния S_i в S_j , проставлена интенсивность λ_{ij} потока событий, переводящего систему по данной стрелке (рис.5.6). Такой граф состояний называют *размеченным*.

Вероятность того, что система S , находящаяся в состоянии S_i , за элементарный промежуток времени $(t, t + dt)$ перейдет в состояние S_j (элемент вероятности перехода из S_i в S_j), есть вероятность того, что за это время dt появится хотя бы одно событие потока, переводящего систему S из S_i в S_j . С точностью до бесконечно малых высших порядков эта вероятность равна $\lambda_{ij} dt$.

Потоком вероятности перехода из состояния S_i в S_j называется величина $\lambda_{ij} p_i(t)$ (здесь интенсивность λ_{ij} может быть как зависящей, так и независящей от времени).

Рассмотрим случай, когда система S имеет конечное число состояний S_1, S_2, \dots, S_n . Для описания случайного процесса, протекающего в этой системе, применяются вероятности состояний

$$p_1(t), p_2(t), \dots, p_n(t), \quad (5.13)$$

где $p_i(t)$ — вероятность того, что система S в момент t находится в состоянии S_i :

$$p_i(t) = P\{S\{t\} = S_i\}. \quad (5.14)$$

Очевидно, для любого t

$$\sum_{i=1}^n p_i(t) = 1. \quad (5.15)$$

Для нахождения вероятностей (5.13) нужно решить систему дифференциальных уравнений (уравнений Колмогорова), имеющих вид

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^n \lambda_{ij} p_j(t) - p_i(t) \sum_{j=1}^n \lambda_{ij} \quad (i=1, 2, \dots, n),$$

или, опуская аргумент t у переменных p_i ,

$$\frac{dp_i}{dt} = \sum_{j=1}^n \lambda_{ij} p_j - p_i \sum_{j=1}^n \lambda_{ij} \quad (i=1, 2, \dots, n). \quad (5.16)$$

Напомним, что интенсивности потоков λ_{ij} могут зависеть от времени.

Уравнения (5.16) удобно составлять, пользуясь размеченным графом состояний системы и следующим мнемоническим правилом: *производная вероятности каждого состояния равна сумме всех потоков вероятности, переводящих из других состояний в данное, минус сумма всех потоков вероятности, переводящих из данного состояния в другие*. Например, для

системы S, размеченный граф состояний которой дан на рис. 10.6, система уравнений Колмогорова имеет вид

$$\begin{aligned} dp_1 / dt &= \lambda_{31} p_3 + \lambda_{41} p_4 - (\lambda_{12} + \lambda_{14}) p_1; \\ dp_2 / dt &= \lambda_{12} p_1 - \lambda_{23} p_2; \\ dp_3 / dt &= \lambda_{23} p_2 - (\lambda_{33} + \lambda_{34} + \lambda_{35}) p_3; \\ dp_4 / dt &= \lambda_{14} p_1 + \lambda_{34} p_3 + \lambda_{51} p_5 - \lambda_{41} p_4; \\ dp_5 / dt &= \lambda_{35} p_3 - \lambda_{54} p_5. \end{aligned} \quad (5.17)$$

Так как для любого t выполняется условие (5.15), можно любую из вероятностей (5.13) выразить через остальные и таким образом уменьшить число уравнений на одно.

Чтобы решить систему дифференциальных уравнений (5.16) для вероятностей состояний $p_1(t), p_2(t), \dots, p_n(t)$, нужно задать начальное распределение вероятностей

$$p_1(0), p_2(0), \dots, p_i(0), \dots, p_n(0), \quad (5.18)$$

сумма которых равна единице.

Если, в частности, в начальный момент $t = 0$ состояние системы S в точности известно, например, $S(0) = S_i$, и $p_i(0) = 1$, то остальные вероятности выражения (5.18) равны нулю.

Во многих случаях, когда процесс, протекающий в системе, длится достаточно долго, возникает вопрос о предельном поведении вероятностей $p_i(t)$ при $t \rightarrow \infty$. Если все потоки событий, переводящие систему из состояния в состояние, являются простейшими (т.е. стационарными пуассоновскими с постоянными интенсивностями λ_{ij}), в некоторых случаях существуют *финальные* (или предельные) вероятности состояний

$$p_i = \lim_{t \rightarrow \infty} p_i(t) \quad (i = 1, \dots, n), \quad (5.19)$$

независящие от того, в каком состоянии система S находилась в начальный момент. Это означает, что с течением времени в системе S устанавливается *предельный стационарный режим*, в ходе которого она переходит из состояния в состояние, но вероятности состояний уже не меняются. В этом предельном режиме каждая финальная вероятность может быть истолкована как *среднее относительное время* пребывания системы в данном состоянии.

Систему, в которой существуют финальные вероятности, называют *эргодической*. Если система S имеет конечное число состояний S_1, S_2, \dots, S_n , то для существования финальных вероятностей *достаточно*, чтобы из *любого состояния системы можно было* (за какое-то число шагов) *перейти в любое другое*. Если число состояний S_1, S_2, \dots, S_n бесконечно, то это условие перестает быть достаточным, и существование финальных вероятностей зависит не только от графа состояний, но и от интенсивностей λ_{ij} .

Финальные вероятности состояний (если они существуют) могут быть получены решением *системы линейных алгебраических уравнений*, они получаются из дифференциальных уравнений Колмогорова, если положить в

них левые части (производные) равными нулю. Однако удобнее составлять эти уравнения непосредственно по графу состояний, пользуясь мнемоническим правилом: для каждого состояния суммарный выходящий поток вероятности равен суммарному входящему. Например, для системы S , размеченный граф состояний которой дан на рис. 5.7, уравнения для финальных вероятностей состояний имеют вид

$$\begin{aligned}
 (\lambda_{12} + \lambda_{13})P_1 &= \lambda_{21}P_2; \\
 (\lambda_{21} + \lambda_{24})P_2 &= \lambda_{12}P_1 + \lambda_{42}P_4; \\
 (\lambda_{34} + \lambda_{35})P_3 &= \lambda_{13}P_1; \\
 \lambda_{42}P_4 &= \lambda_{24}P_2 + \lambda_{34}P_3 - \lambda_{54}P_5; \\
 \lambda_{54}P_5 &= \lambda_{35}P_3.
 \end{aligned}
 \tag{5.20}$$

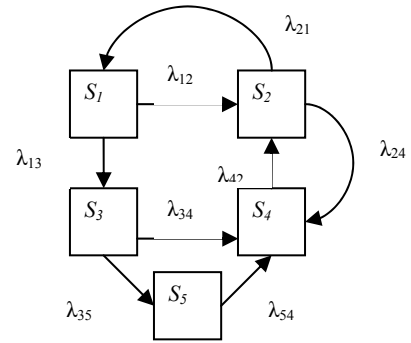


Рисунок.5.7

Таким образом, получается (для системы S с n состояниями) система n однородных линейных алгебраических уравнений с n неизвестными p_1, p_2, \dots, p_n . Из этой системы можно найти неизвестные p_1, p_2, \dots, p_n с точностью до произвольного множителя.

Чтобы найти точные значения p_1, \dots, p_n к уравнениям добавляют нормировочное условие $p_1 + p_2 + \dots + p_n = 1$, пользуясь которым можно выразить любую из вероятностей p_i через другие (и соответственно отбросить одно из уравнений).

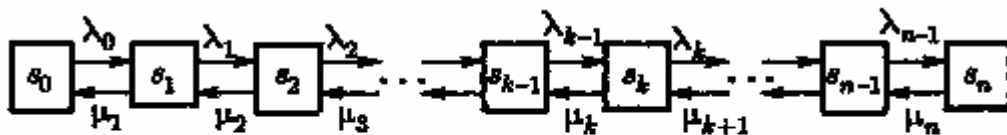


Рисунок 5.8 - Схема гибели и размножения

На практике очень часто приходится встречаться с системами, граф состояний можно вытянуть в цепь, причем каждое из них связано прямой и обратной связью с двумя соседними, кроме двух крайних. Она называется *схемой гибели и размножения*. Это название заимствовано из биологических задач, где состояние популяции S_k означает наличие в ней k единиц. Переход вправо связан с «размножением» единиц, а влево с их «гибелью». На рис. 5.8 «интенсивности размножения» ($\lambda_0, \lambda_1, \dots, \lambda_{n-1}$) и «интенсивности гибели» ($\mu_1, \mu_2, \dots, \mu_{n-1}$) проставлены у стрелок. Для схемы гибели и размножения финальные вероятности состояний выражаются формулами:

$$p_1 = \frac{\lambda_0}{\mu_1} p_0; \quad p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0; \quad \dots; \quad p_k = \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} p_0 \quad (k = 0, \dots, n); \dots;$$

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0; \quad p_0 = \left\{ 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} \right\}^{-1} \quad (5.21)$$

Если процесс описывается схемой гибели и размножения, то можно записать дифференциальные уравнения для математического ожидания и дисперсии случайной функции $X(t)$ — числа единиц в системе в момент времени t :

$$\frac{dm_x(t)}{dt} = \sum_{k=0}^n (\lambda_k - \mu_k) p_k(t); \quad (5.22)$$

$$\frac{dD_x(t)}{dt} = \sum_{k=0}^n [\lambda_k + \mu_k + 2k(\lambda_k - \mu_k) - 2m_x(t)(\lambda_k - \mu_k)] p_k(t) \quad (5.23)$$

В этих формулах нужно полагать $\lambda_n = \mu_0 \equiv 0$. Интенсивности λ_k ($0 \leq k \leq n-1$) и μ_k ($1 \leq k \leq n$) могут быть любыми неотрицательными функциями времени.

При достаточно больших значениях $m_x(t)$ (> 20) и выполнении условия $0 < m_x(t) \pm 3\sqrt{D_x(t)} < n$ можно приближенно полагать, что сечение случайной функции $X(t)$ представляет собой нормальную случайную величину с параметрами $m_x(t)$, $\sqrt{D_x(t)}$ полученными решением уравнений (5.22), (5.23). Формулы (5.22) и (5.23) остаются справедливыми при $n \rightarrow \infty$, если верхний предел в суммах заменить на ∞ .

5.1.3 Непрерывно – вероятностные модели

Системой массового обслуживания называют любую систему, в которой обслуживаются какие-либо заявки или требования, поступающие в случайные моменты времени. Примеры СМО: телефонная станция; бюро ремонта; билетная касса; парикмахерская; ЭВМ. Теория массового обслуживания изучает случайные процессы, протекающие в системах массового обслуживания.

Любое устройство, непосредственно занимающееся обслуживанием заявок, называется *каналом обслуживания*. СМО бывают как одноканальными, так и многоканальными. Пример одноканальной СМО - билетная касса с одним кассиром; пример многоканальной – та же касса с несколькими кассирами.

Различают СМО *с отказами* и СМО *с очередью*. В СМО с отказами заявка, пришедшая в момент, когда все каналы заняты, получает отказ, покидает СМО и в процессе ее дальнейшей работы не участвует. В СМО с очередью заявка, пришедшая в момент занятости всех каналов, не покидает СМО, а становится в очередь и ждет, пока не освободится какой-нибудь канал. Число мест в очереди m может быть как ограниченным, так и неограниченным. При $m=0$ СМО с очередью превращается в СМО с отказами. Очередь может иметь ограничения не только по количеству стоящих в ней заявок (по длине очереди), но и по времени ожидания (такие СМО называются «системами с нетерпеливыми клиентами»).

СМО с очередью различаются не только по ограничениям очереди, но и по *дисциплине обслуживания*: обслуживаются ли заявки в порядке поступления, или в случайном порядке, или же некоторые заявки обслуживаются вне очереди (так называемые «СМО с приоритетом»). Приоритет может иметь несколько градаций или рангов.

Аналитическое исследование СМО проще в случае, если все потоки событий являются простейшими (стационарными, пуассоновскими). Это значит, что интервалы времени между событиями в потоках имеют показательное распределение с параметром, равным интенсивности соответствующего потока. Для СМО это допущение означает, что как поток заявок, так и поток обслуживаний - простейшие. Под потоком обслуживания понимают поток заявок, обслуживаемых одна за другой одним непрерывно занятым каналом. Этот поток оказывается простейшим, только, если время обслуживания заявки $T_{обсл}$ представляет собой случайную величину, имеющую показательное распределение. Параметр этого распределения μ есть величина, обратная среднему времени обслуживания: $\mu = 1 / t_{обс}$, где $t_{обс} = M [T_{обсл}]$. Условимся в дальнейшем всякую СМО, в которой все потоки простейшие, называть *простейшей* СМО. Если все потоки событий простейшие, то процесс, протекающий в СМО, представляет собой случайный марковский процесс с дискретными состояниями и непрерывным временем. При выполнении некоторых условий для этого процесса существует финальный стационарный режим, при котором как вероятности состояний, так и другие характеристики процесса не зависят от времени.

Задачи теории массового обслуживания — нахождение вероятностей различных состояний СМО, а также установление зависимости между заданными параметрами (числом каналов n , интенсивностью потока заявок λ , распределением времени обслуживания и т.д.) и *характеристиками эффективности* работы СМО. В качестве таких характеристик могут рассматриваться:

среднее число заявок A , обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО;

вероятность обслуживания поступившей заявки Q или *относительная пропускная способность* СМО $Q = A / \lambda$;

вероятность отказа $P_{отк}$, т.е. вероятность того, что поступившая заявка не будет обслужена, получит отказ; $P_{отк} = 1 - Q$;

среднее число заявок в СМО (обслуживаемых или ожидающих в очереди) z ;

среднее число заявок в очереди \check{r} ;

среднее время пребывания заявки в СМО (в очереди или под обслуживанием) $t_{сум}$;

среднее время пребывания заявки в очереди $t_{оч}$;

среднее число занятых каналов k .

В общем случае все эти характеристики зависят от времени. Но многие СМО работают в неизменных условиях достаточно долгое время, и поэтому для них успевают установиться режим, близкий к стационарному.

Для любой открытой СМО в предельном стационарном режиме среднее время пребывания заявки в системе $t_{cисм}$ выражается через среднее число заявок в системе с помощью формулы Литтла:

$$\bar{t}_{cисм} = \bar{z} / \lambda, \quad (5.26)$$

где λ — интенсивность потока заявок.

СМО называется открытой, если интенсивность поступающего на нее потока заявок не зависит от состояния самой СМО.

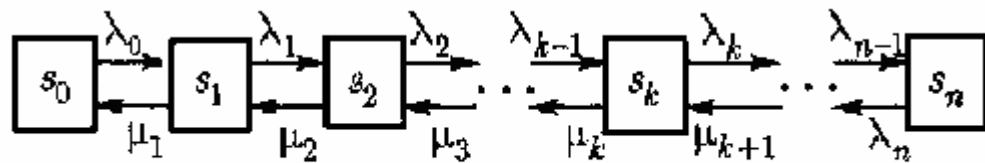


Рисунок 5.9 - Граф состояний СМО

Аналогичная формула (называемая также формулой Литтла) связывает среднее время пребывания заявки в очереди $t_{оч}$ и среднее число r заявок в очереди:

$$\bar{t}_{оч} = r / \lambda. \quad (5.27)$$

Формулы Литтла позволяют вычислять не обе характеристики эффективности (среднее время пребывания и среднее число заявок), а только какую-нибудь одну из них.

Специально подчеркнем, что формулы (5.26) и (5.27) справедливы для любой открытой СМО (одноканальной, многоканальной, при любых видах потоков заявок и обслуживаний); единственное требование к потокам заявок и обслуживаний - чтобы они были стационарными.

Аналогично универсальное значение для открытых СМО имеет формула, выражающая среднее число занятых каналов \bar{k} через абсолютную пропускную способность A / μ ,

$$\bar{k} = A / \mu, \quad (5.28)$$

где μ - интенсивность потока обслуживаний.

Очень многие задачи теории массового обслуживания, касающиеся простейших СМО, решаются с помощью схемы гибели и размножения. Если граф состояний СМО может быть представлен в виде, показанном на рис. 11.1, то финальные вероятности состояний выражаются формулами:

$$p_0 = \left\{ 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} + \dots + \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} \right\}^{-1}$$

$$p_1 = \frac{\lambda_0}{\mu_1} p_0; p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0; \dots; p_k = \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} p_0 \quad (0 \leq k \leq n); \dots;$$

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0. \quad (5.29)$$

При выводе формул для среднего числа заявок (в очереди или в системе) широко используется прием дифференцирования рядов, состоящий в следующем. Если $x < 1$, то

$$\sum_{k=1}^{\infty} kx^k = x \sum_{k=1}^{\infty} \frac{d}{dx} x^k = x \frac{d}{dx} \sum_{k=1}^{\infty} x^k = \frac{d}{dx} \frac{x}{1-x} = \frac{x}{(1-x)^2}$$

и окончательно

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}. \quad (5.30)$$

Ниже приведем без вывода ряд формул, выражающих финальные вероятности состояний и характеристики эффективности для некоторых часто встречающихся типов СМО.

1 Простейшая СМО с отказами (задача Эрланга). На n -канальную СМО с отказами поступает простейший поток заявок с интенсивностью λ , время обслуживания — показательное с параметром μ . Состояния СМО нумеруются по числу заявок, находящихся в СМО (в силу отсутствия очереди, оно совпадает с числом занятых каналов):

S_0 - СМО свободна;

S_1 - занят один канал, остальные свободны; ...;

S_k - занято k : каналов, остальные свободны ($1 \leq k \leq n$); ...;

S_n - заняты все n каналов.

Финальные вероятности состояний выражаются формулами Эрланга:

$$p = \left\{ 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \right\}^{-1}; \quad p_k = \frac{\rho^k}{k!} p_0 \quad (k = 1, 2, \dots, n), \quad (5.31)$$

где $\rho = \lambda / \mu$.

Характеристики эффективности:

$$A = \lambda(1 - p_n); \quad Q = 1 - p_n; \quad P_{омк} = p_n; \quad \bar{k} = \rho(1 - p_n). \quad (5.32)$$

При больших значениях n вероятности состояний (5.33) удобно вычислять через табулированные функции:

$$P(m, a) = \frac{a^m}{m!} e^{-a} \quad (\text{распределение Пуассона}) \quad (5.33)$$

и

$$R(m, a) = \sum_{k=0}^m \frac{a^k}{k!} e^{-a} \quad (5.34)$$

по которым можно найти:

$$P = (m, a) = R(m, a) - R(m-1, a). \quad (5.35)$$

Пользуясь этими функциями, формулы Эрланга (5.31) можно переписать в виде

$$p_k = P(k, \rho) / R(n, \rho) \quad (k = 0, 1, \dots, n), \quad (k = 0, 1, \dots, n). \quad (5.36)$$

2 Простейшая одноканальная СМО с неограниченной очередью. На одноканальную СМО поступает простейший поток заявок с интенсивностью λ . Время обслуживания — показательное с параметром $\mu = 1/t_{обсл}$. Длина очереди не ограничена. Финальные вероятности существуют только при

$\rho = \lambda / \mu < 1$ (при $\rho \geq 1$ очередь растет неограниченно). Состояния СМО нумеруются по числу заявок в СМО, находящихся в очереди или обслуживаемых:

- S_0 — СМО свободна;
- S_1 — канал занят, очереди нет;
- S_2 — канал занят, заявка стоит в очереди; ...;
- S_k — канал занят, $k-1$ заявок стоят в очереди;

Финальные вероятности состояний выражаются формулами:

$$p_0 = 1 - \rho, \quad p_k = \rho^k (1 - \rho) \quad (k = 1, 2, \dots), \quad (5.37)$$

где $\rho = \lambda / \mu < 1$.

Характеристики эффективности СМО:

$$A = \lambda; \quad Q = 1; \quad P_{отк} = 0;$$

$$\bar{Z} = \frac{\mu}{1 - \rho}; \quad \bar{r} = \frac{\rho^2}{1 - \rho}; \quad \bar{t}_{сист} = \frac{\rho}{\lambda(1 - \rho)}; \quad \bar{t}_{оч} = \frac{\rho^2}{\lambda(1 - \rho)};$$

среднее число занятых каналов (или вероятность того, что канал занят)

$$\bar{k} = \lambda / \mu = \rho.$$

3 Простейшая одноканальная СМО с ограниченной очередью. На одноканальную СМО поступает простейший поток заявок с интенсивностью λ ; время обслуживания — показательное с параметром $\mu = 1 / \bar{t}_{обсл}$. В очереди m мест. Если заявка приходит в момент, когда все эти места заняты, она получает отказ и покидает СМО. Состояния СМО:

- s_0 — СМО свободна;
- s_1 — канал занят, очереди нет;
- s_2 — канал занят, одна заявка стоит в очереди; ...;
- s_k — канал занят, $k-1$ заявок стоят в очереди; ...;
- s_{m+1} — канал занят, m заявок стоят в очереди.

Финальные вероятности состояний существуют при любом $\rho = \lambda / \mu$ и равны:

$$p_0 = \frac{1 - \rho}{1 - \rho^{m+2}}; \quad P_k = \rho^k p_0 \quad (k = 1, \dots, m + 1). \quad (5.38)$$

Характеристики эффективности СМО:

$$A = \lambda(1 - p_{m+1}); \quad Q = 1 - p_{m+1}; \quad P_{отк} = p_{m+1}.$$

Среднее число занятых каналов (вероятность того, что канал занят)

$$\bar{k} = 1 - p_0. \quad (5.39)$$

Среднее число заявок в очереди

$$\bar{r} = \frac{\rho^2 [1 - \rho^m (m + 1 - m\rho)]}{(1 - \rho^{m+2})(1 - \rho)}. \quad (5.40)$$

Среднее число заявок в СМО

$$\bar{z} = \bar{r} + \bar{k}.$$

По формуле Литтла среднее время пребывания заявки в СМО равно

$$\bar{t}_{сист} = \bar{z} / \lambda; \quad \bar{t}_{оч} = \bar{r} / \lambda.$$

4 Простейшая многоканальная СМО с неограниченной очередью.

На n -канальную СМО поступает простейший поток заявок с интенсивностью λ ; время обслуживания одной заявки - показательное с параметром μ . Финальные вероятности существуют только при $\rho/n = \chi < 1$, где $\rho = \lambda/\mu$. Состояния СМО нумеруются по числу заявок в СМО:

s_0 — СМО свободна;

s_1 — занят один канал; ...;

s_k — занято k каналов ($1 \leq k \leq n$); ...;

s_n — заняты все n каналов;

s_{n+1} — заняты все n каналов, одна заявка стоит в очереди; ...;

s_{n+r} — заняты все n каналов, r заявок стоят в очереди; Финальные вероятности состояний выражаются формулами:

$$P_0 = \left\{ 1 + \frac{\rho}{1!} + \dots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n * n!} \frac{1}{1 - \chi} \right\}^{-1};$$

$$P_k = \frac{\rho^k}{k!} p_0 (1 \leq k \leq n); \quad p_{n+r} = \frac{\rho^{n+r}}{n^r * n!} p_0 (r \geq 1).$$

С помощью функций $P(m, a)$ и $R(m, a)$ и этой формулы могут быть приведены к виду

$$p_k = \frac{P(k, \rho)}{R(n, \rho) + P(n, \rho) \frac{\chi}{1 + \chi}} (k = 0, \dots, n);$$

$$P_{n+r} = \chi^r p_n (r = 1, 2, \dots).$$

Характеристики эффективности СМО:

$$\bar{r} = \rho^{n+1} p_0 / [n n! (1 - \chi)^2] = \chi * p_n / (1 - \chi)^2;$$

$$\bar{z} = \bar{r} + \bar{k} = \bar{r} + \rho;$$

$$\bar{t}_{\text{сум}} = \bar{z} / \lambda; \quad \bar{t}_{\text{оч}} = \bar{r} / \lambda.$$

5 Простейшая многоканальная СМО с ограниченной очередью.

Условия и нумерация состояний те же, что в п. 4, с той разницей, что число m мест в очереди ограничено. Финальные вероятности состояний существуют при любых λ и μ и выражаются формулами:

$$p_0 = \left\{ 1 + \frac{\rho}{1!} + \dots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n * n!} \frac{1 - \chi^m}{1 - \chi} \right\}^{-1};$$

$$p_k = \frac{\rho^k}{k!} p_0 (1 \leq k \leq n);$$

$$p_{n+r} = \frac{\rho^{n+r}}{n^r * n!} p_0 (1 \leq r \leq m),$$

$$\chi = \rho/n = \lambda/(n\mu).$$

5.2 Имитационное моделирование процессов функционирования систем массового обслуживания

5.2.1 Формирование входных воздействий

Рассмотренные ранее аналитические модели СМО применимы только для простейших систем при условии, что в них потоки событий являются простейшими. Исследование СМО, аналитические модели которых не разработаны, осуществляется на имитационных моделях.

Рассмотрим возможности использования концептуальных моделей в форме *Q-схем* для формального описания процесса функционирования некоторой СМО. Характерная ситуация в работе такой системы — появление заявок на обслуживание и завершение обслуживания в случайные моменты времени, то есть стохастический характер процесса их функционирования. В общем случае моменты поступления заявок в систему из внешней среды образуют входящий поток, а моменты окончания обслуживания образуют выходящий поток обслуженных заявок [8,11].

Формализация на базе Q-схем. Формализуя какую-либо реальную систему с помощью Q-схемы, необходимо построить структуру такой системы. В качестве элементов структуры Q-схем будем рассматривать элементы трех типов: И — источники; Н — накопители; К — каналы обслуживания заявок.

Как отмечалось выше, Q-схему можно считать заданной, если определены:

потоки событий (входящие потоки заявок для каждого накопителя и потоки обслуживаний для каждого канала);

структура системы (число фаз, число накопителей и каналов обслуживания в каждой из фаз обслуживания заявок и связи между И, Н и К);

алгоритмы функционирования системы (дисциплины ожидания заявок в накопителях и выбора их на обслуживание в каналы, правила ухода заявок из Н и К).

Рассмотрим возможности формализации входных воздействий, представляемых в Q-схемах в виде источников. Формирование однородных потоков событий заданных многомерным интегральным законом или плотностью распределения вероятностей, т.е.

$$F(y_1, y_2, \dots, y_k) = P\{\tau_1 < y_1, \tau_2 < y_2, \dots, \tau_k < y_k\},$$
$$f(y_1, y_2, \dots, y_k) = F^k(y_1, y_2, \dots, y_k) / (\partial y_1 \partial y_2 \dots \partial y_k),$$

сводится к машинной имитации k -мерных векторных величин, требующих больших затрат машинных ресурсов. При моделировании систем, формализуемых в виде Q-схем, часто возникают задачи имитации потоков заявок с некоторыми ограничениями, позволяющими упростить как

математическое описание, так и программную реализацию генераторов потоков заявок [2].

Так, для ординарных потоков с ограниченным последствием интервалы между моментами поступления заявок являются независимыми и совместная плотность распределения может быть представлена в виде произведения частных законов распределения $f(y_1, y_2, \dots, y_k) = f_1(y_1)f_2(y_2)\dots, f_k(y_k)$, где $f(y_i), i = \overline{1, k}$ при $i > 1$ являются условными функциями плотности распределения величин y_i при условии, что в момент начала i -го интервала поступит заявка. Относительно начального момента времени t_0 никаких предположений не делается, поэтому функция $f_1(y_1)$ — безусловная.

Если поток с ограниченным последствием удовлетворяет условию стационарности, т. е. вероятность появления k событий на интервале $(t_0, t_0 + \Delta t)$ зависит только от длины интервала Δt , то при $i > 0$ интервалы τ_i распределены одинаково, т. е.

$$f_2(y_2) = f_3(y_3) = \dots = f_k(y_k).$$

Плотность распределения первого интервала $f_1(y_1)$ может быть найдена с использованием соотношения Пальма

$$f_1(y_1) = \lambda(1 - \int_0^{y_1} f(y)dy), \quad (5.41)$$

где λ — интенсивность потока событий.

Порядок моделирования моментов появления заявок в стационарном потоке с ограниченным последствием следующий.

Из последовательности случайных чисел, равномерно распределенных на интервале $(0, 1)$, выбирается случайная величина и формируется первый интервал y_1 в соответствии с (5.41) любым из рассмотренных выше способов формирования случайной величины. Момент наступления первого события равен $t_1 = t_0 + y_1$, а моменты появления следующих событий определяются выражениями

$$t_2 = t_1 + y_2, \dots, t_k = t_{k-1} + y_k, \quad (5.42)$$

где y_k — случайная величина с плотностью $f(y)$.

Пример 1. Пусть при моделировании некоторой системы необходимо сформировать на ЭВМ простейший поток заявок. Распределение интервалов между заявками подчиняется экспоненциальному закону $f(y) = \lambda e^{-\lambda y}, y > 0$. Используем формулу Пальма для определения распределения первого интервала y

$$f_1(y_1) = \lambda(1 - \int_0^{y_1} \lambda e^{-\lambda y} dy) = \lambda e^{-\lambda y_1}.$$

Из этого выражения следует, что $f_1(y_1) = f(y)$, т.е. первый интервал распределен так же, как и остальные. Этого и следовало ожидать ввиду отсутствия последствия в простейшем потоке. Формируя на ЭВМ

равновероятные случайные числа x_i на интервале $(0;1)$, будем преобразовывать их в соответствии с выражением

$$\int_0^{y_i} f(y)dy = x_i.$$

Тогда длина интервала между $(i-1)$ -м и i -м событием будет $y_i = -(1/\lambda)\ln x_i$, а моменты появления заявок в потоке определяются согласно (5.42).
 Пример 2. Пусть при моделировании некоторой системы требуется сформировать на ЭВМ поток событий, равномерно распределенных на интервале (a,b) . Функция плотности распределения интервалов между событиями $f(y) = 1/(b-a), a \leq y \leq b$. Распределение первого интервала между началом отсчета и первым событием

$$f_1(y_1) = \lambda(1 - \int_0^{y_1} f(y)dy) = \lambda[1 - \int_0^{y_1} dy/(b-a)].$$

Интенсивность потока

$$\lambda = 1/M[y] = 1/\int_a^b yf(y)dy = 2/(a+b).$$

Тогда $f_1(y_1) = 2[1 - y_1/(b-a)]/(a+b)$.

Заметим, что математическое ожидание первого интервала $M[y_1]$ отличается от математического ожидания интервалов при $i > 1$:

$$M[y_1] = \int_a^b y_1 f(y_1) dy_1 = [1/(a+b)] \int_a^b [y_1 - y_1^2/(b-a)] dy_1.$$

Длины интервалов между событиями будут

$$\int_a^{y_1} f(y_1) dy_1 = x_i, \int_a^{y_1} f(y) dy = x_i.$$

Так, например, при $i > 1$ получим

$$\int_a^{y_1} [1/(b-a)] dy = x_i, y_i = a + (b-a)x_i,$$

где x_i - случайная величина, равномерно распределённая на интервале $(0,1)$.

5.2.2 Способы построения моделирующих алгоритмов

Моделирующий алгоритм Q-схемы должен адекватно отражать процесс функционирования системы и в то же время не создавать трудностей при машинной реализации модели. При этом моделирующий алгоритм должен отвечать следующим основным требованиям:

- обладать универсальностью относительно структуры, алгоритмов функционирования и параметров системы;

- обеспечивать одновременную (в один и тот же момент системного времени) и независимую работу необходимого числа элементов системы;

- укладываться в приемлемые затраты ресурсов ЭВМ (машинного времени и памяти) для реализации машинного эксперимента;

- обеспечивать возможность построения блочной структуры алгоритма;

гарантировать выполнение рекуррентного правила — событие, происходящее в момент времени t_k , может моделироваться только после того, как промоделированы все события, произошедшие в момент времени $t_{k-1} < t_k$.

При этом необходимо учитывать, что появление очередной заявки в некоторый момент времени t_i может вызвать изменение состояния не более, чем в одном из элементов Q-схемы, а окончание обслуживания заявки в некотором канале может привести к последовательному изменению состояния нескольких элементов (Н и К), т. е. будет иметь место процесс распространения смены состояний в направлении, противоположном движению заявок в системе.

Известно, что существует два основных принципа построения моделирующих алгоритмов: принцип « Δt » и принцип « δz ». При построении моделирующего алгоритма Q-схемы по принципу « Δt », т. е. алгоритма с детерминированным шагом, необходимо определить минимальный

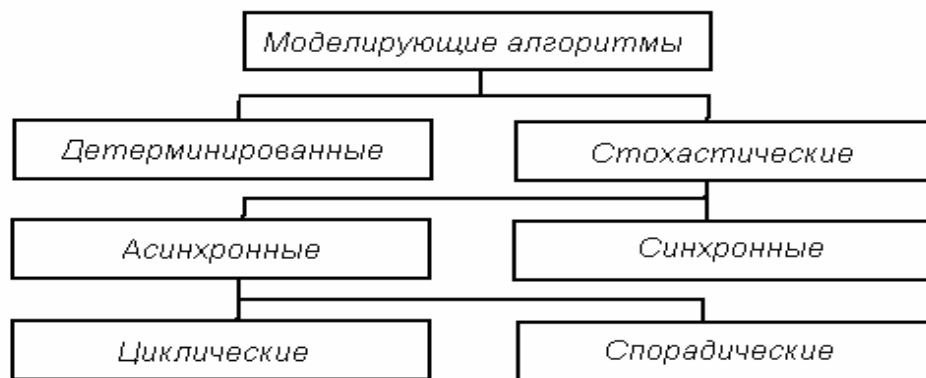


Рисунок 5.10 Классификация способов построения моделирующих алгоритмов Q - схем

интервал времени между соседними событиями $\Delta t' = \min \{u_i\}$ (во входящих потоках и потоках обслуживания) и принять шаг моделирования равным $\Delta t'$.

В моделирующих алгоритмах, построенных по принципу « δz », т. е. в алгоритмах со случайным шагом, элементы Q-схемы просматриваются при моделировании только в моменты особых состояний (в моменты появления заявок из источников или изменения состояний каналов).

При этом длительность шага $\Delta t = \text{var}$ зависит как от особенностей самой системы, так и от входных воздействий. Моделирующие алгоритмы со случайным шагом реализуют синхронным и асинхронным способами. При синхронном способе один из элементов Q-схемы (И, Н или К) выбирается в качестве ведущего и по нему «синхронизируется» весь процесс моделирования. При асинхронном способе построения моделирующего алгоритма ведущий (синхронизирующий) элемент не используется, а очередному шагу моделирования (просмотру элементов Q-схемы) может соответствовать любое особое состояние всего множества элементов схемы. При этом просмотр элементов Q-схемы организован так, что при каждом особом состоянии либо циклически просматриваются все элементы, либо

спорадически — только те, которые могут в этом случае изменить свое состояние (просмотр элементов с прогнозированием) [2]. Классификация возможных способов построения моделирующих алгоритмов Q-схем приведена на рис. 5.10.

Рассмотрим особенности построения и реализацию перечисленных разновидностей моделирующих алгоритмов при моделировании конкретного варианта системы.

5.2.3 Особенности имитации процесса функционирования систем

Математическое обеспечение и возможности современных ЭВМ позволяют достаточно эффективно провести моделирование различных систем, формализуемых в виде Q-схем, используя либо пакеты прикладных программ, созданные на базе алгоритмических языков общего назначения, либо специализированные языки имитационного моделирования. Но прежде чем применять эти средства автоматизации моделирования, необходимо глубже проникнуть в суть процесса построения и реализации моделирующих алгоритмов [2].

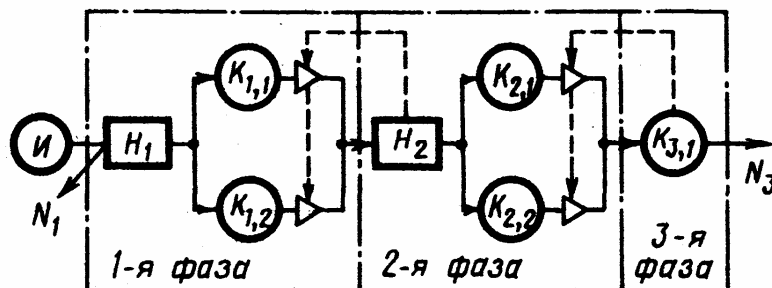


Рисунок 5.11 - Пример Q-схемы общего вида

При моделировании СМО технология машинной имитации процесса функционирования ее зависит от структуры схемы, особенностей построения моделирующего алгоритма и принятого принципа изменения модельного времени. Структура конкретной СМО общего вида (рис. 5.11) представляет собой трехфазную Q-схему ($L^{\phi} = 3$) с блокировкой каналов по выходу в 1-й и 2-й фазах обслуживания (пунктирные линии на рисунке). Выходящими потоками такой Q-схемы являются: поток потерянных заявок из H_1 и поток обслуженных заявок из $K_{3,1}$, (N_1 и N_3 на рис. 5.11).

Для имитационной модели данной Q-схемы можно записать следующие переменные и уравнения:

зависимая переменная P - вероятность потери заявок;

независимые переменные: t_m — время появления очередной заявки из И; $t_{k,j}$ - время окончания обслуживания j -м каналом k -той фазы, $k=1, 2, 3$; $j=1, 2$;

вспомогательные переменные: z_i и $z_{k,j}$ - состояния H_i и $K_{k,j}$, $i=1, 2$; $k=1, 2, 3$; $j=1, 2$;

параметры: L_i - емкость i -го H_i ; L_k^k —число каналов в k -й фазе;
 переменные состояния: N_1 — число потерянных заявок в H_1 ; N_3 —
 число обслуженных заявок, т. е. вышедших из 3-й фазы;
 уравнение модели:

$$P = N_1 / (N_1 + N_3) = N_1 / N .$$

При имитации процесса функционирования Q-схемы на ЭВМ требуется организовать массив состояний. В этом массиве должны быть выделены:

подмассив каналов K для запоминания текущих значений состояний $z_{k,j}$ соответствующих каналов и времени окончания обслуживания очередной заявки $t_{k,j}$;

подмассив накопителей H для записи текущих значений состояний z_i соответствующих накопителей H_i , $i=1, 2$;

подмассив источников I , в который записывается время поступления очередной заявки t_m из источника.

Процедура моделирования процесса обслуживания каждым элементарным каналом $k_{k,j}$ сводится к следующему.

Путем обращения к генератору случайных чисел с законом распределения, соответствующим обслуживанию $k_{k,j}$ данных, определяется длительность обслуживания $\tau_{k,j}$ и вычисляется время окончания обслуживания $t_{k,j}$, а затем фиксируется состояние $z_{k,j} = 1$; при освобождении канала состояние $z_{k,j} = 0$; в случае блокировки канала записывается $z_{k,j} = 2$. При поступлении заявки в H_i к его содержимому добавляется единица, т. е. $z_i = z_i + 1$, а при уходе заявки из H_i , на обслуживание вычитается единица, т. е. $z_i = z_i - 1$, $i=1, 2$.

5.2.4 Моделирующие алгоритмы процессов функционирования системы

Детерминированный моделирующий алгоритм. Укрупненная схема моделирующего алгоритма (МА) с постоянным шагом соответствующего системе, изображенной на рис.5.11, приведена на рис. 5.12. Особенностью ее является наличие как блока системного времени, вычисляющего значения текущих моментов времени $t_n = t_{n-1} + \Delta t$, так и блока, определяющего момент окончания моделирования по условию $t_n \geq T$, где T - время моделирования. Кроме вспомогательных блоков общего назначения: ввод исходных данных (ВИД), установка начальных условий (УНУ), обработка (ОРМ) и выдача результатов моделирования (ВРМ), моделирующий алгоритм содержит блоки, отражающие специфику детерминированного подхода (блоки 4...9). Рассмотрим детальные схемы алгоритмов этих блоков (рис.5.13,...,5.16) с учетом принятых обозначений в них: $ZN(I) \equiv z_i$; $Z(K, J) \equiv z_{k,j}$; $TM \equiv t_m$; $TN \equiv t_n$;

$T(K, J) \equiv t_{k,j}; L(I) \equiv L_i; NO1 \equiv N_1; NO3 \equiv N_3$. Процедура обслуживания заявки каналом $K_{k,j}$ оформлена в виде подпрограммы $WORK[K(K, J)]$, позволяющей обратиться к генератору

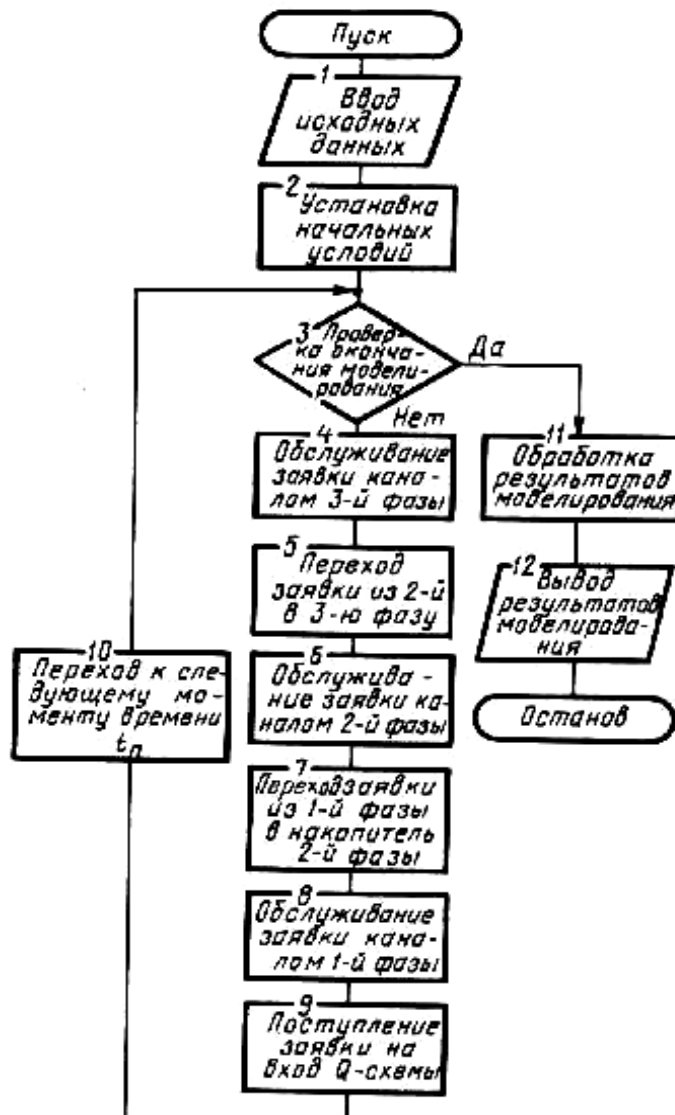


Рисунок 5.12 – Укрупненная схема МА

моделирования переходят к имитации обслуживания заявки каналом 3-й фазы. Если закончено обслуживание в канале 3-й фазы, то фиксируется выход из системы очередной заявки (подсчет их) и производится освобождение этого канала (рис.5.13а).

случайных чисел с соответствующим данному каналу законом распределения, генерирующей длительности интервалов обслуживания очередных заявок $t_{k,j}$. Процедура генерации заявок источником также оформлена в виде подпрограммы, но $D(TM)$, которая определяет момент поступления t_m очередной заявки в систему.

Окончание обслуживания заявки каналом $K_{k,j}$ в момент t_n может вызвать процесс распространения изменений состояний элементов системы в направлении, обратном движению заявки, поэтому все накопители и каналы должны просматриваться при моделировании, начиная с обслуживающего канала последней фазы по направлению к накопителю первой фазы (рис.5.11). После проверки условия окончания

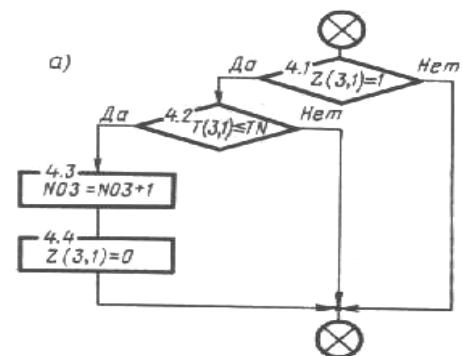


Рисунок 5.13 – Схема алгоритма блока 4а

Далее имитируется переход заявки из 2-й в 3-ю фазу (рис.5.14б). При этом осуществляется просмотр каналов 2-й фазы и определение заявок, ожидающих обслуживания в канале $K(3,1)$ и, если он свободен, то согласно дисциплине обслуживания выбирается одна из заявок и имитируется ее обслуживание в канале $K(3,1)$, занятость его и освобождение канала 2-й фазы. Если канал $K(3,1)$ занят, то осуществляется блокировка заявки в канале 2-й фазы.

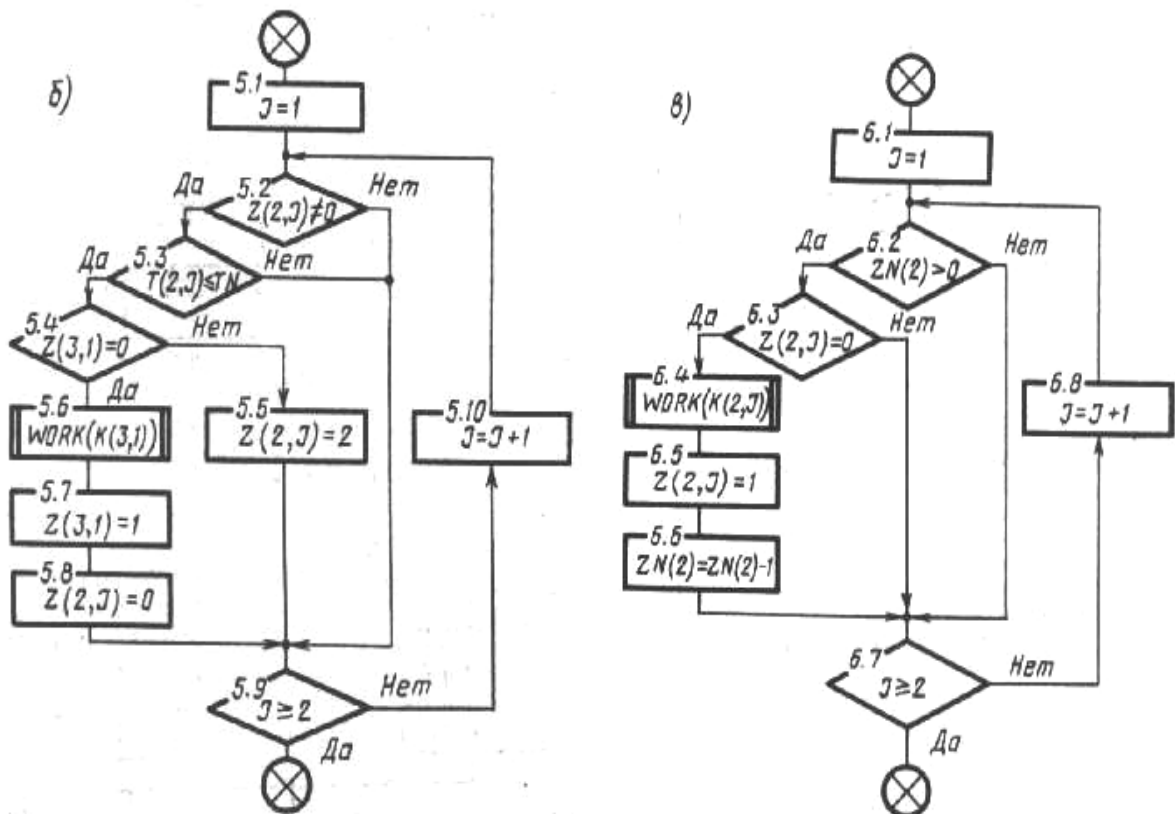


Рисунок 5.14 – Схемы алгоритмов блока 5(б), блока 6(в)

Имитация обслуживания заявок в каналах 2-й фазы проводится последовательно для каждого канала (рис.5.14в). Если в накопителе $H(2)$ имеются заявки и хотя бы один канал 2-й фазы свободен, то имитируется обслуживание заявки в одном из этих каналов и уменьшение числа заявок в накопителе.

Затем имитируется переход заявки из каналов 1-й фазы в накопитель и каналы 2-й фазы (рис. 5.15г).

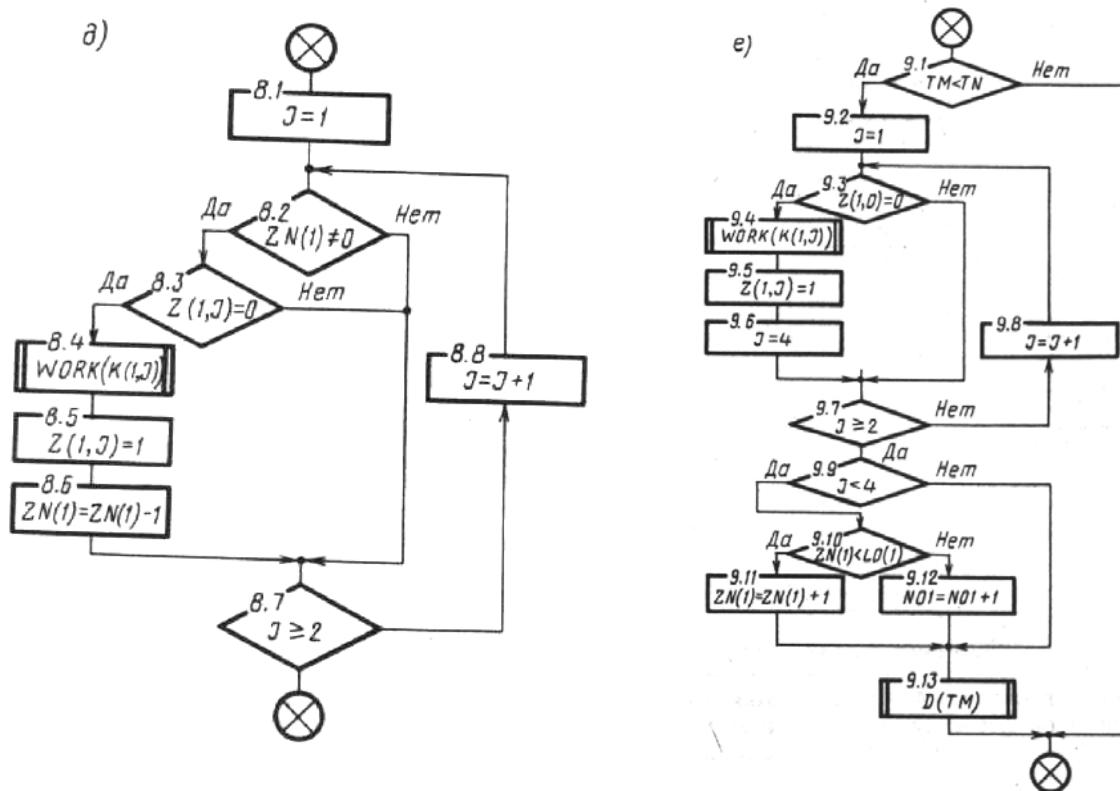


Рисунок 5.16 – Схемы алгоритмов блоков 8(д), 9(е)

Потом имитируется обслуживание заявок каналами 1-й фазы (рис.5.16д). Для этого проверяется наличие заявок в накопителе $H(1)$ и отсутствие их в каналах 1-й фазы, а также организуется обслуживание заявок каналами 1-й фазы и изменение состояний каналов и накопителя.

Алгоритм имитации поступления заявок из источника в накопитель 1-й фазы (рис.5.16е) с учетом занятости каналов заключается в последовательном выполнении следующих операций:

- проверка поступления заявки из источника в текущий момент времени;
- проверка отсутствия заявки в канале 1-й фазы;
- организация обслуживания поступившей заявки в канал первой фазы (вызов процедуры WORK и изменения состояния канала);
- генерация момента поступления очередной заявки.

Если каналы первой фазы заняты, то при наличии мест поступившая заявка помещается в накопитель первой фазы или получает отказ в обслуживании. После чего опять осуществляется генерация момента поступления очередной заявки и управления передается блоку 10, который определяет момент очередного шага моделирования (рис. 5.12).

Затем управление снова передается блоку 3, который при наборе необходимой статистики проводит обработку и выдачу результатов моделирования, а потом и остановку моделирования.

Синхронный моделирующий алгоритм. При построении схемы алгоритма для этой же СМО (Рис. 5.11) в случае, если синхронизирующим

элементом является источник заявок, необходимо учитывать следующие особенности.

В момент времени t_n на вход первой фазы СМО поступает очередная заявка из источника. С момента t_{n-1} до момента t_n могли произойти изменения состояний накопителя и каналов первой фазы. Эти изменения необходимо промоделировать раньше, чем поступит заявка в эту фазу. Это справедливо и для других фаз СМО. Укрупненная схема синхронного МА (рис.5.17) существенно отличается от схемы детерминированного алгоритма.

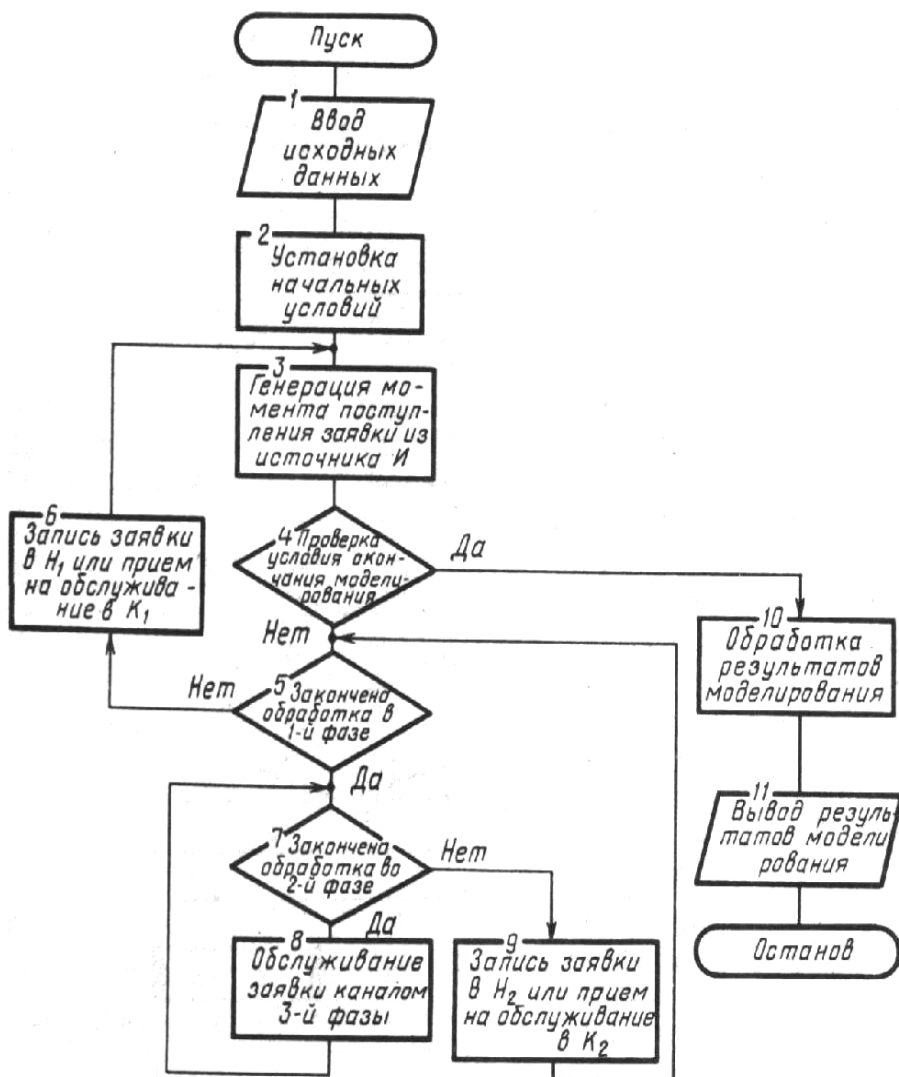


Рисунок 5.17 – Укрупненная схема синхронного МА

Особенность схемы этого алгоритма заключается в том, что генерация момента поступления заявки производится сразу после установки начальных условий, после чего проверяются: условие окончания моделирования и

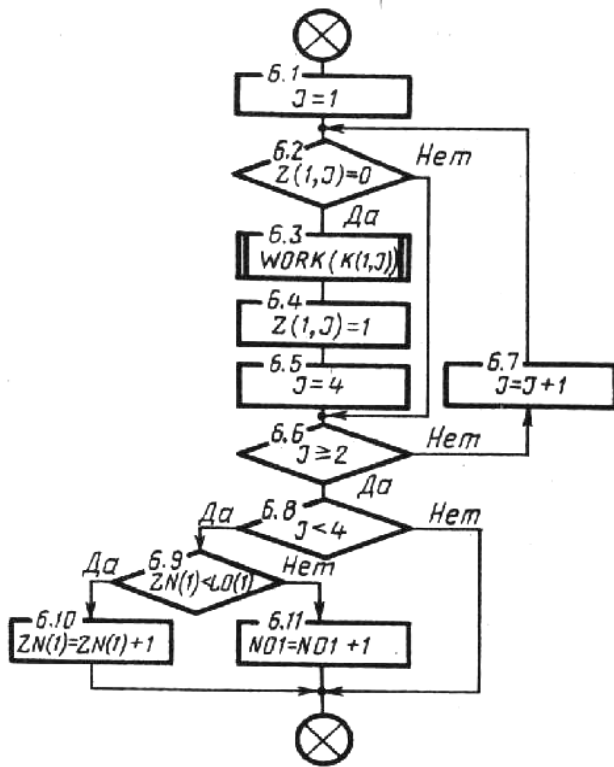


Рисунок 5.18 – Схема алгоритма блока 6

Асинхронный моделирующий алгоритм.

Этот алгоритм не имеет синхронизации. В нем очередному шагу соответствует особое состояние, то есть момент окончания обслуживания одной из заявок любым каналом любой фазы или момент поступления заявки из источника. Циклический просмотр элементов системы на каждом шаге позволяет определить, какие переходы заявок из элемента в элемент имеют место на данном шаге.

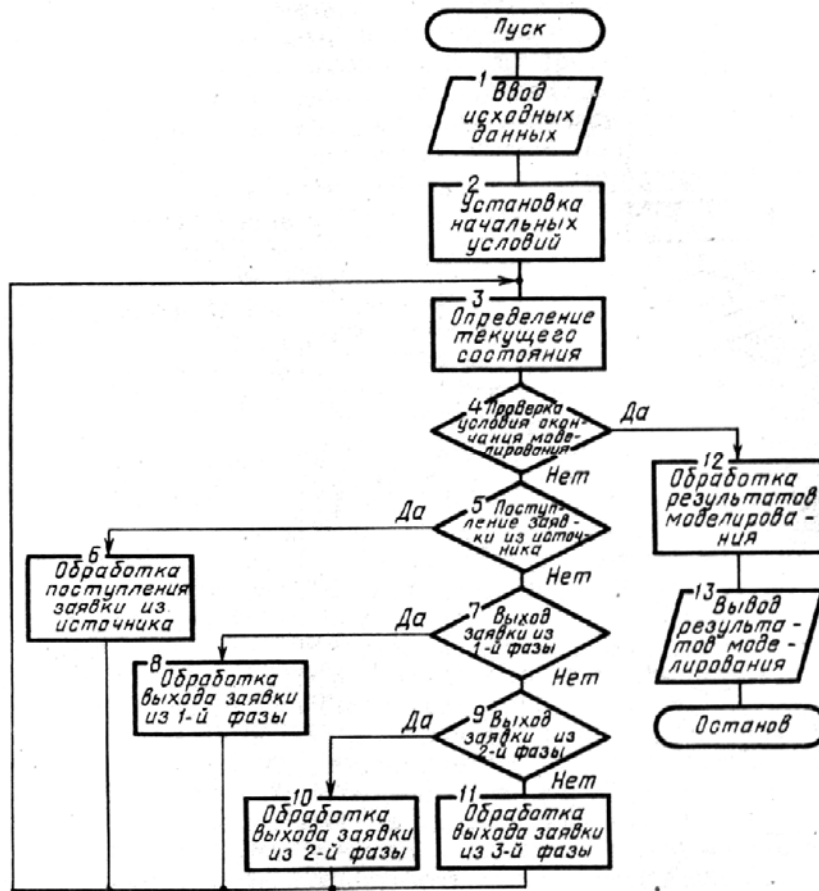


Рисунок 5.19 – Укрупненная схема асинхронного спорадического МА

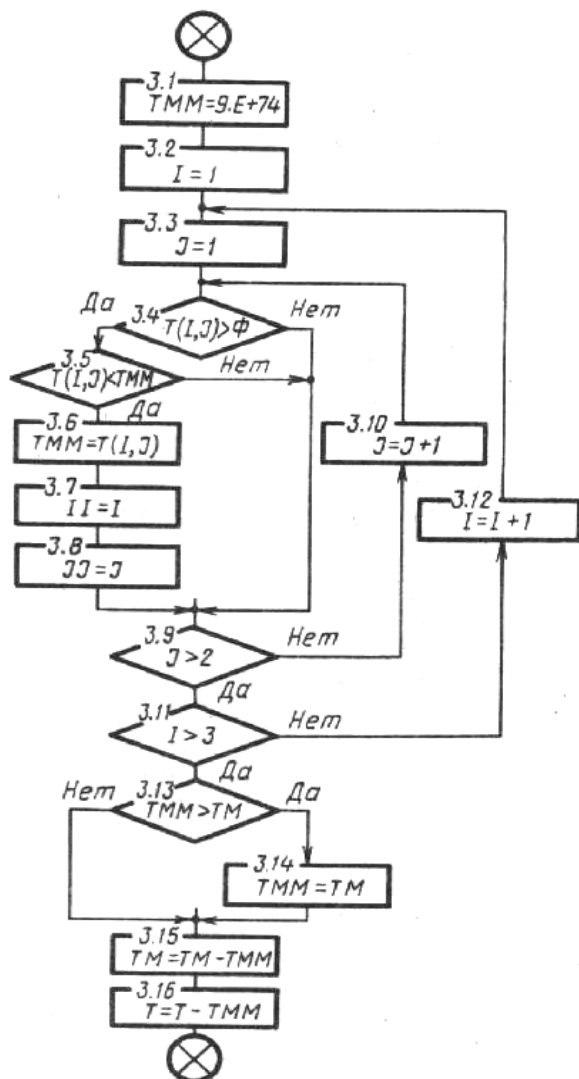


Рисунок 5.20-Схема МА блока3 изменения состояний элементов схемы поиском с помощью операторов 3.1,...3.12. В момент наступления ближайшего события изменение состояния осуществляется операторами 3.15 и 3.16.

Рассмотренный подход в построении МА позволяет алгоритмизировать процессы функционирования достаточно сложных систем. Сложность алгоритма возрастает, если учитываются такие особенности СМО как: наличие нескольких потоков заявок обслуживаемых своими каналами; наличие приоритетов при постановке заявки в очередь и выборе их на обслуживание; ограничение времени пребывания заявки в системе; отказ элементов системы и их восстановление.

Кроме сложности реализации алгоритмов необходимо учитывать затраты машинного времени и необходимый объем памяти ЭВМ. В этом случае целесообразно использовать асинхронный спорадический алгоритм, а также блочный принцип построения моделей. Моделирующий алгоритм процесса функционирования СМО может строиться как на использовании универсального алгоритмического языка, так и на языке имитационного моделирования GPSS.

Такой алгоритм аналогичен детерминированному алгоритму за исключением отсчета системного времени, определяемого выражением

$$t_n = \min(\min t_{k,j}, \min t_m),$$

и позволяющего определять моменты наступления следующих событий. Блок, выполняющий эту задачу, располагается перед блоком проверки окончания моделирования в детерминированном алгоритме.

Укрупненная схема (рис.5.19) асинхронного спорадического МА включает в себя особый блок 3 (рис. 5.20) для определения моментов текущего состояния по выражению

$$t_n = \min(\min t_{k,j}, t_m),$$

а также блоки 4,5,7,9 для проверки: условия окончания моделирования, поступления заявки из источника и выхода заявок из 1-й и 2-й фаз, и блоки 6,8,10,11 для их обработки.

Блок 3 определяет временной интервал до момента ближайшего

5.3 Имитационное моделирование систем массового обслуживания в среде «MATLAB»

5.3.1. Имитация потоков заявок и обслуживаний

В общем случае любая СМО состоит из источников заявок, источников обслуживаний, накопителей и каналов. Поскольку аналитические методы исследования СМО разработаны только для простейших СМО с очередью для однородных, ординарных и стационарных потоков заявок без последствия, то, чаще всего, исследование более сложных СМО осуществляют на имитационных моделях.

Одним из достаточно легко реализуемых способов такого исследования является структурное имитационное моделирование СМО в среде «Matlab» с использованием типовых элементов встроенной библиотеки «Simulink». Для создания имитационной модели (ИМ) СМО в этом случае необходимо сформировать ИМ типовых и структурных элементов СМО.

Основными особенностями функционирования элементарных приборов в СМО (накопитель - канал) являются: поступление заявок (требований) на обслуживание в случайные моменты времени; длительность обслуживания каждой заявки также случайно и приводит к окончанию обслуживания в канале в случайные моменты времени. Поскольку переходы СМО из одних состояний в другие происходят под действием потоков заявок, поступающих на вход системы, и потока обслуживаний, то процессы функционирования таких систем относят к случайным процессам с дискретными состояниями. Отсюда следует, что ИМ элементов должны изменять свое состояние, либо с постоянным шагом, либо со случайным шагом. Потоки заявок и обслуживаний, являясь потоками событий, чаще всего, могут иметь экспоненциальное (показательное) распределение, нормальное и равномерное, а также пуассоновское распределение.

Для создания ИМ источника заявок с равномерными или нормальным законом распределения интервалов между заявками можно воспользоваться последовательным соединением блока «Matlab Fcn» из подраздела «Функции и таблицы» и блока «Hit Crossing» (индикатор пересечения) из подраздела «Сигналы и системы» библиотеки «Simulink» (рис.5.21).

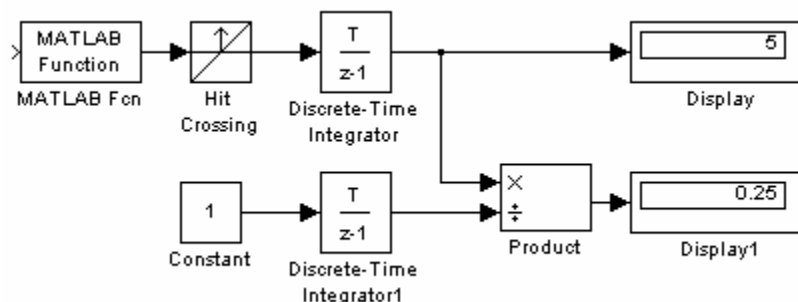


Рисунок 5.21 – Структурная схема исследования источника заявок с нормальным или равномерным распределением

После набора схемы ИМ источника заявок необходимо дополнить ее типовыми блоками для подсчета числа заявок «n» и установления требуемой интенсивности λ потока заявок в единицу времени.

Блок «*Matlab Fcn*» должен формировать последовательность случайных чисел с равномерным или нормальным распределением, для чего в его окне записывают $unifrnd(\alpha, \beta)$ или $normrnd(m, \sigma)$ соответственно, здесь α и β - границы интервала равномерного распределения случайных чисел; m и σ - математическое ожидание и средне - квадратичное отклонение, вводятся также в виде чисел отделенных запятой. Блок «*Hit Crossing*» осуществляет формирование импульсов единичной амплитуды и длительностью равной шагу моделирования Δt в моменты пересечения случайными числами уровня блока. Установка параметров этого блока заключается в установке уровня пересечения записью числа и выборе направления пересечения (*Rising, Falling, Either*).

Дискретный сумматор времени 1 производит подсчет поступивших заявок, поэтому его настройка сводится к установке шага моделирования Δt в окне «*Sample time*». В аналогичном окне второго сумматора времени обычно устанавливают масштабную единицу реального времени, поскольку интенсивность потока заявок обычно задается в реальном времени.

По завершении установки параметров блоков необходимо установить параметры моделирования в меню «*Simulation*» (моделирование). В вызванном окне устанавливают $t_{start}=0$, $t_{stop}=T_m$, шаг- фиксированный и величину шага Δt .

Требуемую интенсивность потока заявок устанавливают путем регулирования уровня пересечения в блоке «*Hit Crossing*».

Схема ИМ источника заявок (обслуживаний) с экспоненциальным законом распределения отличается от рассмотренной модели наличием преобразователя ПСЧ из равномерной в экспоненциальную (рис. 5.22) на блоке «*Fcn*».

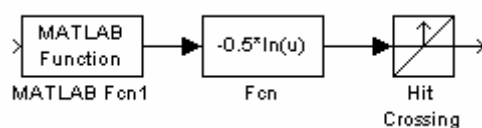


Рисунок 5.22 – Структурная схема ИМ источника заявок с экспоненциальным распределением

Настройка блоков модели включает в себя: в блоке «*Matlab*» запись выражения $unifrnd(0, 1)$; в блоке «*Fcn*» - $-\frac{1}{\lambda} \ln(u)$; в блоке «*Hit Crossing*» - регулировку уровня пересечения для достижения требуемой интенсивности потока.

Схема ИМ источника заявок с пуассоновским распределением может быть получена, если использовать оператор отношения (рис. 5.23).

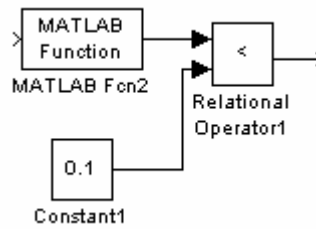


Рисунок 5.23 – Структурная схема ИМ источника заявок, имеющего пуассоновское распределение

Для этого распределения в блоке «*Matlab Fcn*» записывают также выражение, обеспечивающее формирование блоком ПСЧ с равномерным распределением в интервале $(0, 1)$, т.е. *unifrnd(0, 1)*.

В блоке «*Constant*» в зависимости от необходимой интенсивности $\lambda = NP$, где N – число опытов; p – вероятность появления события, устанавливают значение вероятности $p \leq 0,1 \dots 0,2$. Закон будет выполняться тем точнее, чем больше число опытов.

С помощью рассмотренных блоков можно создать ИМ источников заявок и обслуживания с другими законами распределения.

5.3.2 Модели накопителей и каналов обслуживания

Для формирования накопителя заявок в виде очереди можно использовать блок «*Queue*» (очередь), обеспечивающий сохранение значений входного сигнала в регистре, имеющем формат FIFO. При этом входной сигнал может быть скаляром, вектором или матрицей (рис. 5.24).

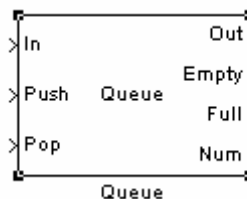


Рисунок 5.24 – Блок «*Queue*» (Очередь)

Блок имеет: входной порт «*In*», на который подается сигнал в виде числа; управляемый импульсами порт «*Push*» (постановка заявки в очередь); управляемый также импульсами порт «*Pop*» (выталкивание очередной заявки на обслуживание). К выходам блока относят порты: «*Out*», «*Empty*», «*Full*» и «*Number*». Через порт «*Out*» заявки покидают очередь при наличии управляющего импульса на входе «*Pop*». При этом значение сигнала на выходе сохраняется до следующего импульса на входе «*Pop*», это дает возможность моделировать с помощью блока элементарный прибор обслуживания. Выход «*Empty*» (пустой) выдает «1», если очередь пуста и «0» - в противном случае. Выход «*Full*» (полный) выдает «1», если в очереди находится число заявок равное емкости очереди, и «0» - в противном случае. Выход «*Number*» показывает число заявок в очереди на данном шаге.

Выходы: «Empty», «Full», «Number» устанавливаются флажками на панели блока. Подготовка блока «Очередь» производится путем: установки емкости очереди; выбора режима (*Rising, Falling, Either*); выбора режима при условии заполнения регистра очереди; выбора режима при условии, что регистр пуст, а следует вытолкнуть очередной элемент из очереди.

Накопление сигналов можно осуществлять и блоком «Stack» (*стек*), работающим в формате LIFO (последним вошел, первым вышел). Его использование в качестве ИМ накопителя аналогично блоку «Queue». Кроме рассмотренных блоков для накопления значений сигнала могут использоваться блоки «Сумматор» и «Дискретный сумматор времени».

5.3.3. Модели типовых схем систем массового обслуживания

В основе способа имитации канала обслуживания лежит блок «Очередь». На основе блока «Очередь» можно построить схему ИМ канала обслуживания с потерями или схему одноканальной СМО с очередью. Схема ИМ их одинакова (рис. 5.26), а подготовка схемы отличается лишь числом мест в очереди. В этой схеме источник заявок и источник обслуживаний представлены соответственно блоками: «Subsystem» и «Subsystem1». Для одноканальной СМО с отказами должна быть установлена емкость очереди $L=1$.

Информационный вход блока «Очередь» должен соединяться с выходом дискретного сумматора времени для того, чтобы заявки в очереди различались по уровню.



Рисунок 5.25 – Q-схемы а) одноканальной СМО с отказами; б) одноканальной СМО с очередью

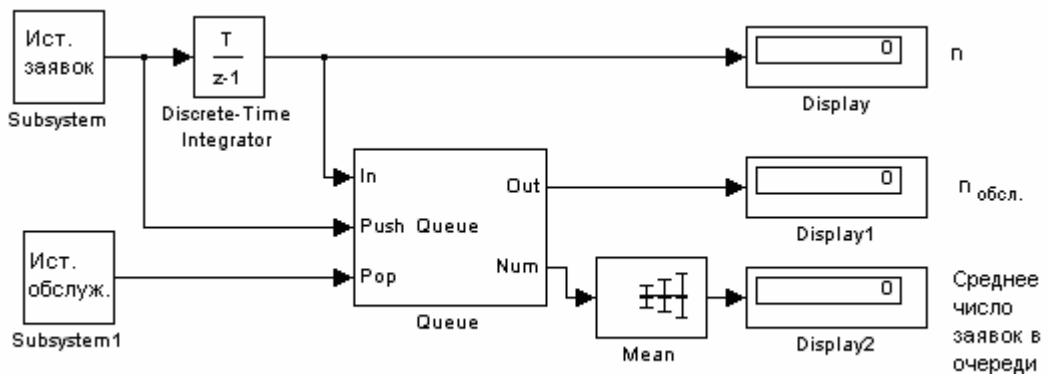


Рисунок 5.26 – Структурная схема ИМ одноканальной СМО с очередью

К входу «Pop» подключают выход источника обслуживаний. Для фиксации среднего числа заявок в очереди к выходу «Number» подключают блок «Mean» (МО) с дисплеем.

В практике моделирования элементы СМО могут соединяться последовательно, параллельно и с помощью обратной связи в случае необходимости подать обслуженную с недостаточным качеством заявку на повторное обслуживание в ту или иную фазу.

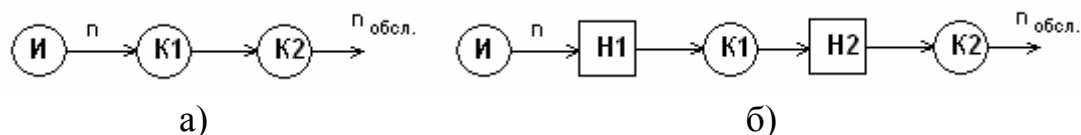


Рисунок 5.27 – Q-схемы а) последовательного соединения каналов; б) последовательного соединения каналов с накопителями

Для осуществления последовательного соединения блоков «Очередь» необходимо информационный выход предыдущего блока соединить как с информационным входом последующего блока «Очередь», так и с входом «Push» (постановка в очередь) с помощью блока производной (рис.5.28).

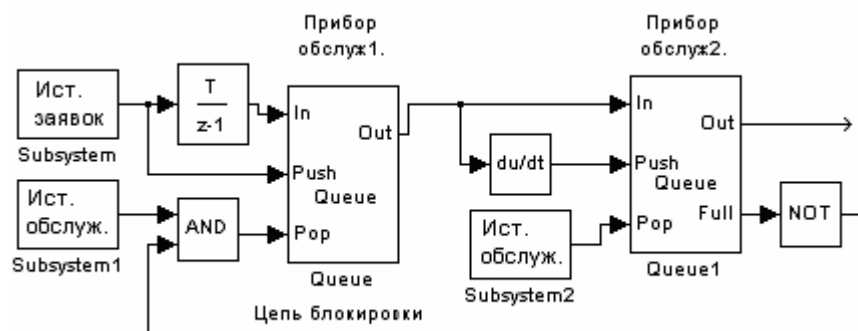


Рисунок 5.28 – Последовательное соединение блоков «Очередь»

Блокировку (задержку обслуженных заявок) в каналах можно выполнять отключением потока обслуживания в предыдущем канале с помощью блока «AND», управляемого сигналом «Full» (полный регистр) через блок «NOT».

Имитация параллельного соединения приборов обслуживания оказывается сложнее и требует реализации процедуры распределения заявок между ними. Она зависит от числа каналов в многоканальной фазе и от заданного закона распределения заявок между каналами. В случае равномерного распределения заявок по каналам соединение одноканальной фазы с многоканальной можно произвести посредством схемы случайного выбора канала (рис. 5.29), обеспечивающей равномерное распределение заявок по трем каналам.

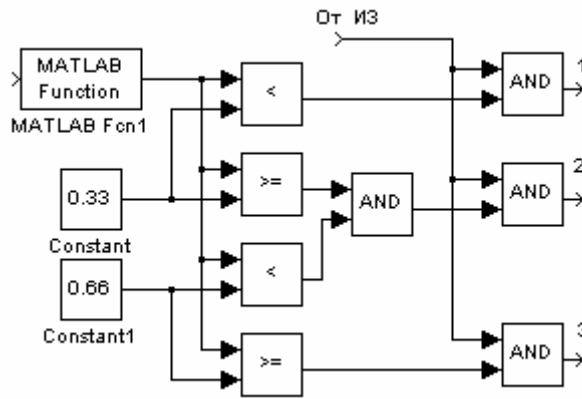
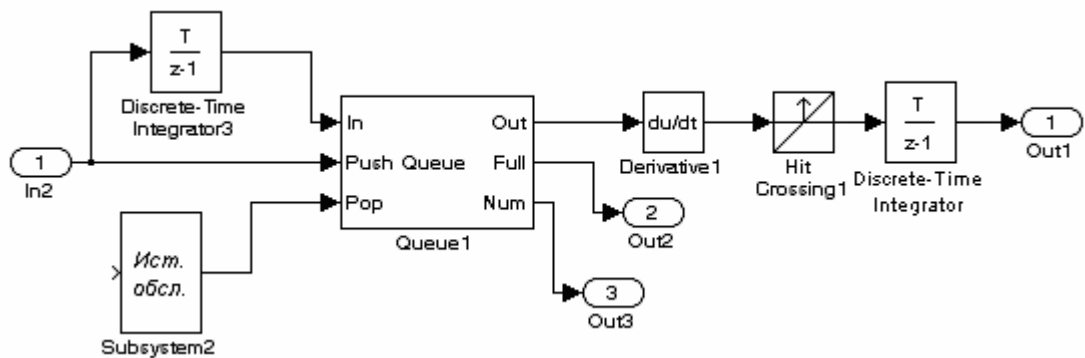


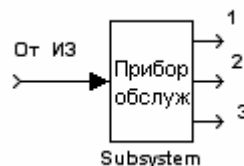
Рисунок 5.29 – Схема случайного выбора канала

Работа схемы заключается в генерации на каждом шаге случайного числа с равномерным распределением в интервале $(0, 1)$, сравнение которого с постоянными числами позволяет выработать один импульс единичной амплитуды и шаговой длительности, подключающий на каждом шаге к магистрали заявок λ только один из трех каналов. Эта схема легко преобразуется как для большего числа каналов, так и для других законов распределения заявок между каналами.

Для упрощения схем ИМ СМО схему прибора обслуживания будем представлять в виде подсистемы (рис. 5.30).



а)



б)

Рисунок 5.30 – Схема прибора обслуживания – а), в виде подсистемы – б)

Рассмотрим пример ИМ трехканальной СМО с отказами.

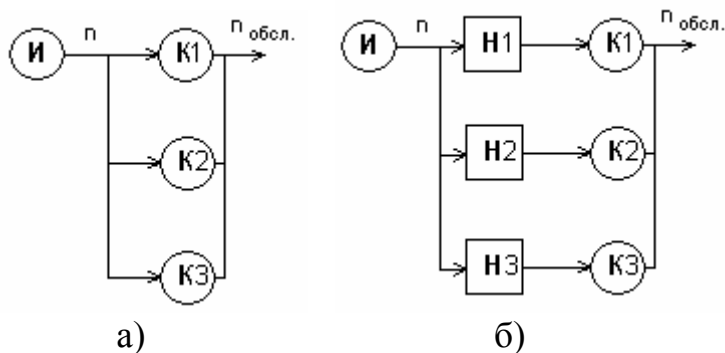


Рисунок 5.31- Q-схемы трехканальной СМО а) без накопителей; б) с накопителями

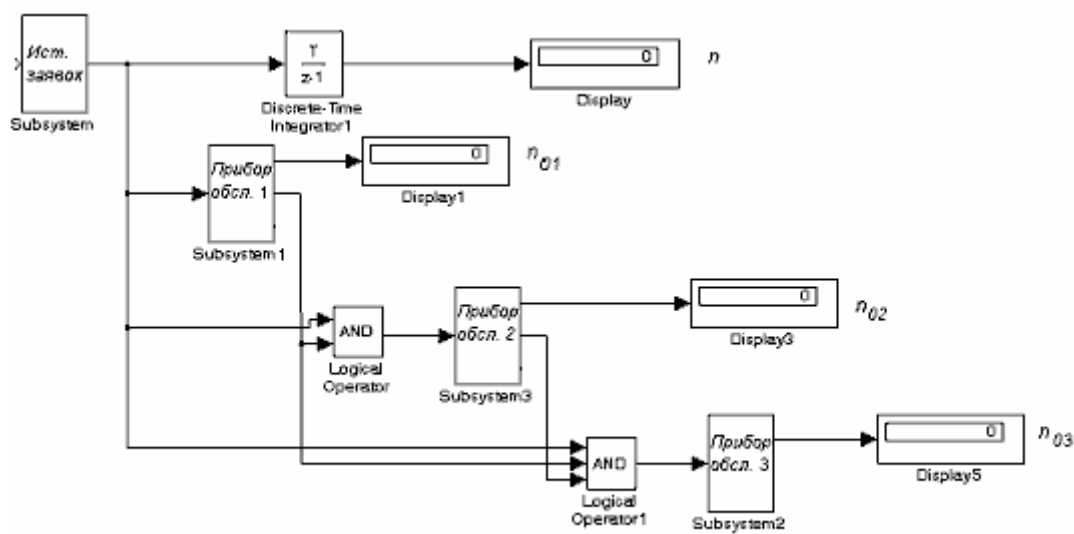
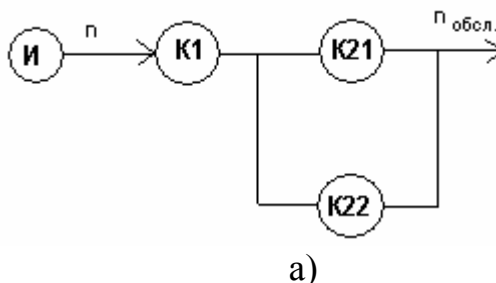
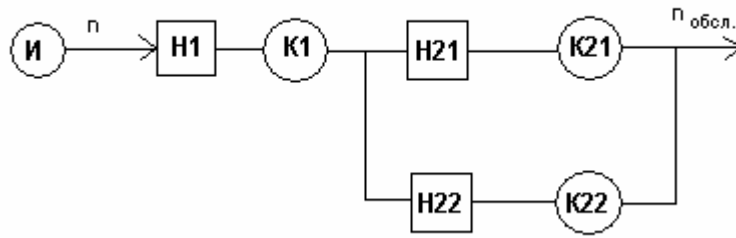


Рисунок 5.32 – схема ИМ трехканальной СМО

В данном случае заявки, сформированные источником заявок, попадут на обслуживание во второй канал только в том случае, если первый канал ($L=1$) занят или первый накопитель полон ($L=k$). Аналогично для третьего канала.

Среда «Matlab» позволяет исследовать СМО не только состоящие из нескольких каналов, но и из нескольких фаз. Рассмотрим пример схемы ИМ двухфазной СМО.





б)

Рисунок 5.33 - Q-схемы двухфазной СМО а) без накопителей;
б) с накопителями

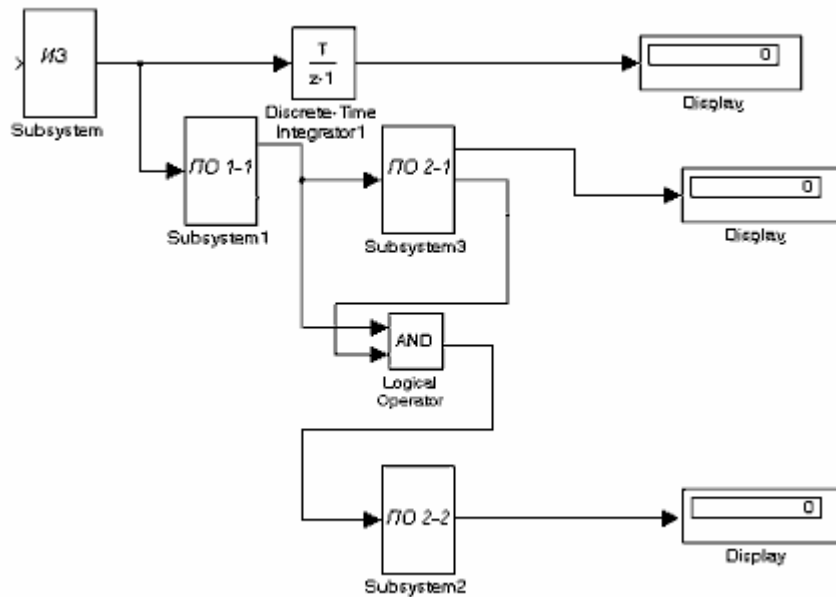


Рисунок 5.34 – схема ИМ двухфазной СМО

Основываясь на выше изложенном материале, можно создавать ИМ СМО большей сложности и исследовать на них процессы, протекающие в реальных системах.

Вопросы для самоконтроля

1. Какими свойствами должен обладать поток событий, чтобы быть простейшим?
2. Какая операция позволяет превратить простейший поток событий в поток событий с одинаковыми интервалами между ними?
3. Чем различаются графы состояний дискретной и непрерывной марковских цепей?
4. Объясните причину появления очередей в системах массового обслуживания.
5. Как выбирается величина шага расчета в детерминированном и асинхронном моделирующих алгоритмах?

6. Почему моделирующие алгоритмы непременно должны удовлетворять рекуррентному правилу?

7. Почему моделирующие алгоритмы систем массового обслуживания должны быть основаны на просмотре элементов схемы в направлении обратном движению заявок в системе?

8. В среде «MATLAB» создайте модель источника простейшего потока событий (заявок) с интенсивностью две заявки в единицу времени.

9. Какие параметры необходимо установить в блоке «Очередь», чтобы он мог выполнять функции канала обслуживания?

10. Как осуществляется параллельное соединение приборов обслуживания, созданных на основе блока «Очередь»?

Список использованных источников

- 1 Бусленко Н.П. Моделирование сложных систем/Н.П.Бусленко. – М.: Наука, 1988.
- 2 Советов Б.Я. Моделирование систем: Учебник для вузов/ Б.Я. Советов, С.А. Яковлев. – М.: Высш. шк., 2001. – 343с.
- 3 Гультаев А. Визуальное моделирование в среде «Matlab»: Учебный курс/А. Гультаев. – СПб.: Питер, 2000.
- 4 Вентцель Е.С. Задачи и упражнения по теории вероятностей: Учебное пособие для студ. Втузов/Е.С. Вентцель, Л.А. Овчаров. – М.: Издательский центр «Академия», 2003. – 448с.
- 5 Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учебное пособие для вузов/В.Е. Гмурман. – М.: Высш. шк., 2003. – 405с.
- 6 Лебедев А.Н. Моделирование в научно-технических исследованиях/ А.Н. Лебедев. – М.: Радио и связь, 1989. – 224с.
- 7 Х. Шенк. Теория инженерного эксперимента. Перевод с английского Е.Г. Коваленко/ Х. Шенк. – М.: Мир, 1972. – 382 с.
- 8 Романцев В.В. Моделирование систем массового обслуживания/ В.В. Романцев, С.А. Яковлев. – СПб.: Поликом, 1995.
- 9 Ермаков С.М. Математический эксперимент с моделями сложных стохастических систем/ С.М. Ермаков, В.Б. Мелос. – СПб.: изд. ГУ, 1993.
- 10 Клейнен Дж. Статистические методы в имитационном моделировании/ Клейнен Дж. – М.: Статистика, 1978.
- 11 Шеннон Р. Имитационное моделирование систем. Искусство и наука/ Шеннон Р. – М.: Мир, 1978.