

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
«Волгоградский государственный технический университет»

**С.А. Фоменков,**

**В.А. Камаев,**

**Ю.А. Орлова**

**МАТЕМАТИЧЕСКОЕ  
МОДЕЛИРОВАНИЕ  
СИСТЕМНЫХ ОБЪЕКТОВ**

*Учебное пособие*

Волгоград - 2014

УДК 519.8

Рецензенты:

Кафедра «Информационные системы и технологии»  
Волгоградского государственного аграрного университета,  
заведующая кафедрой д-р техн. наук, проф. Кочеткова О.Ю.;  
профессор кафедры «Радиофизика»  
Волгоградского государственного университета  
д-р техн. наук, проф. Захарченко В.Д.;  
Ученый Совет Федерального государственного бюджетного учрежде-  
ния науки «Институт проблем точной механики и управления Россий-  
ской академии наук».

Печатается по решению редакционно-издательского совета  
Волгоградского государственного технического университета

**Фоменков, С.А.**

Математическое моделирование системных объектов: учебное пособие  
/ Фоменков С.А., Камаев В.А., Орлова Ю.А.; ВолгГТУ, Волгоград, 2014.-  
340 с.

В учебном пособии представлен теоретический раздел курса «Модели-  
рование систем». Основное внимание уделено типовым математическим  
моделям элементов систем. Предназначено для студентов факультета элек-  
троники и вычислительной техники очной и заочной форм обучения.

Волгоградский государственный  
технический университет, 2014

© Фоменков С.А., Камаев В.А., Ор-  
лова Ю.А.

## ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ.....</b>	<b>7</b>
<b>ГЛАВА 1. ПОНЯТИЕ О МОДЕЛЯХ И МОДЕЛИРОВАНИИ.....</b>	<b>9</b>
1.1. Определение понятия модель. Свойства моделей .....	9
1.2. Классификация моделей .....	13
1.3. Классификация математических моделей по свойствам обобщенного объекта моделирования.....	24
1.4. Адекватность и эффективность математических моделей.....	29
1.5. Методы построения моделей.....	31
1.5.1. Общая логика процесса моделирования .....	31
1.5.2. Аналитические модели .....	37
1.5.3. Идентифицируемые модели .....	40
1.6. Вопросы для самопроверки .....	43
<b>ГЛАВА 2. ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ.....</b>	<b>45</b>
2.1. Постановка задачи идентификации .....	45
2.2. Идентификация моделей с помощью регрессионного метода .....	47
2.2.1. Идентификация статических линейных систем с несколькими входами.....	49
2.2.2. Идентификация нелинейных систем .....	52
2.2.3. Достоверность (адекватность) регрессионной модели.....	53
2.3. Построение моделей идентификации поисковыми методами.....	55
2.4. Вопросы для самопроверки .....	60
<b>ГЛАВА 3. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ НЕОДНОРОДНЫХ СИСТЕМ.....</b>	<b>61</b>
3.1. Математические модели элементов.....	61
3.2. Математические модели взаимодействия элементов сложной системы .....	67
3.3. Вопросы для самопроверки .....	71
<b>ГЛАВА 4. МОДЕЛИРОВАНИЕ ПО СХЕМЕ МАРКОВСКИХ СЛУЧАЙНЫХ ПРОЦЕССОВ .....</b>	<b>72</b>
4.1. Классификация марковских процессов .....	73
4.2. Расчет марковской цепи с дискретным временем.....	74
4.3. Марковские цепи с непрерывным временем .....	76

4.3.1. Уравнения Колмогорова .....	76
4.3.2. Поток событий. Простейший поток и его свойства.....	80
4.3.3. Пуассоновские потоки событий и непрерывные марковские цепи .....	83
4.3.4. Предельные вероятности состояний .....	84
4.3.5. Схема гибели и размножения.....	85
4.4. Вопросы для самопроверки .....	88
<b>ГЛАВА 5. ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ .....</b>	<b>90</b>
5.1. Классификация СМО и их основные характеристики.....	93
5.2. Одноканальная СМО с отказами.....	96
5.3. Многоканальная СМО с отказами .....	99
5.4. Одноканальная СМО с ограничением по длине очереди.....	103
5.5. Многоканальная СМО с ограничением по длине очереди.....	112
5.6. Многоканальная СМО с ограниченным временем ожидания .....	115
5.7. Замкнутые системы массового обслуживания .....	118
5.8. Вопросы для самопроверки .....	124
<b>ГЛАВА 6. АГРЕГАТИВНОЕ ОПИСАНИЕ СИСТЕМ .....</b>	<b>125</b>
6.1. Понятие «агрегат» в теории систем .....	125
6.2. Процесс функционирования агрегата.....	127
6.2.1. Функционирование агрегата общего вида .....	127
6.2.2. Функционирование СМО как агрегата.....	130
6.3. Алгоритм процесса функционирования агрегата общего вида ....	134
6.4. Кусочно-линейные агрегаты.....	141
6.4.1. Понятие о кусочно-линейном агрегате. ....	142
6.4.2. Процесс функционирования КЛА. ....	143
6.4.3. Примеры представления систем в виде КЛА. ....	146
6.5. Вопросы для самопроверки .....	151
<b>ГЛАВА 7. СЕТИ ПЕТРИ.....</b>	<b>152</b>
7.1. Основные определения .....	152
7.2. Сети Петри для моделирования .....	159
7.3. Анализ сетей Петри .....	165
7.3.1. Задачи анализа сетей Петри .....	165
7.3.2. Методы анализа .....	170
7.4. Обобщения сетей Петри.....	179
7.4.1. Временные сети событий (ВСС).....	182
7.4.2. Е-сети .....	184
7.4.3. Комби-сети .....	188
7.5. Вопросы для самопроверки .....	190

## **ГЛАВА 8. МОДЕЛИРОВАНИЕ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ МЕТОДОМ СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ..... 191**

8.1. Основные положения .....	191
8.2. Способы организации единичного жребия .....	197
8.3. Приемы построения и эксплуатации имитационных моделей .....	203
8.3.1. Технология дискретной имитации.....	205
8.3.2. Определение характеристик стационарного случайного процесса по одной реализации.....	212
8.3.3. Методы сбора статистических данных .....	217
8.3.4. Структурная схема имитационной модели для многоканальной СМО с ожиданием .....	224
8.4. Моделирование на универсальных и специализированных языках .....	228
8.5. Современное содержание терминов «имитация», «имитационная модель» .....	237
8.6. Вопросы для самопроверки .....	240

## **ГЛАВА 9. НЕЧЕТКОЕ ОПИСАНИЕ СИТУАЦИЙ..... 241**

9.1. Основные понятия теории нечетких множеств.....	242
9.2. Операции над нечеткими множествами.....	252
9.3. Нечеткие отношения .....	265
9.4. Понятие нечеткой и лингвистической переменной .....	273
9.3. Построение функций принадлежности. Прямые и косвенные методы.....	277
9.4. Арифметические операции над нечеткими числами .....	292
9.5. Сравнение нечетких чисел.....	294
9.6. Примеры использования теории нечетких множеств.....	297
9.6.1. Оценка и упорядочение альтернатив при аддитивности критериев.....	297
9.6.2. Нечеткие сети Петри .....	298
9.7. Вопросы для самопроверки .....	308

## **ГЛАВА 10. КОГНИТИВНЫЕ ПОДХОДЫ К РЕШЕНИЮ СЛАБОСТРУКТУРИРОВАННЫХ И ПЛОХО ФОРМАЛИЗОВАННЫХ ЗАДАЧ..... 310**

10.1 Когнитивные модели.....	310
10.2. Суть когнитивного подхода. Когнитивный анализ.....	314
10.3. Моделирование динамики слабоструктурированной системы .	318
10.4. Аналитическая программная система «СТРАТЕГ».....	325
10.5. Пример. Модель потребления электроэнергии в регионе (модель Робертса) .....	329
10.6. Вопросы для самопроверки .....	333

<b>СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....</b>	<b>334</b>
--	------------

## ВВЕДЕНИЕ

Учебное пособие создано в результате чтения лекций по дисциплине «Моделирование систем» для студентов, обучающихся по направлению 230100 «Информатика и вычислительная техника» факультета электроники и вычислительной техники Волгоградского государственного технического университета (ВолгГТУ).

Материал учебного пособия изложен в соответствии с Федеральным государственным образовательным стандартом по указанному направлению и сложившейся системой его преподавания в ВолгГТУ.

Основное внимание уделено типовым математическим схемам и подходам, недостаточно освещенным в доступной учебной литературе. В настоящее время вышло сравнительно много учебных пособий, учебников и монографий, в совокупности полностью раскрывающих содержание дисциплины «Моделирование систем».

Однако трудно назвать один-два учебника, которые единственно можно рекомендовать для данной дисциплины. Кроме того, многие источники издавались в ограниченном объеме и в библиотеках часто присутствуют в единичных экземплярах или вообще отсутствуют.

Отметим, что все объекты, о которых идет речь в учебном пособии, рассматриваются как системы, что является общепринятым в современной методологии моделирования. Вопросы, касающиеся систем, в настоящем учебном пособии не излагаются, поскольку они входят в содержание дисциплины «Системный анализ».

По аналогичной причине очень кратко представлен переход от математических моделей к компьютерным и свойства компьютерных моделей. Вопросы такого рода рассматриваются в дисциплине «Аналитическое программное обеспечение» и ей подобных дисциплинах.

Статус издания как учебного пособия позволяет (по крайней мере, так считают авторы) широко привлекать без детальных ссылок материалы из изданных уже учебников, учебных пособий и монографий. Все они указаны в списке литературы.

Настоящая работа является вторым изданием учебного пособия, выпущенным в 2006 году. По сравнению с первым изданием существенно расширено содержание глав 6, 8, 10, введена новая глава 9 и внесены ряд изменений в главы 1, 4, 7.

# ГЛАВА 1. ПОНЯТИЕ О МОДЕЛЯХ И МОДЕЛИРОВАНИИ

## 1.1. Определение понятия модель. Свойства моделей

Термин модель неоднозначен и охватывает чрезвычайно широкий круг материальных и идеальных объектов. Признаком, объединяющим такие, казалось бы, несопоставимые объекты как система дифференциальных уравнений математической физики и пара дамских туфель, выставленных на витрине, является их информационная сущность. Любая модель – идеальная или материальная, используемая в научных целях, на производстве или в быту – несет информацию о свойствах и характеристиках исходного объекта (объекта - оригинала), существенных для решаемой субъектом задачи. Модели – отражение знаний об окружающем мире.

Пусть имеется некоторая конкретная система. Лишь в единичных случаях мы имеем возможность провести с самой этой системой все интересующие нас исследования. С ростом сложности системы возможности натурального эксперимента резко падают. Он становится дорогим, трудоемким, длительным по времени, в слабой степени вариативным. Тогда предпочтительнее работа с моделью. В ряде же случаев мы вообще не имеем возможности наблюдать систему в интересующем нас состоянии. Например, разбор аварии на техническом объекте приходится вести по ее протокольному описанию. Специалист по электронной технике будет изучать большинство типов ЭВМ по литературе, и только часть из них опробует на практике. В этих примерах доступна лишь модель, но это не мешает нам эффективно познавать систему.

Рассмотрение вместо самой системы (явления, процесса, объекта) ее модели практически всегда несет идею упрощения. Мы огрубляем представления о реальном мире, так как оперировать категорией модели экономичнее, чем действительностью. Но вопрос выделения и формальной фиксации тех особенностей, которые существенны для целей рассмотрения, весьма непросто. Известно большое количество удачных моделей, составляющих предмет гордости человеческой мысли, — от конечноэлементной модели в прикладных задачах математической физики до модели генетического кода. Однако велико количество процессов и явлений, для которых на настоящий момент нет

удовлетворительного описания. В области техники положение с моделированием можно считать удовлетворительным, но и здесь имеются «узкие» места, связанные с плохо определяемыми параметрами, коэффициентами, а также слишком грубые описания.

*Определение.* Модель в общем смысле есть создаваемый с целью получения и (или) хранения информации специфический объект (в форме мысленного образа, описания знаковыми средствами либо материальной системы), отражающий свойства, характеристики и связи объекта – оригинала произвольной природы, существенные для задачи, решаемой субъектом.

Модель есть материально или теоретически сконструированный объект, который заменяет (представляет) объект исследования в процессе познания, находится в отношении сходства с последним (изоморфизм, аналогия, физическое сходство и т.п.) и более удобен для исследования. Как следствие, изучение модели и выполнение над ней операций позволяют получить информацию о реальном объекте исследования.

Заметим, что в настоящее время в качестве объекта-оригинала практически всегда рассматривается некоторая система.

Под системой понимается совокупность объектов, компонентов или элементов произвольной природы, образующая некоторую целостность в том или ином контексте. Определяющим принципом рассмотрения некоторой совокупности объектов как системы является появление у нее новых свойств, которые не имеют составляющие ее элементы. Системы различной физической природы окружают нас повсеместно – это и конкретные предметы и объекты: солнечная система, человек, персональный компьютер, автомобиль, самолет, аэропорт. Это и более абстрактные сущности, такие как компьютерная программа, естественный язык, коммерческая фирма, культура, политика, наука, экономика. Наиболее ортодоксальная точка зрения предполагает, что все окружающие нас предметы являются системами.

Непосредственно из структуры принятого определения вытекают ряд общих свойств моделей, которые обычно принимаются во внимание в практике моделирования.

1. Модель представляет собой «четырёхместную конструкцию», компонентами которой являются субъект; задача, решаемая субъектом; объект-оригинал и язык описания или способ воспроизведения модели. Осо-

бую роль в структуре обобщенной модели играет решаемая субъектом задача. Вне контекста задачи или класса задач понятие модели не имеет смысла.

2. Каждому материальному объекту, вообще говоря, соответствует бесчисленное множество в равной мере адекватных, но различных по существу моделей, связанных с разными задачами.

3. Паре задача-объект тоже соответствует множество моделей, содержащих в принципе одну и ту же информацию, но различающихся формами ее представления или воспроизведения.

4. Модель по определению всегда является лишь относительным, приближенным подобием объекта-оригинала и в информационном отношении принципиально беднее последнего. Это ее фундаментальное свойство.

5. Произвольная природа объекта-оригинала, фигурирующая в принятом определении, означает, что этот объект может быть материально-вещественным, может носить чисто информационный характер и, наконец, может представлять собой комплекс разнородных материальных и информационных компонентов. Однако независимо от природы объекта, характера решаемой задачи и способа реализации модель представляет собой информационное образование.

6. Частным, но весьма важным для развитых в теоретическом отношении научных и технических дисциплин является случай, когда роль объекта-моделирования в исследовательской или прикладной задаче играет не фрагмент реального мира, рассматриваемый непосредственно, а некий идеальный конструкт, т.е. по сути дела другая модель, созданная ранее и практически достоверная. Подобное вторичное, а в общем случае *n*-кратное моделирование может осуществляться теоретическими методами с последующей проверкой получаемых результатов по экспериментальным данным, что характерно для фундаментальных естественных наук. В менее

развитых в теоретическом отношении областях знания (биология, некоторые технические дисциплины) вторичная модель обычно включает в себя эмпирическую информацию, которую не охватывают существующие теории.

Свойства любой модели таковы:

- конечность: модель отображает оригинал лишь в конечном числе его отношений и, кроме того, ресурсы моделирования конечны;
- упрощенность: модель отображает только существенные стороны объекта;
- приближительность: действительность отображается моделью грубо или приблизительно;
- адекватность: модель успешно описывает моделируемую систему (правильное соответствие изучаемому объекту относительно выбранной системы его свойств);
- информативность: модель должна содержать достаточную информацию о системе - в рамках гипотез, принятых при построении модели и должна давать возможность получить новую информацию);
- потенциальность – предсказательность с позиций возможности получения новых знаний об исследуемом объекте;
- полнота – в модели должны быть учтены все основные связи и отношения, необходимые для обеспечения цели моделирования;
- адаптивность - модель должна быть приспособлена к изменениям внешней среды и внутренних параметров.

Целями моделирования могут являться:

- осмысление действительности (познание и разработка теории исследуемых систем);
- постановка над моделью экспериментов с последующей интерпретацией их результатов применительно к моделируемой системе;
- прогнозирование будущего поведения системы (выходных данных, ситуаций, состояний системы);
- проектирование систем;
- управление системами;

- общение с другими лицами, общественными группами, устройствами обработки информации;
- обучение и тренаж специалистов.

## **1.2. Классификация моделей**

Каждая модель характеризуется тремя признаками:

- принадлежностью к определённому классу задач (по классам задач);
- указанием класса объектов моделирования (по классам объектов);
- способом реализации (по форме представления и обработки информации).

Рассмотрим более подробно последний вид классификации. По этому признаку модели делятся на материальные и идеальные.

Материальные модели:

- геометрически подобные масштабные, воспроизводящие пространственно-геометрические характеристики оригинала безотносительно его субстрату (макеты зданий и сооружений, учебные муляжи и др.);
- основанные на теории подобия субстратно подобные, воспроизводящие с масштабированием в пространстве и времени свойства и характеристики оригинала той же природы, что и модель, (гидродинамические модели судов, продувочные модели летательных аппаратов);
- аналоговые приборные, воспроизводящие исследуемые свойства и характеристики объекта оригинала в моделирующем объекте другой природы на основе некоторой системы прямых аналогий (разновидности электронного аналогового моделирования).

Рассмотрим более подробно два последних пункта. Для парохода правильный выбор обводов, подбор гребного винта и согласование с характеристиками винта и корпуса мощности и скорости вращения вала – пробле-

ма №1. По существу речь идет о необходимости оптимизировать взаимодействие системы корпус – винт – двигатель с обтекающей судно жидкой средой по критерию максимального КПД. Решение проблемы опытным путем невозможно по очевидным экономическим соображениям, не поддается она и теоретическому решению. Выход был найден на пути синтеза теории и масштабного гидродинамического моделирования, т.е. экспериментальное исследование малых геометрически подобных моделей проектируемых судов в специальных бассейнах на основе теории подобия. Теория обеспечивала возможность достоверного переноса данных, полученных на модели, на «натуру», на свойства и характеристики реального, но еще не существующего судна. И сегодня методы масштабного физического моделирования сохраняют свое значение.

Аналоговое моделирование основано на том, что свойства и характеристики некоторого объекта воспроизводятся с помощью модели иной, чем у оригинала физической природы. Целый ряд явлений и процессов существенно различной природы описывается аналогичными по структуре математическими выражениями. Описываемые аналогичными математическими структурами разнородные объекты можно рассматривать как пару моделей, которые с точностью до свойств, учитываемых в математическом описании, взаимно отображают друг друга, причем коэффициенты, связывающие соответственные (сходственные) параметры, являются в этом случае размерными величинами.

$$1. \frac{\partial T}{\partial t} = \alpha * \frac{\partial^2 T}{\partial x^2}$$

$$2. \frac{\partial C}{\partial t} = D * \frac{\partial^2 C}{\partial x^2}$$

$$3. \frac{\partial u}{\partial t} = \frac{1}{RC} * \frac{\partial^2 U}{\partial x^2}$$

1-уравнение теплопроводности (закон Фурье), 2-уравнение диффузии (закон Фика), 3-уравнение электропроводности (закон Ома).

Идеальные модели:

- неформализованные модели, т.е. системы представлений об объекте оригинале, сложившиеся в человеческом мозгу;

- частично формализованные:

- вербальные – описание свойств и характеристик оригинала на некотором естественном языке (текстовые материалы проектной документации, словесное описание результатов технического эксперимента);

- графические иконические – черты, свойства и характеристики оригинала, реально или хотя бы теоретически доступные непосредственно зрительному восприятию (художественная графика, технологические карты);

- графические условные – данные наблюдений и экспериментальных исследований в виде графиков, диаграмм, схем;

- вполне формализованные (математические) модели.

Основное отличие этого типа моделей от остальных состоит в вариативности — в кодировании одним знаковым описанием огромного количества конкретных вариантов поведения системы. Так, линейные дифференциальные уравнения с постоянными коэффициентами описывают и движение массы на пружине, и изменение тока в колебательном контуре, и измерительную схему системы автоматического регулирования, и ряд других процессов. Однако еще более важно то, что в каждом из этих описаний одни и те же уравнения в буквенном (а вообще говоря, и в числовом) виде соответствуют бесконечному числу комбинаций конкретных значений параметров. Скажем, для процесса механических колебаний — это любые значения массы и жесткости пружины.

В знаковых моделях возможен дедуктивный вывод свойств, количество следствий в них обычно более значительно, чем в моделях других ти-

пов. Они отличаются компактной записью, удобством работы, возможностью изучения в форме, абстрагированной от конкретного содержания. Все это позволяет считать знаковые модели наивысшей ступенью и рекомендовать стремиться к такой форме моделирования.

В сравнении с натурным экспериментом математическое моделирование имеет следующие преимущества:

1. Экономичность (в частности, сбережение ресурса реальной системы).
2. Возможность моделирования гипотетических (то есть не реализованных в природе) объектов и процессов.
3. Возможность реализации режимов, опасных или труднопроизводимых в природе (критический режим ядерного реактора, работа системы ПРО).
4. Возможность изменения масштаба времени.
5. Легкость многоаспектного анализа.
6. Сжатие информации и ее единообразное представление, способствующие усмотрению основных закономерностей.
7. Отсутствие погрешностей измерения, которые в реальных условиях могут замаскировать искомый эффект.
8. Наличие строго сформулированных правил, позволяющее:
  - вскрыть ложность некоторых предубеждений;
  - заменить содержательные рассуждения формальным преобразованием выражений – в частности, выполнить оптимизацию;
  - выявить глубинные (сущностные) свойства и отношения и как следствие получить большую прогностическую силу и построить разнообразные аналогии.
9. Универсальность технического обеспечения проводимой работы (ЭЦВМ, системы программирования и пакеты прикладных программ специального назначения).

Часто в качестве основания для классификации идеальных моделей берется вид языка, на котором они формулируются:

- содержательная модель формулируется на естественном языке;

- формальная модель воплощается с помощью одного или нескольких формальных языков (например, языков математических теорий или языков программирования).

Если в естественно-научной среде моделирование нередко считают только математическим, то в гуманитарной сфере чаще используют содержательные модели. Чтобы разобраться во взаимоотношениях моделей различного типа, рассмотрим рис. 1.1.

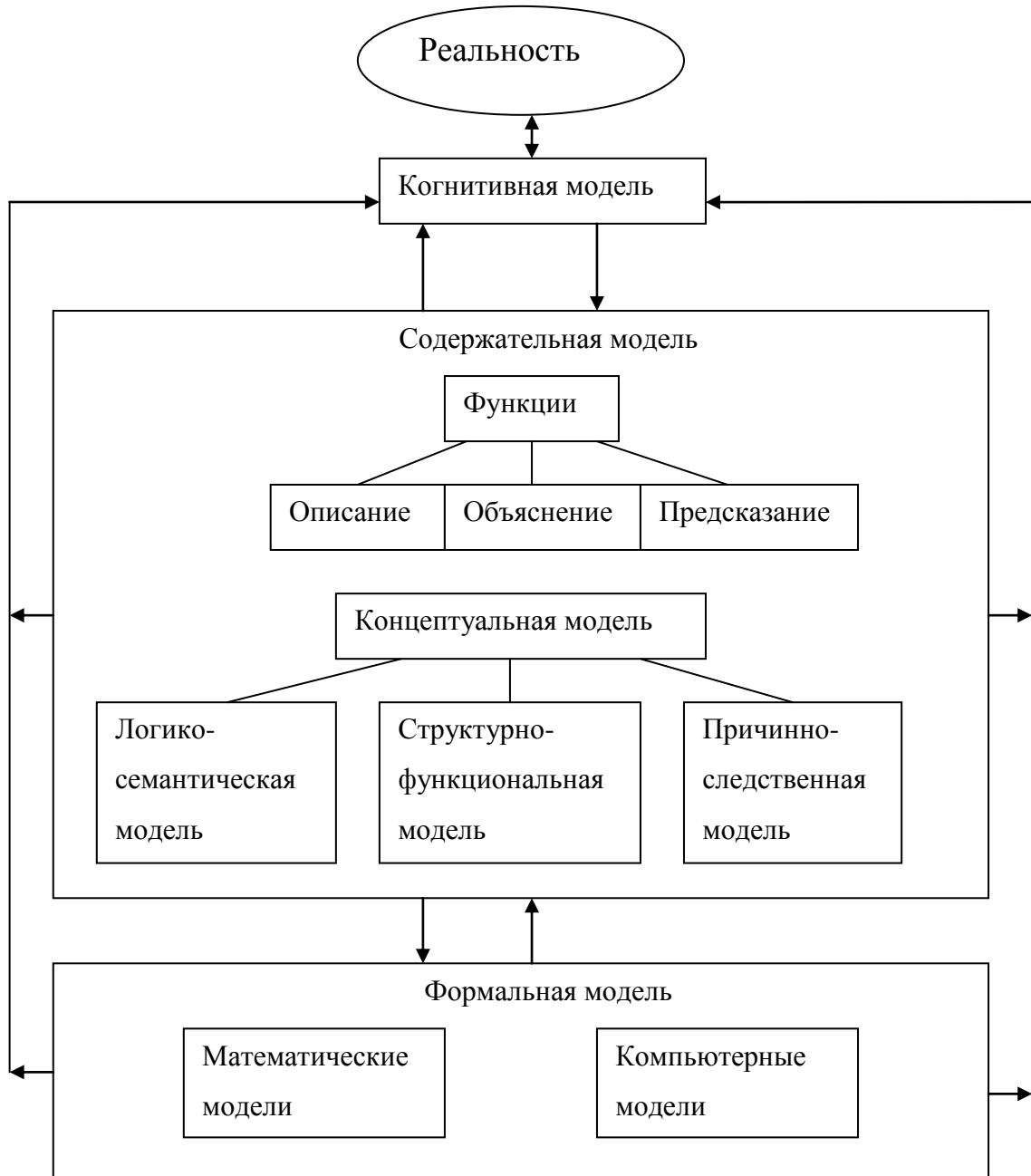


Рис. 1.1. Типы идеальных моделей

Любая модель является, в конечном счете, моделью объекта, фрагмента реальности (верхний уровень на указанной схеме). Наблюдая за объектом, индивид формирует в голове некий мысленный образ объекта, который будем называть когнитивной моделью. В данной случае когнитологи используют также термин «ментальная» модель, понимая под когнитивной моделью модель взаимодействия с объектом. В ряде работ термин ментальная модель относится только к индивиду, а термин когнитивная модель используется при описании других видов когнитивных систем.

Формируя когнитивную модель объекта, индивид, как правило, стремится ответить на определенные, конкретные вопросы, поэтому от бесконечно сложной реальности отсекается все ненужное с целью получения более компактного и лаконичного описания объекта. Когнитивная модель объекта формируется на основе «картины мира» индивида – особенностей его восприятия, установок, ценностей, интересов.

Следующий этап моделирования – построение содержательной модели. При этом нельзя утверждать, что эта модель является просто вербализированной копией когнитивной модели. Дело в том, что когнитивная модель может содержать элементы, которые индивид не может или не хочет сформулировать. Возможна и обратная ситуация. Если содержательная модель сформулирована кем-то другим или является продуктом коллективного творчества, то ее интерпретация, уровень понимания, степень доверия по отношению к отдельным элементам модели могут существенно варьироваться.

Построение содержательной модели позволяет получить новую информацию о поведении объекта, выявить взаимосвязи и закономерности, которые не удастся обнаружить при других способах анализа. По функ-

циональному признаку содержательные модели подразделяются на описательные, объяснительные и прогностические.

Описательной моделью можно назвать любое описание объекта. Объяснительные модели призваны ответить на вопрос, почему что-либо происходит. Прогностические модели должны описывать будущее поведение объекта, т.е. отвечать на вопрос, к каким изменениям приводит то или иное воздействие на объект. Заметим, что прогностические модели совсем не обязаны включать в себя объяснительные модели. Нередко удается получить удовлетворительный прогноз на основе эмпирических обобщений, т.е. используя только данные описательной модели.

Концептуальной моделью называется содержательная модель, при формулировке которой используются теоретические концепции и конструкции данной предметной области знания. В более широком смысле под концептуальной моделью понимают содержательную модель, базирующуюся на определенной концепции или точке зрения. Формулировка концептуальной модели нередко представляет собой достижение определенного уровня абстрагирования на пути от предварительного описания объекта к его формальной модели.

Концептуальные модели воплощаются либо в чисто вербальной форме, либо в смешанном вербально-визуальном представлении. Выделяют три вида концептуальных моделей: логико-семантические, структурно-функциональные и причинно-следственные.

Элементами логико-семантической модели являются все утверждения и факты, включенные в вербальное описание объекта. Анализ такой модели осуществляется средствами логики с привлечением знаний, накопленных в данной предметной области.

При построении структурно-функциональных моделей объект обычно рассматривается как целостная система, которую следует расчленить на

составные части, компоненты, элементы, подсистемы. Части системы связываются структурными отношениями, описывающими подчиненность, логическую и временную последовательность решения отдельных задач. Структурно-функциональные связи целесообразно визуализировать в виде разного рода схем, карт и диаграмм. С помощью таких схем удобно анализировать служебную роль и назначение отдельных подсистем по отношению к целому, оценивать взаимозависимость отдельных элементов.

Причинно-следственные модели часто используются для объяснения и прогнозирования поведения объекта. В отличие от структурно-функциональных моделей они ориентированы в основном на описание динамики исследуемых процессов, при этом время далеко не всегда учитывается в явном виде. Дело в том, что причинно-следственные отношения, связывающие элементы модели, подразумевают развитие процессов и событий во времени. Элементами таких моделей могут быть понятия, категории, концепты, конструкторы, показатели, индикаторы и переменные, описывающие поведение исследуемого объекта. Элемент причинно-следственных содержательных моделей лучше называть фактором (от лат. factor – делающий, производящий) – причиной, движущей силой исследуемых процессов и явлений. Термин «фактор» напоминает исследователю, что после расчленения объекта на части и компоненты необходимы следующие операции:

- выявление главных взаимосвязей;
- определение того, как изменение одних факторов влияет на другие компоненты модели;
- понимание того, как будет в целом функционировать собранный механизм, будет ли он адекватно описывать динамику интересующих исследователя параметров.

Нередко говорят об изучении причинно-следственного механизма явления, отбрасывая при этом все лишнее и оставляя только минимальный набор факторов, взаимодействие которых позволяет понять устройство и спрогнозировать поведение изучаемого объекта.

В процессе построения, изучения и совершенствования содержательной модели когнитивная модель непрерывно модифицируется и усложняется. В гуманитарных науках цикл моделирования на этом обычно и заканчивается, но в некоторых случаях модель удается формализовать до такой степени, что становится возможным построение и изучение формальной модели объекта. Формальные модели в свою очередь делятся на две группы: математические и компьютерные.

Создание формальной модели дает возможность постичь сущность исследуемых явлений, выявить основные взаимосвязи и закономерности. Использование формальных средств анализа позволяет изучить поведение модели, получить новые, неочевидные результаты. В любом случае результаты формального моделирования используются для уточнения содержательной модели и, главное, когнитивной модели.

Введем «прагматическое» определение математической модели, удобное для практических приложений. Для этого используем хорошо известное из кибернетики представление объекта в виде «черного ящика».

Первым шагом к осознанному построению модели во всех случаях является уяснение и четкая формулировка исследования или иной задачи, ради решения которой осуществляется моделирование. Этот шаг базируется на содержательном анализе исходной проблемы, предполагает сбор и осмысление всех уже имеющихся данных, относящихся к задаче. Следующий шаг, с которого начинается процедура собственно моделирования, заключается в определении границ объекта, подлежащего модельному описанию и исследованию с целью решения задачи. Здесь возможен очень

широкий диапазон различных ситуаций (зависит от характера задачи, степени сложности и изученности). Будем считать, что в соответствии с имеющейся информацией мы приняли некоторую гипотезу о границах объекта, подлежащего модельному исследованию. Исходя из принципа всеобщей взаимосвязи и взаимозависимости, можно утверждать, что в общем случае выявленный объект, с одной стороны, подвергается воздействиям со стороны окружающей среды, с другой – сам воздействует на эту среду, изменяя её состояние. Связи среда – объект будем именовать, как это принято, входными воздействиями или входами  $X$  (часто вводят разделение входных воздействий на управления ( $U$ ) и возмущения ( $V$ )), а воздействия объект – среда ( $Y$ ) – выходными.

Очевидно, что достаточно полный (с точки зрения решаемой задачи) учет входных и выходных связей объекта с более широкой системой (средой), компонентом которой он является, есть необходимое условие правомерности выделения объекта из среды. Каждая упущенная исследованием существенная связь создает угрозу того, что состояние и свойства выявленного объекта уже не будут соответствовать тем, которые имели место в исходной реальной системе и модель, базирующаяся на подобном представлении, окажется заведомо неадекватной. С другой стороны, по практическим соображениям в модели желательно учитывать возможно меньшее число факторов, ибо её сложность и громоздкость являются не менее серьезными недостатками, чем неполнота. Разрешение данного противоречия, т.е. выбор подлежащих учету в модели существенных входных и выходных воздействий и абстрагирование от прочих, предположительно незначимых, представляет собой весьма ответственный момент при построении любой модели, т.к. решающим образом влияет на её качество и эффективность. Здесь необходимо глубокое понимание существа решаемой задачи, тщательное изучение воспроизводимой в модели исходной реальной сис-

темы, необходим опыт и эвристические способности. Если моделируемый объект представляет собой реально существующую материальную систему, его связями, очевидно, являются также вполне реальные материальные факторы: силы различной природы, пространственные перемещения с их производными, потоки вещества, потоки энергии, а в некоторых случаях потоки информации. Все они должны быть исследованы и описаны в качественном и количественном отношении, оценены посредством «числа и меры», после чего превращаются в информационные конструкты и приобретают статус переменных модели.

Использование математической модели в современном смысле слова не связано с материальным воспроизведением подлежащих исследованию свойств и характеристик объекта и не предполагает экспериментальных процедур. Объект, описанный на языке математики, представляется некоторой математической структурой (дифференциальными или конечно-разностными уравнениями, передаточной функцией, графом и т.п.) с определенными параметрами, а процесс исследования (так называемое решение математической модели) заключается в применении к этой структуре совокупности математических преобразований и операций в соответствии с некоторым алгоритмом. Результатом вычислительного процесса является новая информация об объекте, разумеется, в той части его свойств, которые нашли отражение в исходном математическом описании. Возможности современных ЭВМ и программных средств позволяют исследовать эти свойства при всевозможных вариациях параметров, входящих в исходную модель, определять присущие ей вероятностно-статистические характеристики, находить значения параметров, оптимальных по тому или иному критерию и решать множество других самых разнообразных задач.

Под словами “модельное описание” или “модель” понимается математически формализованное описание некоторого явления или объекта в терминах определенной группы его характеристик. Математическая модель сложных управляемых процессов

содержит очень много величин различной природы. Все эти величины естественным образом можно разделить на три группы:

- к первой группе относятся величины, которые принято называть эндогенными (внутренними), или фазовыми; они являются искомыми величинами, т. е. подлежат определению, вычислению в силу связей модели;
- ко второй группе относятся так называемые экзогенные (внешние) величины, они полагаются известными в рамках данной модели;
- к третьей группе относятся управления — величины, находящиеся в распоряжении органов управления, с помощью которых можно оказать влияние на течение процесса.

Само слово “модель” означает совокупность связей между всеми этими величинами. Если эта совокупность связей позволяет определить на данном отрезке времени все эндогенные величины при условии, что на нем заданы управления, экзогенные величины, а также начальные для этого отрезка (и, возможно, граничные – в пространственном смысле) значения фазовых переменных, то модель называется замкнутой.

Разделение на внешние и внутренние величины можно выполнить не единственным образом, оно является в известной мере условным и связано со способом использования модели и целями моделирования.

### **1.3. Классификация математических моделей по свойствам обобщенного объекта моделирования**

Будем предполагать, что возможно, хотя бы в принципе, установить и на некотором языке описания (например, средствами математики) охарактеризовать зависимость каждой из выходных переменных от входных. Связь между входными и выходными переменными моделируемого объекта в принципе может характеризоваться графически, аналитически, т.е. посредством некоторой формулы общего вида, или алгоритмически. Независимо от формы представления конструкта, описывающего эту связь, будем именовать его оператором вход-выход и обозначать через  $\hat{A}$ .

Пусть  $M=M(X,Y,Z)$ , где  $X$  - множество входов,  $Y$  - выходов,  $Z$  - состояний системы. Схематически можно это изобразить:  $X \rightarrow Z \rightarrow Y$ .

Рассмотрим теперь наиболее существенные с точки зрения моделирования внутренние свойства объектов разного класса. При этом придется использовать понятие структура и параметры моделируемого объекта. Под структурой понимается совокупность учитываемых в модели компонентов и связей, содержащихся внутри объекта, а после формализации описания объекта – вид математического выражения, которое связывает его входные и выходные переменные (например:  $y=au+bv$ ). Параметры представляют собой количественные характеристики внутренних свойств объекта, которые отражаются принятой структурой, а в формализованной математической модели они суть коэффициенты (постоянные переменные), входящие в выражения, которыми описывается структура (а и b).

Первое свойство *непрерывность* и *дискретность*. Все те объекты, переменные которых (включая, при необходимости, время) могут принимать несчетное множество сколь угодно близких друг к другу значений называются непрерывными или континуальными. Подавляющее большинство реальных физических и теоретических объектов, состояние которых характеризуется только макроскопическими физическими величинами (температура, давление, скорость, ускорение, сила тока, напряженность электрического или магнитного полей и т.д.) обладают свойством непрерывности. Математические структуры, адекватно описывающие такие объекты, тоже должны быть непрерывными. Поэтому при модельном описании таких объектов используется главным образом, аппарат дифференциальных и интегро-дифференциальных уравнений. Дифференциальные уравнения как инструмент модельного описания физических и технических объектов настолько широко распространены в приложениях, что некоторые специалисты, главным образом инженеры, только их и рассматривают, как полно-

ценные модели. Эта точка зрения, естественно, не имеет под собой никаких оснований. Объекты, переменные которых могут принимать, практически всегда конечное число наперед известных значений, называются дискретными. Характерными примерами дискретных объектов являются релейно-контактные и бесконтактные переключательные схемы, коммутационные системы АТС, конечные автоматы, ЭВМ. Основой формализованного описания дискретных объектов является аппарат математической логики (логические функции, аппарат булевой алгебры, алгоритмические языки). В связи с развитием ЭВМ дискретные методы анализа получили широкое распространение также для описания и исследования непрерывных объектов.

Непрерывность или дискретность. Это свойство выражается в структуре множеств (совокупностей), которым принадлежат параметры состояния, параметр процесса и входы, выходы системы. Таким образом, дискретность множеств  $Z, Y, T, X$  ведет к модели, называемой дискретной, а их непрерывность — к модели с непрерывными свойствами. Дискретность входов (импульсы внешних сил, ступенчатость воздействий и др.) в общем случае не ведет к дискретности модели в целом. Важной характеристикой дискретной модели является конечность или бесконечность числа состояний системы и числа значений выходных характеристик. В первом случае модель называется дискретной конечной. Дискретность модели также может быть как естественным условием (система скачкообразно меняет свое состояние и выходные свойства), так и искусственно внесенной особенностью. Типичный пример последнего — замена непрерывной математической функции на набор ее значений в фиксированных точках.

Следующее свойство модели — *детерминированность* или *стохастичность*. Если в модели среди величин имеются случайные, т. е. определяемые лишь некоторыми вероятностными характеристиками, то модель называется стохастической (вероятностной, случайной). В этом случае и все результаты, полученные при рассмотрении модели, имеют стохастический характер и должны быть соответственно интерпретированы.

Здесь подчеркнем, что с точки зрения практики граница между детерминированными и стохастическими моделями выглядит расплывчатой. Так, в технике про любой размер или массу можно сказать, что это не точное значение, а усредненная величина типа математического ожидания, в связи с чем и результаты вычислений будут представлять собой лишь математические ожидания исследуемых величин. Однако такой взгляд представляется крайним. Удобный практический прием состоит в том, что при малых отклонениях от фиксированных значений модель считается детерминированной, а отклонение результата исследуется методами оценок или анализа ее чувствительности. При значительных же отклонениях применяется методика стохастического исследования.

Свойства *сосредоточенности* или *распределённости* характеризуют объекты с точки зрения роли, которую играет в их модельном описании пространственная протяженность (на фоне скорости распространения физических процессов). Если пространственной протяженностью объекта можно пренебречь и считать, что независимой переменной является только время (протекающих в нем процессов), принято говорить об объекте с сосредоточенными параметрами. К числу таких объектов, которые описываются (в случае детерминированности и непрерывности) обыкновенными дифференциальными уравнениями, относится подавляющее большинство механизмов, машин и вообще локальных технических устройств (расстояния между компонентами практически не влияют на исследуемые свойства и характеристики). В пространственно протяженных объектах адекватное описание требует учета не только времени, но и пространственных координат. В таком случае говорят о классе объектов с распределенными параметрами. Типичными примерами одномерных объектов с распределенными параметрами служат всевозможные «длинные линии»: проводная линия связи, описываемая так называемым телеграфным уравнением, длинные трубопроводы, технологические линии в непрерывном производстве. Пример моделей двумерного объекта с распределенными параметрами являются собой синоптические карты. Наконец, электромагнитное поле с его обоб-

щенной математической моделью - уравнениями Максвелла - представляет собой классический пример трехмерного объекта с распределенными параметрами. Непрерывные и детерминированные объекты с распределенными параметрами описываются дифференциальными уравнениями в частных производных.

*Статические* и *динамические* модели. Статические модели относятся к объектам, практически неизменяющимся во времени или рассматриваемым в отдельные временные сечения. Динамические модели воспроизводят изменения состояний («движение») объекта с учетом как внешних, так и внутренних факторов.

Для динамических моделей часто вводят понятия *стационарность* и *нестационарность*. Чаще всего стационарность выражается в неизменности во времени некоторых физических величин: стационарным является поток жидкости с постоянной скоростью, стационарна механическая система, в которой силы зависят только от координат и не зависят от времени.

Под стационарным объектом, в более общем смысле, подразумевают неизменность структуры и параметров объекта. Поэтому он описывается выражением, которое включает в себе только постоянные коэффициенты. Нестационарность может иметь место относительно параметров, относительно структуры и одновременно. Чаще имеет место нестационарность относительно параметров, т.е. рассматривается объект с переменными коэффициентами, что усложняет исследование. Общей теории и специального математического аппарата для описания существенно нестационарных объектов переменной структуры еще не существует. Исследование таких объектов проводится на основе некоторых методов прикладного системного анализа, которые сочетают формализованные математические процедуры с эвристикой и здравым смыслом, а также широко используют прием декомпозиции и последующего объединения частных решений.

С точки зрения общности методов анализа, возможностей математического аппарата и трудоемкости исследования чрезвычайно существенно деление объектов на *линейные* и *нелинейные*. Для первых справедлив

принцип суперпозиций, когда каждый из выходов объекта характеризуется некоторой линейной формулой, связывающих его со значениями соответственных входных переменных. С точки зрения математического аппарата линейность объекта относительно переменных означает, что среди коэффициентов, входящих в его математическое описание, отсутствуют величины, зависящие от переменных, их производных и интегралов. Если коэффициенты не зависят и от времени, то это самый благоприятный и наиболее распространенный в технических приложениях случай: описание объекта в классе линейных стационарных моделей.

Линейность (нелинейность) обычно расшифровывается как линейная (нелинейная) зависимость от входов операторов состояний или выходов. Линейность может являться как естественным, хорошо соответствующим природе, так и искусственным (вводимым для целей упрощения) свойством модели.

#### **1.4. Адекватность и эффективность математических моделей**

Вопрос о необходимой и достаточной степени соответствия объекту – оригиналу или адекватности модели относится к числу важнейших в сфере модельной методологии. Под эффективностью понимают практическую полезность. Процесс моделирования неизбежно протекает в условиях диалектического взаимодействия двух противостоящих друг другу тенденций. С одной стороны, исследователь всегда стремится к возможно более полному и точному воспроизведению в модели свойств и характеристик объекта. Неизбежным следствием такого подхода является рост сложности, которая проявляется в числе переменных, числе учитываемых связей и влияний, повышении требования к точности исходных данных и т.д. Именно эта сторона дела – требование полноты соответствия модели объекту – оригиналу акцентируется в большинстве философских работ и даже рассматривается некоторыми авторами как мера совершенства модели.

Однако практика показала неопровержимо: эффективность модели находится в обратной зависимости от её сложности, быстро убывая с ростом последней.

Определить математическим путем наилучшее сочетание полноты-точности создаваемой модели с одной стороны и простоты с другой, практически никогда не удастся из-за неформализуемости и неоднозначности большей части подлежащих учету факторов.

Пара задача-объект в основном определяет номенклатуру подлежащих учету переменных объекта; параметры, входящие в модель, число и характер связей между ними, требования к точности данных и ряд других важнейших характеристик модели. Решающим фактором эффективности сейчас оказывается математический аппарат. Эффективность модели зависит и от такого субъективного момента, как профессиональные качества и уровень подготовки исследователя – исполнителя.

Таким образом, можно сделать заключение: наилучшее в практическом отношении качество или эффективность любой модели достигается как разумный компромисс между близостью модели к оригиналу (адекватностью) и простотой, обеспечивающей возможность и удобство использования модели по её прямому назначению; чрезмерная точность модели на практике не менее вредна, чем её неполнота и грубость.

Математическая модель изучаемого процесса или объекта является основой, фундаментом всего исследования. Тем не менее на сегодняшний день не существует и, по-видимому, не может существовать науки о моделировании реальных процессов и явлений окружающего мира — точно так же, как не существует науки о том, как совершать открытия, изобретения, создавать новые методы научного поиска. Даже математика — одна из наук, которая в большей, чем другие науки, степени использует дедукцию, своему прогрессу обязана в громадной степени таким “ненаучным” приемам, как интуиция, догадка, фантазия, т. е. индуктивному способу мышления. Моделирование объектов и явлений реальности (на сегодняшний день) в большой степени представляет искусство, а искусству учат на опыте. Человечество обладает таким опытом. Это опыт

классиков естествознания, опыт представителей естественных наук, эксплуатирующих для своих целей математический аппарат, и т. д.

В каждом конкретном случае качество модели во многом зависит от способностей исследователя понять существо, физику изучаемого процесса и создать его адекватное математическое описание. Математику привлекают, когда сложен изучаемый или управляемый процесс. Сложность обычно состоит в огромном числе характеристик, его описывающих, и большом числе связей между ними. И задача заключается не только в том, чтобы создать адекватное математическое описание изучаемого процесса, т. е. его модель, но и разработать методику работы с нею. С громоздкими многопараметрическими моделями трудно проводить исследования, поэтому математики вынуждены были при формализации реального процесса отбрасывать многие, на их взгляд менее существенные связи, загроублять математическое описание. Необходимо обладать незаурядной интуицией для определения, что важно с точки зрения интересующих исследователя вопросов, что — нет. Однако при решении серьезных практических задач невозможно полагаться лишь на интуицию и опыт небольшой группы исследователей, необходима методика, позволяющая с большой степенью достоверности определить адекватность модели и реальности, ею описываемой, область возможного ее применения и круг вопросов, для исследования которых они пригодны. Необходима “система знаний”, которая позволила бы, используя накопленный опыт и определенные принципы, выработанные на его основе, а также доказанные или установленные на их базе положения, создавать модели изучаемых процессов, проводить их анализ и определять пути их дальнейшего использования.

## **1.5. Методы построения моделей**

### **1.5.1. Общая логика процесса моделирования**

В самом общем случае процесс системного моделирования может быть представлен в форме взаимосвязанных этапов, на каждом из которых выполняются определенные действия, направленные на построение и последующее использование моделей систем (рис. 1.2). Характерной особенностью данного процесса является его циклический или итеративный харак-

тер, который отражает современные требования к анализу и проектированию сложных систем. Ниже дается краткая характеристика каждого из этапов, конкретное содержание которых зависит от специфических особенностей решаемых задач в той или иной предметной области.

*Анализ проблемной ситуации.* Одним из основных принципов системного моделирования является проблемная ориентация процессов построения и использования моделей. Другими словами, та или иная модель конкретной системы строится в контексте решения некоторой проблемы или достижения некоторой цели.



Рис. 1.2. Общая концептуальная схема процесса моделирования

Главное назначение первого этапа – логическое осмысление решаемой проблемы в контексте методологии системного моделирования. При этом

выполняется анализ всех доступных ресурсов (материальных, финансовых, информационных и др.), необходимых для построения модели, ее использования и реализации полученных результатов с целью решения имеющейся проблемы.

В случае отсутствия требуемых ресурсов на данном этапе может быть принято решение либо о сужении (уменьшении масштаба) решаемой проблемы, либо вообще об отказе от использования средств системного моделирования. На этом этапе также выполняется анализ требований, предъявляемых в той или иной форме к результату решения проблемы.

Первоначальный анализ решаемой проблемы и соответствующей проблемной области является наименее формализуемым с точки зрения применения известных аналитических подходов и средств. Поэтому на данном этапе рекомендуется применять так называемые эвристические или неформальные методы системного анализа.

*Структуризация предметной области и построение модели.* Целью данного этапа является построение адекватной модели системы и соответствующей предметной области в наиболее общем контексте решения исходной проблемы. Структуризация проблемной области предполагает определение и последующее уточнение ее границ, а также установление границ и состава систем, которые потенциально могут участвовать в решении исходной проблемы. Соответствующая информация представляется в форме модели системы или проблемной области в целом на некотором формально-логическом языке.

Речь идет о том, чтобы вся доступная информация о решении проблемы должна быть зафиксирована в виде некоторой модели системы. При этом модель должна удовлетворять принципу адекватности отражения основных особенностей системы-оригинала. Другими словами, модель не должна быть ни поверхностной (неполной), которая не учитывает существ-

венные аспекты структуры или поведения системы-оригинала, ни излишне сложной или избыточной, в рамках которой разработчики пытаются учесть даже несущественные с точки зрения исходной проблемы детали системы-оригинала.

Данный этап построения модели предполагает выполнение следующей последовательности действий:

1. Построение концептуальной или информационной модели системы и проблемной области, которая содержит наиболее общую информацию и отражает структурные взаимосвязи системы-оригинала с другими объектами окружающей среды.

2. Построение аналитической или математической модели системы, которая детализирует отдельные аспекты структуры и поведения системы-оригинала в форме текста с использованием специальной математической нотации (символики).

3. Построение имитационной или программной модели системы, которая непосредственно реализует информационно-логическую модель в форме, специально предназначенной для ее исследования с использованием компьютеров.

Один из принципов системного моделирования заключается в том, что для построения адекватной модели сложной системы может потребоваться не одна, а несколько моделей системы-оригинала. В этом случае каждая из подобных моделей будет являться отдельным представлением сложной системы, а полная модель системы будет состоять из комплекса взаимосвязанных моделей. Этот принцип получил специальное название – принцип многомодельности системного моделирования. С точки зрения системного аналитика все частные модели системы равноправны, поэтому корректно вести речь лишь об их адекватности. При этом выбор типа модели должен зависеть от характера решаемой проблемы, а не от профессиональной специализации прикладных математиков и системных аналитиков, участвующих в решении проблемы.

Процесс разработки адекватных моделей и их последующего конструктивного применения требует не только знания общей методологии сис-

темного анализа, но и наличия соответствующих изобразительных средств или языков для фиксации результатов моделирования и их документирования. Очевидно, что естественный язык не вполне подходит для этой цели, поскольку обладает неоднозначностью и неопределенностью. Поэтому для построения моделей используются формально-теоретические методы, основанные на дальнейшем развитии математических и логических средств моделирования. Для этой цели также предложены различные графические нотации и языки моделирования, в той или иной степени отражающие специфику решаемых задач на основе применения соответствующих программных инструментариев.

*Выполнение вычислительных экспериментов с моделью.* Модель системы разрабатывается для получения некоторой новой информации о системе-оригинале с целью решения исходной проблемы. В этом случае базовым объектом для получения такой информации является программная модель сложной системы, реализованная на одном из языков программирования или построенная с использованием соответствующих программных инструментариев.

Реализация данного этапа в контексте методологии системного моделирования означает выполнение серии экспериментов с программной моделью системы на той или иной вычислительной платформе. При этом возможна следующая последовательность действий, отражающая содержание собственно процесса планирования экспериментов:

1. Формирование конкретных значений наборов исходных данных (входных переменных), которые характеризуют отдельный вычислительный эксперимент с программной моделью системы.

2. Выполнение расчетов или, в общем случае, выполнение отдельной итерации с имитационной моделью системы с целью получения конкретных значений выходных параметров (переменных) модели.

3. Оценка точности и верификация полученных результатов на основе проверки согласованности отдельных компонентов вычислительных расчетов с использованием аналитической модели.

4. Интерпретация полученных результатов в форме управляющих воздействий или альтернатив решения исходной проблемы.

5. Оценка потенциальной возможности реализации полученных результатов применительно к системе-оригиналу.

*Применение результатов вычислительных экспериментов.* Содержанием данного этапа является материальное или информационное воздействие на систему-оригинал с целью решения исходной проблемы. При этом может потребоваться планирование организационных мероприятий по реализации подобных воздействий и контроль их выполнения.

После реализации рекомендаций выполненных исследований, что оказывается возможным только после окончания этапа вычислительных экспериментов с моделью, может сложиться одна из двух ситуаций.

Исходная проблема полностью решена – тем самым цели системного моделирования достигнуты. В этом случае можно перейти к решению очередной проблемы из данной предметной области, что характеризует начало нового цикла системного моделирования.

Исходная проблема не решена или решена не полностью – тем самым цели системного моделирования не достигнуты. В этом случае необходимо тщательно проанализировать сложившуюся ситуацию и причины неудачи. После этого можно перейти либо к коррекции исходной модели системы, либо вообще отказаться от построенной модели и реализовывать цикл системного моделирования заново.

Следует заметить, что процесс системного моделирования при решении сложных проблем занимает достаточно продолжительное время, в течение которого, вообще говоря, может измениться как само содержание исходной проблемы, так и наличие необходимых для ее решения ресурсов. Эти особенности зачастую не учитываются при реа-

лизации сложных проектов, что является источником их неудачного завершения. Именно для исключения или ослабления негативного влияния данных факторов на схеме системного моделирования должен быть предусмотрен отдельный этап – коррекция или доработка модели, который может начать выполняться с любого момента изменения исходной ситуации или в результате возникновения признаков неадекватности модели на любом из рассмотренных выше этапов.

*Коррекция или доработка модели.* Цель данного этапа неявно была уже сформулирована выше, а именно – внесение изменений в существующую модель, которые направлены на обеспечение ее адекватности решаемой проблеме. Речь может идти как о включении в состав исходной модели дополнительных компонентов, так и о радикальном изменении структуры и содержания модели. Важно отметить проблемно-ориентированный характер этих изменений, т.е. коррекция или доработка модели должны выполняться в непосредственном контакте с решаемой проблемой.

Рассмотрим особенности двух основных классов моделей:

- аналитических;
- идентифицируемых.

### **1.5.2. Аналитические модели**

Одним из главных результатов многовекового развития науки является познание и объяснение бесчисленного множества объективно существующих явлений и процессов, протекающих на разных уровнях живой и неживой природы. Компоненты теоретического арсенала современной науки – картины мира, теории, законы, принципы – все они от наиболее общих практически универсальных, таких как законы сохранения вещества и энергии, начала термодинамики, закон всемирного тяготения и других, до сугубо локальных, относящихся к узкому классу объектов или явлений, носят модельный характер. Таким образом в распоряжении исследователя,

решающего на основе моделирования конкретную исследовательскую или прикладную задачу, сегодня находится огромное множество моделей-заготовок, которые, очевидно, могут и должны быть использованы. Наиболее благоприятной является ситуация, когда подлежащие описанию и исследованию свойства объекта удается представить непосредственно на основе ранее разработанных и практически достоверных модельных конструкций, являющихся элементами соответствующих областей теоретического знания (механики, термодинамики, электротехники и т.п.). В этом случае создаваемая конкретная модель должна быть охарактеризована как аналитическая (теоретическая). Она, как правило, не только описывает свойства и характеристики объекта, но вскрывает и в терминах соответствующих теорий выявляет сущность процессов, протекающих в исследуемом объекте. Все допущения и ограничения переносятся на модель.

На практике теоретические модели выступают в двух основных ролях. Прежде всего, они образуют структурную основу и являются главным исходным материалом всех без исключения теоретических построений. Любая теория, относящаяся к сфере точных наук, есть не что иное, как система взаимосвязанных аналитических моделей, подчиненная регулятивным принципам и универсальным зависимостям более высокого уровня. В поисковых областях научного знания теоретические модели, предназначенные для объяснения и описания явлений, не укладывающихся в существующие теоретические представления, играют роль главного инструмента познания.

В сложившихся областях научного знания, главным образом, прикладного характера, таких как теоретическая механика, теоретическая электротехника и др. аналитические модели, в большей или меньшей мере дополняемые обобщенными экспериментальными данными, носят типовой канонический характер, они являются важнейшей составной частью

понятийного аппарата, специфического языка и профессионального мышления. Вместе с тем, модели этого класса являются основой для решения множества конкретных прикладных задач, в частности инженерно-технического характера, относящимся к хорошо изученным, не слишком сложным объектам и носящим типовой или рутинный характер. Расчет прочностных характеристик конструкций, расчеты параметров и характеристик электрических цепей. В каждом конкретном случае модель исследуемого явления строится с учетом специфики природы и свойств объекта. Вместе с тем можно указать и некоторые общие методы и приемы.

В основе аналитических моделей, как правило, лежат так называемые балансовые соотношения, связывающие входные и выходные переменные или некоторые функционалы от этих переменных, имеющие смысл обобщенных сил, обобщенных потоков или координат. Типичные примеры: условие равновесия сил или моментов, действующих на некоторую механическую систему, равенство масс исходных и конечных продуктов некоторой химической реакции, равенство нулю суммы ЭДС и падений напряжений в электрической цепи и т.п. Все эти и прочие им подобные соотношения по существу представляют собой частные проявления законов сохранения вещества и энергии. К этой основе добавляется необходимая дополнительная информация, не вытекающая из этих соотношений, источником которой может быть либо специфическая для данного класса объектов теория, либо эксперимент.

Возможности чисто теоретического решения вопроса уменьшаются с ростом сложности и новизны исследуемого объекта. Впрочем, опыт показывает, что нередко даже для широко используемых на практике и казалось бы, хорошо изученных объектов и процессов, например металлургических, чисто аналитическим путем построить удовлетворительную модель не удается и это побуждает исследователя к формированию модели

преимущественно на экспериментальной основе, т.е. в классе идентифицируемых моделей.

### **1.5.3. Идентифицируемые модели**

В основе всех ныне весьма многочисленных методов идентификации или опытного отождествления модели с объектом-оригиналом, лежит идея мысленного эксперимента с «черным ящиком» (Н. Винер). В предельном (теоретическом) случае «черный ящик» представляет собой некую систему, о структуре и внутренних свойствах которой неизвестно решительно ничего. Зато входы, т.е. внешние факторы, воздействующие на этот объект, и выходы, представляющие собой реакции на входные воздействия, доступны для наблюдений (измерений) в течение неограниченного времени. Задача заключается в том, чтобы по наблюдаемым данным о входах и выходах выявить внутренние свойства объекта или, иными словами, построить модель.

Задача идентификации формулируется следующим образом: по результатам наблюдений над входными и выходными переменными системы должна быть построена оптимальная в некотором смысле модель, т.е. формализованное представление этой системы.

Решение задачи допускает применение двух стратегий:

Осуществляется активный эксперимент. На вход подаются специальные сформированные тестовые сигналы, характер и последовательность которых определена заранее разработанным планом. Преимущество: за счет оптимального планирования эксперимента необходимая информация о свойствах и характеристиках объекта получается при минимальном объеме первичных экспериментальных данных и соответственно при минимальной трудоемкости опытных работ. Но цена за это достаточно высока:

объект выводится из его естественного состояния (или режима функционирования), что не всегда возможно.

Осуществляется пассивный эксперимент. Объект функционирует в своем естественном режиме, но при этом организуются систематические измерения и регистрация значений его входных и выходных переменных. Информацию получают ту же, но необходимый объем данных существенно, на 2-3 порядка больше, чем в первом случае.

На практике при построении идентифицируемых моделей часто целесообразна смешанная стратегия эксперимента. По тем входным переменным конкретного объекта, которые это допускают (по условиям безопасности, техническим, экономическим соображениям и пр.), проводится активный эксперимент. Его результаты дополняют данными пассивного эксперимента, охватывающего все прочие значимые переменные. «Черный ящик» - теоретически граничный случай. На деле есть объем исходной информации. На практике приходится иметь дело с «серым», отчасти прозрачным ящиком. Поэтому различают три основных класса постановки задачи идентификации объекта:

1. Для сложных и слабо изученных объектов системного характера достоверные исходные данные о внутренних свойствах и структурных особенностях исчезающе малы, почти отсутствуют. Поэтому задача идентификации, казалось бы, должна включать в себя с одной стороны, определение зависимостей, связывающих входы и выходы (обобщенного оператора), с другой - определение внутренней структуры объекта. В такой постановке задача не разрешима даже теоретически. Непосредственным результатом идентификации является только определение зависимостей входы-выходы, причем не в параметрической форме - в виде таблиц или кривых. Для того, чтобы говорить о структуре модели, необходимо перейти к параметрической форме их представлений. Однако, как известно, од-

нозначной связи между функциональной зависимостью и порождающей эту зависимость математической структурой не существует. Каждую непараметрическую зависимость вход-выход можно аппроксимировать различными способами и соответственно построить ряд практически равноценных моделей, характеризующихся собственной структурой, собственным набором параметров и их значений. Основанием для предпочтения той или иной параметрической модели могут быть только данные, внешние по отношению к процессу идентификации, например, основанные на теоретических соображениях. Если таких данных нет, то в рассматриваемой ситуации мы получаем чисто функциональную или имитационную модель, которая воспроизводит с тем или иным приближением характеристики объекта.

2. Второй класс задач идентификации характеризуется тем, что априорные данные о структуре моделируемого объекта, в принципе имеются. Однако, какой вклад в характеристики объекта или его модели вносит тот или иной компонент, заранее не известно и это надлежит определить на основе эксперимента наряду со значением соответствующих параметров. Типичный пример – исследование влияния на характеристики динамической системы, описанной в классе стационарных линейных моделей, старших членов соответствующих дифференциальных уравнений ради того, чтобы исключить малые, практически незначимые переменные, и снизив порядок уравнений, упростить модель. Задачи этого класса, связанные с уточнением структуры и оценивания параметров, часто встречаются на практике и характерны для объектов и процессов средней сложности, в частности технологических.

3. Третий класс задач связан с относительно простыми и хорошо изученными объектами, структура которых известна точно и речь идет только о том, чтобы по экспериментальным данным оценить значения всех или

некоторых входящих в исследуемую структуру параметров (параметрическая идентификация). Очевидно, что модели данного класса тесно смыкаются с требующими экспериментального доопределения аналитическими моделями и четкой границы между ними не существует. Это наиболее массовый класс задач.

Независимо от характера решаемой на основе идентификации задачи, построение модели базируется на результатах измерений соответствующих величин переменных. С этим связаны два существенных обстоятельства:

- эксперимент должен быть обеспечен необходимыми средствами измерения надлежащей точности (датчики, измерительные преобразователи, средства регистрации);

- измерительный комплекс со всеми его компонентами требует метрологического обеспечения, т.е. градуировки, аттестации и периодичности проверки.

Реальные свойства подавляющего большинства сложных объектов, а также неизбежные случайные погрешности измерений, лежащих в основе идентификации, придают последней статистический характер, что влечет за собой необходимость получения больших объемов первичных экспериментальных данных с их последующей обработкой. Поэтому на практике построение моделей путем идентификации неизбежно связано с использованием информационно-вычислительной техники как при получении первичных данных (автоматизация эксперимента), так и для их обработки и использования.

## **1.6. Вопросы для самопроверки**

1. Назовите компоненты, которые входят в общее определение понятия «модель».

2. Какими свойствами обладает модель?
3. Можно ли построить модель, отражающую свойства и характеристики некоторой другой модели?
4. По каким признакам можно классифицировать модели?
5. В чем заключаются преимущества математических моделей перед другими видами моделей?
6. Какие наиболее существенные свойства объектов моделирования влияют на классификацию математических моделей?
7. Чем отличаются модели статических объектов от моделей стационарных объектов?
8. В какой взаимосвязи находятся свойства адекватности и эффективности математических моделей?
9. Назовите общие этапы процесса математического моделирования.

# ГЛАВА 2. ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

## 2.1. Постановка задачи идентификации

Задача идентификации формулируется следующим образом. Пусть в результате каких-либо экспериментов над объектом замерены его входные  $X = (x_1, x_2, \dots, x_n)$  и выходные переменные  $Y = (y_1, y_2, \dots, y_m)$  как функции времени. Требуется определить вид (структуру) и параметры некоторого оператора  $\hat{A}$ , ставящего в соответствие переменные  $X$  и  $Y$ .

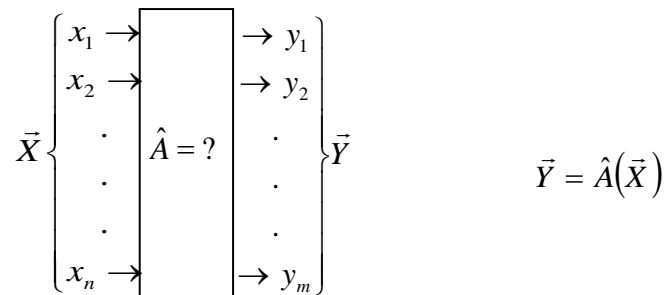


Рис. 2.1. Общая схема постановки задачи идентификации

В реальных условиях переменные  $\vec{X}$  и  $\vec{Y}$  замеряются с погрешностью  $\xi_x$  и  $\xi_y$ . Чаще всего эти погрешности считаются некоррелированными и аддитивными с полезной информацией, т.е. имеют вид

$$x_i = x_i^{ucm} \pm \xi_x \quad (i=1, 2 \dots n)$$

$$y_j = y_j^{ucm} \pm \xi_y \quad (j=1, 2 \dots m).$$

В зависимости от априорной информации об объекте различают задачи идентификации в узком и широком смысле. Задача идентификации в узком смысле (параметрическая идентификация) состоит в оценивании параметров и состояния системы по результатам наблюдений над входными и выходными переменными, полученными в условиях функционирования объ-

екта. При этом известна структура системы и задан класс моделей, к которому данный объект относится. Априорная информация об объекте точна велика.

Априорная информация об объекте при идентификации в широком смысле (структурная идентификация) отсутствует или очень мала, поэтому приходится предварительно решать большое число дополнительных задач. К этим задачам относятся: выбор структуры системы и задание класса моделей, оценивание степени стационарности и линейности объекта и действующих переменных, оценивание степени и формы влияния входных переменных на выходные, выбор информативных переменных и другие.

Задача параметрической идентификации сводится к отысканию таких оценок параметров математической модели  $\hat{A}$ , которые обеспечивают в каком либо смысле близость расчетных  $\vec{Y}_p$  и экспериментальных  $\vec{Y}_e$  значений выходных переменных при одинаковых входных  $\vec{X}$ . Отметим, что в общем случае необходимы измерения «m» компонент вектора  $\vec{Y}$ , которые могут производиться при «k» повторениях эксперимента при «l» дискретных отметках времени (если идентифицируемый объект функционирует во времени). В качестве критериев количественной меры близости модели и оригинала чаще всего используются максимальные  $\delta_y$ , средние  $m_y$  и среднеквадратичные  $\sigma_y$  величины погрешностей рассогласования расчетных и экспериментальных значений  $y_{pj}$  и  $y_{ej}$ , соответственно, т.е

$$\begin{aligned} \delta_y &= \max |y_{pj} - y_{ej}|, \\ m_y &= 1/N \sum_{j=1}^N (y_{pj} - y_{ej}), \\ \sigma_y &= \sqrt{\frac{1}{N} \sum_{j=1}^N (y_{pj} - y_{ej})^2}, \end{aligned} \tag{2.1}$$

где:  $j = 1, 2, \dots, N = m + 1 + k$  - номер опыта по измерению компоненты  $U_{эj}$ .

Таким образом, задача параметрической идентификации сводится к минимизации одной из функций вида (2.1). Для минимизации могут быть использованы известные численные методы решения экстремальных задач.

Обилие существующих методов идентификации отражает разнообразие используемых математических моделей и методов их исследования. Очевидно, что идентифицировать модель детерминированного, линейного, стационарного процесса (модель считается стационарной, если её параметры либо постоянные, либо меняются медленно по сравнению со временем, необходимым для их идентификации) известной размерности, с одним входом – существенно проще, чем аналогичного стохастического процесса неизвестного порядка и степени стационарности.

## **2.2. Идентификация моделей с помощью регрессионного метода**

Регрессионный анализ представляет собой классический статистический метод. Благодаря своим широким возможностям регрессионные методы давно и успешно используются в инженерной практике. В последнее время в связи с развитием и внедрением быстродействующих ЭВМ они широко используются для идентификации моделей, в том числе для идентификации динамических, многомерных процессов, систем диагностики и управления в реальном масштабе времени. Регрессионный анализ основывается на двух главных принципах.

1. *Методы применяются для линейных по идентифицируемым параметрам моделям.* Структура математической модели процесса представляется функцией вида:

$$y = \sum_{i=1}^k a_i f_i(\bar{x}), \quad (2.2)$$

где  $a_i$  –  $i$ -тый оцениваемый параметр;  $f_i(\bar{x})$  –  $i$ -тая известная функция,  $\bar{x}$  – вектор входных воздействий,  $y$  – выходная переменная (можно рассматривать только одну выходную переменную в предположении, что все  $y_j$  не зависят друг от друга).

Возможно представление идентифицируемой модели в следующей форме:

$$y = \frac{\sum_{i=1}^k a_i f_i(\bar{x})}{1 + \sum_{j=1}^p b_j \psi_j(\bar{x})}, \quad (2.3)$$

где  $a_i, b_j$  – оцениваемые параметры;  $f_i(\bar{x})$  и  $\psi_j(\bar{x})$  – априори известные (заданные) функции. После несложных математических преобразований на основе этих функций можно формировать невязки, линейно зависящие от идентифицируемых параметров  $a_i, b_j$ .

На практике чаще всего в качестве  $f_i(\bar{x})$  и  $\psi_j(\bar{x})$  выбираются степенные функции, а соответственно выражения (2.2) и (2.3) являются полиномиальными, либо дробно-рациональными зависимостями. При этом точность описания достигается увеличением числа членов полинома, обеспечивающих их сходимость к реальному процессу. Заметим, что получающаяся модель практически никогда не соответствует физической сущности моделируемого реального процесса, его истинному виду, однако инженерная простота вычислений, удобство практического использования модели, возможность получения результата без «особых размышлений» служит основной причиной широкого распространения на практике регрессионных методов.

Естественно, и в этом случае с помощью удачно выбранного вида полинома можно существенно сократить размер модели, а значит и трудоемкость вычислительного процесса, как при идентификации, так и при использовании модели.

2. *Минимизируемой функцией ошибки (разности между прогнозируемой моделью и данными эксперимента) при регрессионном анализе является сумма квадратов ошибок выходного фактора.* Благодаря этому удастся применить метод наименьших квадратов, математический аппарат которого предельно прост, а вычислительные методы сводятся к методам линейной алгебры.

Регрессионные модели могут быть как линейными, так и нелинейными с любым числом входов и выходов.

### **2.2.1. Идентификация статических линейных систем с несколькими входами**

Пусть необходимо идентифицировать систему с «n» входами  $x_1, x_2, \dots, x_n$  и одним выходом  $y$ . Представим структуру модели в виде линейного алгебраического уравнения вида:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad , \quad (2.4)$$

где  $a_0, a_1, \dots, a_n$  - параметры модели, подлежащие идентификации. В результате идентификации мы должны получить вектор оценок  $\hat{a}$  истинного вектора  $\bar{a}$ . Этому вектору будет соответствовать оценка значения выходной величины  $\hat{y}$ .

Для определения значений  $\hat{a}_i (i = 0, 1, 2, \dots, n)$  произведем N последовательных измерений величины  $y$ , соответствующих в определенном смысле произвольным набором величин  $x_i (i=1, 2, \dots, n)$ . В результате получим век-

тор  $\bar{y} = (y_1, y_2, \dots, y_N)$ . По N наборам входных величин  $x_i$  ( $i=1, 2, \dots, n$ ) будет соответственно N оценок выходных величин

$$\begin{aligned}\hat{y}_1 &= \hat{a}_0 + \hat{a}_1 x_1^{(1)} + \dots + \hat{a}_n x_n^{(1)} \\ \hat{y}_2 &= \hat{a}_0 + \hat{a}_1 x_1^{(2)} + \dots + \hat{a}_n x_n^{(2)} \\ \hat{y}_N &= \hat{a}_0 + \hat{a}_1 x_1^{(N)} + \dots + \hat{a}_n x_n^{(N)}.\end{aligned}\tag{2.5}$$

Разница  $y_j - \hat{y}_j$  характеризует погрешность каждой модели в каждом из N измерений. Суммарную погрешность будем характеризовать величиной:

$$J = \sum_{j=1}^N (y_j - \hat{y}_j)^2.\tag{2.6}$$

Определение оценок  $\hat{a}$  производят из условия минимума величины суммарной погрешности J. Таким образом, основой идентификации регрессионными методами служит метод наименьших квадратов. Используя аппарат математического анализа, оценка вектора  $\hat{a}$  должна удовлетворять необходимому условию экстремума (подразумевается, что условие положительной определенности гессиана - матрицы из вторых производных - выполнено)

$$\frac{\partial J}{\partial \hat{a}_i} = 0 \quad (i=0, 1, 2, \dots, n)\tag{2.7}$$

Уравнения (2.7) позволяют построить вычислительный процесс идентификации вектора  $\bar{a}$  на основе N групп измерений  $y$  и  $\bar{x}$ . Для получения эффективных и несмещенных оценок  $\hat{a}^*$  необходимо, чтобы  $N \gg n$ . Если  $N = n+1$ , то в оценке  $\hat{y}$  шум измерений не будет сглажен, окажет негативное влияние и случайность наборов  $\bar{x}$ . Мерой ошибки регрессионной модели обычно служит величина среднеквадратичного отклонения

$\sigma_y = \sqrt{\frac{1}{N-n-1} J}$ . С увеличением N уменьшается флуктуация  $\sigma_y$ , ее величи-

на и является определяющей для выбора N в рамках принятой структуры модели.

*Псевдолинейная (внутренне линейная) модель.*

К этому классу относятся нелинейные по идентифицируемым параметрам модели, которые простыми преобразованиями и подстановками сводятся к линейным. В качестве примера псевдолинейной модели приведем мультипликативную модель:

$$y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_n^{a_n} . \quad (2.8)$$

Логарифмируя это выражение

$$\ln y = \ln a_0 + a_1 \ln x_1 + \dots + a_n \ln x_n$$

и производя замены переменных:

$$W = \ln y, z_i = \ln x_i \quad (i = 1, 2, \dots, n), \ln a_0 = b_0$$

получаем линейную форму

$$W = b_0 + a_1 z_1 + \dots + a_n z_n .$$

После нахождения коэффициентов по процедурам линейной модели необходимо сделать обратное преобразование  $a_0 = \exp(b_0)$ . Зависимости (2.8) широко используются при построении моделей по эмпирическим данным в гидравлике, термодинамике, обработке металлов давлением и т.д.

Другими примерами псевдолинейных моделей являются следующие:

$$y = \exp(a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n) ,$$

$$y = \frac{1}{a_0 + a_1 x_1 + \dots + a_n x_n} ,$$

$$y = \frac{1}{1 + \exp(a_0 + a_1 x_1 + \dots + a_n x_n)} , \quad (2.9)$$

$$y = a_0 + \frac{a_1}{x_1} + \dots + \frac{a_n}{x_n} .$$

## 2.2.2. Идентификация нелинейных систем

Если величина среднеквадратичного отклонения идентифицируемой системы  $\sigma_y$  слишком велика, значит, структура модели неудовлетворительна и необходимо учитывать нелинейные члены алгебраического полинома. Рассмотрим порядок идентификации параметров нелинейной системы второго порядка с одним выходом  $y$  и входами  $Z_1, Z_2, \dots, Z_n$

$$y = L_0 + \sum_{i=1}^n L_i Z_i + \sum_{j=1}^n L_{1j} Z_1 Z_j + \sum_{j=2}^n L_{2j} Z_2 Z_j + \dots + L_m Z_n^2. \quad (2.10)$$

Обозначим переменные  $Z_i$  и произведения  $Z_i Z_j$  переменных в выражении (2.10) через новые переменные  $x_i$  ( $i=1, 2, \dots, m$ ). С учетом этого, выражение (2.10) можно записать в виде:

$$y = a_0 + \sum_{i=1}^m a_i x_i, \quad (2.11)$$

что совпадает с (2.4).

Для большинства систем 20% факторов определяют 80% свойств. Поэтому в процессе идентификации несущественные параметры, влияние которых на величину среднеквадратичного отклонения мало, отсеивают. Тем самым определение значений постоянных коэффициентов в (2.10) и в эквивалентном ему выражении (2.11) представляет собой процесс структурно-параметрической идентификации.

Отсеивание несущественных параметров можно проводить с помощью, например корреляционного анализа, т.е. вычисления так называемых коэффициентов парной корреляции между входными и выходной переменными  $r_{x_i y}$  (здесь  $x_i$  - компонент входного вектора  $\vec{X}$ ,  $i=1, \dots, n$ ). Коэффициент парной корреляции характеризует меру линейной зависимости между двумя переменными  $x_i$  и  $y$  и меняется от -1 до +1. Чем ближе модуль  $r_{x_i y}$  к

1, тем ближе изучаемая зависимость к линейной. Значение  $r_{x,y} \approx 0$  свидетельствует об отсутствии линейной связи между  $x_i$  и  $y$ , т. е. при построении линейной регрессионной модели переменную  $x_i$  учитывать не следует. Коэффициент парной корреляции вычисляют по формуле:

$$r_{x,y} = \frac{\sum_{k=1}^N (x_{ik} y_k) - \frac{1}{N} \sum_{k=1}^N x_{ik} \sum_{k=1}^N y_k}{\sqrt{\left[ \sum_{k=1}^N x_{ik}^2 - \frac{1}{N} \left( \sum_{k=1}^N x_{ik} \right)^2 \right] \cdot \left[ \sum_{k=1}^N y_k^2 - \frac{1}{N} \left( \sum_{k=1}^N y_k \right)^2 \right]}}, \text{ где} \quad (2.12)$$

$N$  - количество опытов (повторений эксперимента),  
 $x_{ik}$  - значение (экспериментальное)  $i$ -ой входной переменной для  $k$ -го опыта,  
та,

$y_k$  - значение (экспериментальное) выходной переменной для  $k$ -го опыта.

Оценку значимости коэффициента парной корреляции (проверку наличия корреляции) проводят разными способами. В самом простом случае значение коэффициента корреляции сравнивается с некоторым его критическим значением, если  $r_{x,y} \geq r_{крит}$ , то переменную  $x_i$  необходимо учитывать в зависимости для  $y$ .

### 2.2.3. Достоверность (адекватность) регрессионной модели

Обычно мерой ошибки регрессионной модели служит стандартное (среднеквадратичное) отклонение  $\sigma_y$ . Для процессов, подчиняющихся закону нормального распределения, приблизительно 66% точек находится в пределах одного стандартного отклонения от модели и 95% точек в пределах двух стандартных отклонений.

Стандартное отклонение – важный показатель для решения вопроса о достоверности модели. Большая ошибка может означать, что модель не соответствует процессу, который послужил источником экспериментальных данных. Однако большая ошибка модели может быть вызвана и другой причиной: большим разбросом данных измерений. В этом случае, возможно, потребуется взять большее количество выборок.

Проверку значимости (качества предсказания) множественного уравнения регрессии можно осуществить на основе F-критерия Фишера. Вычисляют дисперсию среднего:

$$\bar{S}_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2$$

Вычисляют так называемую остаточную дисперсию  $\bar{S}_{ao}^2$  (дисперсию адекватности):

$$\bar{S}_{ao}^2 = \frac{1}{N-n-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Сравнивают  $F = \frac{\bar{S}_y^2}{\bar{S}_{ao}^2}$  с числом степеней свободы в числителе  $\nu_1 = N-1$ ,

в знаменателе  $\nu_2 = N-n-1$ . Считают, что уравнение регрессии предсказывает результаты опытов лучше среднего, если F достигает или превышает границу значимости при выбранном уровне значимости  $p$  (обычно принимают  $p = 1 - q = 5\%$ ,  $q$  – вероятность предсказания). Другими словами, F – критерий Фишера показывает во сколько раз уравнение регрессии предсказывает результаты опытов лучше, чем среднее «у».

### 2.3. Построение моделей идентификации поисковыми методами<sup>1</sup>

При параметрической идентификации моделей с сугубо нелинейными параметрами дело обстоит сложнее. Как и в случае линейного и псевдолинейного моделирования, идентификацию таких моделей можно выполнить методом наименьших квадратов, приравнивая к нулю частные производные (см. 2.7). В результате получается система нелинейных уравнений, общего аналитического метода решения которой не существует. Применение же численных методов на практике часто оказывается неэффективным вследствие общих недостатков, присущих этим методам:

1. успешность применения того или иного метода зависит от качества начального решения - приближения, используемого на первом шаге итерационного процесса: многие методы эффективны только в том случае, если начальное решение выбрано достаточно близко от искомого результата;
2. в отличие от одномерного случая (решение одного независимого нелинейного уравнения) нельзя быть уверенным, что корень системы находится в той ограниченной области, где происходит итерационный процесс;
3. нельзя быть также полностью уверенным, что система имеет строго одно решение: может оказаться, что она имеет несколько решений или же вообще их не имеет.

Помимо этого, дополнительные трудности возникают вследствие самой структуры задачи параметрической идентификации. Так, для решения системы (2.7) необходимо, чтобы функция  $J$ , параметры  $\bar{a}$  которой мы

---

<sup>1</sup> Параграф написан совместно с доцентом кафедры САПР и ПК ВолгГТУ, к.т.н. Крыжановским Д.И.

ищем, была дважды дифференцируема в области поиска решения. Однако в общем случае мы не можем быть уверены, что это условие будет выполняться. Кроме того, даже если вторая производная функции  $J$  существует на всём интересующем нас множестве, вычисление её и первой производной может оказаться трудоёмким. Надеяться на изобретение более эффективных численных методов также не приходится – все издержки численных методов решения систем нелинейных уравнений обусловлены самой природой задачи (многомерность и нелинейность), а не техническими недостатками существующих методов.

Задачу нахождения значений параметров сугубо нелинейных моделей можно решить, непосредственно оптимизируя сумму квадратов невязок (2.6). Преимуществом такого подхода по сравнению с традиционным методом наименьших квадратов являются менее строгие ограничения, накладываемые на вид функции моделирования  $y$ . Кроме того, использование традиционного метода требует предварительного составления системы нормальных уравнений (2.7), вывода выражений для первой и второй производных и т.п. Подход, основанный на применении методов оптимизации критерия (2.6), в этом отношении гораздо более универсален. Одна и та же алгоритмическая процедура без предварительных вычислений, выполняемых человеком, может быть использована для очень широкого класса математических зависимостей.

Нами составлены методические указания для решения задачи параметрической идентификации с использованием методов Нелдера-Мида и поиска глобального экстремума функции многих переменных на сетке кода Грея (RRR). Эти методы хорошо себя зарекомендовали при решении различных задач, где традиционный метод наименьших квадратов не давал приемлемых результатов, но они обладают и серьёзными недостатками. Так, алгоритм Нелдера-Мида является методом локальной оптимизации.

Это означает, что найденное решение может располагаться достаточно далеко от глобального оптимума минимизируемой функции. Кроме того, он эффективно работает только при размерности пространства поиска  $k < 6$  (в нашем случае для функций, у которых число параметров меньше 6). Наконец, метод RRR, представляющий собой метод глобальной оптимизации на дискретной сетке, сильно ограничен в точности при большом числе входных переменных.

В последнее время всё большее внимание привлекают к себе стохастические методы оптимизации, которые, несмотря на свою вероятностную природу, применяются с большим успехом для всё более широкого круга задач. Так, можно предложить комбинированный алгоритм нелинейной параметрической идентификации, который сначала осуществляет глобальный поиск оптимума, а затем, после сужения области поиска (локализации решения), запускается процедура локального поиска (например, упомянутый выше алгоритм Нелдера-Мида) для уточнения решения.

В качестве базового алгоритма для глобального поиска можно использовать *алгоритм имитации отжига*. Несмотря на то, что этот алгоритм успешно используется в различных областях науки и техники, многократно был описан в литературе, в отечественной системе образования он практически не изучается. Поэтому ниже мы кратко изложим его суть.

Как правило, алгоритмы оптимизации носят итерационный характер. На начальном шаге у нас есть некоторое начальное приближение решения  $x_0$ , мы оцениваем его качество (штраф, отличие значения целевой функции от оптимального), после чего выбираем направление дальнейшего движения к оптимуму и шаг. Так мы получаем следующее приближение решения  $x_1$  и снова повторяем описанную выше процедуру. Недостатком такого подхода к оптимизации является то, что, находя на каждом новом шаге лучшее решение, мы легко можем оказаться в области локального оптимума.

ма, из которой итерационный процесс уже не выйдет, так как все точки в его окрестности будут ухудшать значение целевой функции. Основная идея стохастических методов глобальной оптимизации состоит в том, чтобы иногда в качестве следующего решения выбиралась точка с худшим значением целевой функции. Одним из таких алгоритмов и является алгоритм имитации отжига.

Алгоритм основывается на имитации физического процесса, который происходит при кристаллизации расплавов (в металлургии этот процесс называется отжигом металлов). Процесс протекает при постепенно понижающейся температуре. Переход атома металла из одной ячейки кристаллической решётки в другую происходит с некоторой вероятностью, причём вероятность уменьшается с понижением температуры. Устойчивая кристаллическая решётка соответствует минимуму энергии атомов, поэтому атом либо переходит в состояние с меньшим уровнем энергии, либо остаётся на месте. Теперь поясним, как эта аналогия с физикой работает для поиска оптимальных решений.

Мы задаёмся некоторым начальным решением  $x_0$  и начальным значением «температуры»  $T_0$ . Температура в данном случае – это некий параметр итерационной процедуры, который, как и температура в физическом процессе отжига, постоянно понижается. Случайным образом в окрестности начального решения  $x_0$  радиусом  $R_0$  мы генерируем точку  $x^*$ . Затем вычисляем значение целевой функции  $U^*$  в этой точке и сравниваем его со значением целевой функции  $U_0$  в точке  $x_0$ . Если  $U^* < U_0$  (для задачи минимизации), то  $x^*$  берётся в качестве следующего приближения решения. В противном случае разыгрывается случайная величина  $\omega$  на полуинтервале  $[0; 1)$ . Если  $\omega > P$  ( $P$  – вероятность перехода), то мы также совершаем переход к точке  $x^*$ , хотя значение целевой функции в ней хуже, чем в  $x_0$ . По-

сле перехода новая текущая точка  $x^*$  обозначается за  $x_0$ , генерируется следующая точка  $x^*$  и процесс повторяется.

Вероятность перехода  $P$  зависит от текущего значения температуры  $T$  – чем ниже температура, тем ниже вероятность:

$$P = \exp\left(-\frac{U(x^*) - U(x_0)}{T}\right).$$

Таким образом, по мере «остывания» системы (по мере осуществления новых итераций) переходы в состояния с худшим решением становятся всё менее вероятными. Существует много различных модификаций алгоритма имитации отжига, которые различаются функциями зависимости вероятности  $P(T)$ , правилами изменения температуры и радиуса генерации новых точек, а также введением некоторых эвристических приёмов, улучшающих работу алгоритма в целом.

Итак, в поисковых методах идентификации принятый критерий невязки (показатель качества идентификации) формируется из выходных характеристик объекта и его идентифицируемой модели и минимизируется с помощью численных методов. Итерационный процесс изменения вектора идентифицируемых параметров определяется используемым алгоритмом (методом) поиска и текущей ситуацией. Поисковые методы необходимо применять при «безысходности» (когда сложность ситуации намного превосходит возможности аналитических методов) или когда другие методы идентификации оказываются неэффективными. Это наблюдается в объектах, имеющих существенно нелинейную природу, особенно, когда нелинейности не являются непрерывными и аппроксимация полиномами или дробно-рациональными функциями невозможна. Поисковые методы также применяются к объектам, описываемым сложными нелинейными функциями априорно известной формы (например, полученными физиками - теоретиками или инженерами - исследователями из анализа физических

представлений и требующими уточнения численных значений коэффициентов).

#### **2.4. Вопросы для самопроверки**

1. Каким образом формулируется задача построения моделей идентификации?
2. В чем заключается отличие структурной идентификации от параметрической идентификации?
3. На каких принципах основывается регрессионный метод построения моделей идентификации?
4. Назовите основные процедуры идентификации статических линейных объектов.
5. Что понимают под «внутренне линейными» моделями? Приведите примеры таких моделей.
6. Перечислите процедуры проверки модели на адекватность по критерию Фишера. Какие ограничения существуют для этого способа проверки?
7. Какова основная особенность поисковых методов идентификации?
8. В каких случаях целесообразно применение поисковых методов идентификации?

## ГЛАВА 3. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ НЕОДНОРОДНЫХ СИСТЕМ

При построении математических моделей сложных неоднородных систем эффективным оказывается их последовательное расчленение на подсистемы (декомпозиция системы) с сохранением связей между выявленными подсистемами. Процедура декомпозиции осуществляется до получения таких подсистем, которые в условиях рассматриваемой задачи будут признаны достаточно простыми и удобными для непосредственного математического описания. Эти подсистемы, не подлежащие дальнейшей декомпозиции, называются элементами сложной системы.

Таким образом, в общем случае сложная система является многоуровневой иерархической конструкцией из взаимодействующих элементов, объединяемых в подсистемы различных уровней. Представление моделируемого объекта в виде многоуровневой системы называется его структуризацией. Математическая модель сложной системы образуется композицией (в рамках выделенной структуры) математических моделей элементов и взаимодействий между ними.

Построение простой и изящной математической модели, достаточно точно описывающей процесс функционирования сложной системы, требует немалого искусства. Необходимо знать типичные математические схемы.

### 3.1. Математические модели элементов

Математические модели широкого класса детерминированных объектов (при описании которых влияние случайных факторов не учитывается), функционирующих в непрерывном времени, описываются чаще всего

дифференциальными уравнениями в обыкновенных или частных производных.

Детерминированные объекты, функционирующие в дискретном времени, описываются математическими моделями, сводящимися к различным типам конечных автоматов.

Автомат можно представить как некоторое устройство (черный ящик), на которое подаются дискретные входные воздействия (сигналы) и с которого снимаются дискретные выходные воздействия; оно имеет также некоторые внутренние состояния. Автомат функционирует в дискретном автоматном времени, моментами которого являются такты, т.е. примыкающие друг к другу равные интервалы времени, каждому из которых соответствуют постоянные значения входного и выходного сигналов и внутренние состояния. Обозначим состояние, а также входной и выходной сигналы, соответствующие  $t$ -тому такту при  $t = 0, 1, 2, \dots$  через  $z(t)$ ,  $x(t)$ ,  $y(t)$ . При этом, по условию  $z(0) = z_0$ , а  $z(t) \in Z$ ,  $x(t) \in X$ ,  $y(t) \in Y$ . Абстрактный конечный автомат имеет один входной и один выходной каналы. В каждый момент  $t = 0, 1, 2, \dots$  дискретного времени автомат находится в определенном состоянии  $z(t)$  из множества  $Z$  состояний автомата, причем в начальный момент времени  $t = 0$  он всегда находится в начальном состоянии  $z(0) = z_0$ . Конечным автоматом называется такой автомат, у которого множества входных сигналов, состояний и выходных сигналов являются конечными множествами. Абстрактно конечный автомат представляет собой математическую схему, характеризующуюся 6 элементами:

- конечным множеством  $X$  входных сигналов (входной алфавит);
- конечным множеством  $Z$  внутренних состояний (алфавитом состояний);
- конечным множеством  $Y$  выходных сигналов (выходным алфавитом);
- начальным состоянием  $z_0$ ;

- функцией переходов  $\varphi(z,x)$ ;
- функцией выходов  $\psi(z,x)$ .

В момент  $t$ , будучи в состоянии  $z(t-1)$ , автомат способен воспринять на входном канале сигнал  $x(t) \in X$  и выдать на выходном канале сигнал  $y(t) = \psi[z(t-1),x(t)]$ , переходя в состояние  $z(t) = \varphi[z(t-1),x(t)]$ ,  $z(t) \in Z$ ,  $y(t) \in Y$ . Сказанное выше можно описать следующими уравнениями: для автомата, называемого автоматом Мили

$$z(t) = \varphi[z(t-1),x(t)], t=1,2,\dots, \quad (3.1)$$

$$y(t) = \psi[z(t-1),x(t)], t=1,2,\dots \quad (3.2)$$

Автомат, для которого

$$y(t) = \psi[z(t)], t=1,2, \quad (3.3)$$

т.е. функция выходов не зависит от входной переменной, называется автоматом Мура.

По числу состояний различают конечные автоматы с памятью и без памяти. Автоматы с памятью имеют более одного состояния, а автоматы без памяти (комбинационные или логические) обладают лишь одним состоянием. При этом, согласно (3.2) работа комбинационной схемы заключается в том, что она ставит в соответствие каждому входному сигналу  $x(t)$  определенный выходной сигнал, т.е. реализует логическую функцию вида

$$y(t) = \psi[x(t)], t = 1,2,\dots$$

Это функция называется булевой, если алфавиты  $X$  и  $Y$ , которым принадлежат значения сигналов  $x$  и  $y$ , состоят из двух букв.

По характеру отсчета дискретного времени конечные автоматы делятся на синхронные и асинхронные. В синхронных автоматах моменты времени, в которые автомат «считывает» входные сигналы, определяются принудительно синхронизирующими сигналами. После очередного синхронизирующего сигнала с учетом «считанного» и в соответствии с уравнениями (3.1-3.3) происходит переход в новое состояние и выдача сигнала на выходе, после чего автомат может воспринимать следующее значение вход-

ного сигнала. Таким образом, реакция автомата на каждое значение входного сигнала заканчивается за один такт, длительность которого определяется интервалом между соседними синхронизирующими сигналами. Асинхронный автомат считывает входной сигнал непрерывно, и поэтому, реагируя на достаточно длинный входной сигнал постоянной величины  $x$ , он может, как следует из (3.1-3.3), несколько раз изменять состояние, выдавая соответствующее число выходных сигналов, пока не перейдет в устойчивое состояние, которое уже не может быть изменено данным входным сигналом.

Чтобы задать конечный автомат, необходимо описать все элементы кортежа  $\langle Z, X, Y, \varphi, \psi, z_0 \rangle$ . Существует несколько способов задания работы автомата, но наиболее часто используются табличный, графический и матричный.

Простейший табличный способ задания конечного автомата основан на использовании таблиц переходов и выходов, строки которых соответствуют входным сигналам автомата, а столбцы – его состояниям. На пересечении  $i$ -той строки и  $k$ -того столбца помещаются соответствующие значения  $\varphi(x_i, z_k)$  и  $\psi(x_i, z_k)$ .

Для автомата Мура обе таблицы можно совместить, получив так называемую отмеченную таблицу переходов, в которой над каждым состоянием  $z_k$  автомата, обозначающим столбец таблицы, стоит соответствующий этому состоянию, согласно (3.3), выходной сигнал  $\psi(z_k)$ .

При другом способе задания конечного автомата используется понятие направленного графа. Граф автомата представляет собой набор вершин, соответствующих различным состояниям автомата и соединяющих вершины дуг графа, соответствующих тем или иным переходам автомата. Если входной сигнал  $x_k$  вызывает переход из состояния  $z_i$  в состояние  $z_j$ , то на графе автомата дуга, соединяющая вершину  $z_i$  с вершиной  $z_j$ , обозначается  $x_k$ . Для того, чтобы задать функцию выходов, дуги графа необходимо отметить соответствующими выходными сигналами. Для автоматов Мили – дуга из  $z_i$  помечается парой  $x_k$  и  $y_s = \psi(z_i, x_k)$ . Для автоматов Мура – дуга,

направленная в  $z_j$ , отмечается только  $x_k$ , а  $y_s = \psi(z_j)$  приписывается узлу  $z_j$ . При решении задач моделирования систем часто более удобной формой является матричное задание конечного автомата. При этом матрица соединений автомата есть квадратная матрица  $C = \|c_{ij}\|$ , строки которой соответствуют исходным состояниям, а столбцы – состояниям перехода. Элемент  $c_{ij} = x_k/y_s$ , стоящий на пересечении  $i$ -той строки и  $j$ -того столбца, в случае автомата Мили соответствует входному сигналу  $x_k$ , вызывающему переход из состояния  $z_i$  в состояние  $z_j$ , и выходному сигналу  $y_s$ , выдаваемому при этом переходе. Если переход из состояния  $z_i$  в состояние  $z_j$  происходит под действием нескольких сигналов, элемент матрицы  $c_{ij}$  представляет собой множество пар «вход-выход» для этого перехода, соединенный знаком дизъюнкции.

Для автомата Мура элемент  $c_{ij}$  равен множеству входных сигналов на переходе  $(z_i, z_j)$ , а выход описывается вектором выходов,  $i$ -тая компонента которого – выходной сигнал, отмечающий состояние  $z_i$ .

Для детерминированных автоматов выполняется условие однозначности переходов: автомат, находящийся в некотором состоянии, под действием любого входного сигнала не может перейти более чем в одно состояние. Применительно к графическому способу задания автомата это означает, что в графе автомата из любой вершины не могут выходить два ребра и более, отмеченные одним и тем же входным сигналом. Аналогично этому в матрице соединений автомата  $C$  в каждой строке любой входной сигнал не должен встречаться более одного раза.

Понятие автомата в дискретно-детерминированном подходе к исследованию на моделях свойств объектов является математической абстракцией, удобной для описания широкого класса процессов функционирования реальных объектов в автоматизированных системах обработки информации и управления. В качестве таких объектов в первую очередь следует назвать

элементы и узлы ЭВМ, устройства контроля, регулирования и управления, системы временной и пространственной коммутации в технике обмена информацией и т.д. Для всех перечисленных объектов характерно наличие дискретных состояний и дискретный характер работы во времени. Но эта схема не универсальна. Данный подход не применим для описания процесса принятия решений, процессов в динамических системах с наличием переходных процессов и стохастических элементов.

Стохастические объекты (при моделировании которых учитываются случайные факторы), функционирующие в дискретном времени, можно представить вероятностными автоматами. Для такого автомата состояние  $z(t-1)$  и входной сигнал  $x(t)$  определяют не конечное состояние  $z(t)$ , а распределение вероятностей  $P_{ij}$  перехода автомата из состояния  $z_i = z(t-1)$  в одно из возможных состояний  $z_j(t)$  в момент времени  $t$  под воздействием входного сигнала  $x(t)$ .

Функция переходов вероятностного автомата определяет не одно конкретное состояние, а лишь распределение вероятностей на множестве состояний (автомат со случайными переходами), а функция выходов – распределение вероятностей на множестве выходных сигналов (автомат со случайными выходами). Функционирование вероятностных автоматов изучается при помощи аппарата цепей Маркова.

Математическими моделями стохастических объектов с непрерывным временем являются системы массового обслуживания (представители марковских случайных процессов).

### **3.2. Математические модели взаимодействия элементов сложной системы**

Взаимодействие элементов в процессе функционирования сложной системы рассматривается как результат совокупности воздействий каждого элемента на другие элементы. Воздействие представленное набором своих характеристик, часто называют сигналом, т.о. взаимодействие элементов сложной системы может быть рассмотрено в рамках механизма обмена сигналами. Сигналы передаются по каналам связи между элементами. Начало данного канала – выходной полюс, конец канала – входной полюс элемента. Идеальным каналом называется канал, в котором передача сигнала осуществляется мгновенно и без искажений. Полностью и правильно формализованная система имеет только идеальные каналы связей. Физические каналы связи не являются идеальными. Такие каналы связи необходимо рассматривать как самостоятельные элементы системы (электрические соединительные провода – отдельные резисторы), функционирование которых сводится к соответствующим задержкам и искажениям сигнала. Т.о. окончательная формализация моделируемого объекта может привести к сложной системе, которая по составу элементов и конфигурации связей между ними отличается от конструкции, полученной в результате первоначальной структуризации этого объекта. При построении математической модели сложной системы необходимо учитывать взаимодействие её с внешней средой. Внешняя среда рассматривается как некоторая совокупность объектов, воздействующих на элементы сложной системы, а также испытывающих воздействия, поступающие от элементов сложной системы. Механизм обмена сигналами и формализованная схема взаимодействия элементов сложной системы между собой и с объектами внешней среды включает наборы следующих составляющих:

- процесс формирования выходного сигнала соответствующим элементом системы;
- определение адреса передачи для каждого выходного сигнала;
- прохождение сигналов по каналам связи и компоновка входных сигналов для элементов системы, принимающих сигналы.
- реагирование элементов на поступающие входные сигналы.

Первая и четвертая составляющие описываются в рамках математических моделей элементов. Третья составляющая связана с заменой реальных физических каналов идеальными. Вторая составляющая механизма обмена сигналами в сложной системе обеспечивает адресацию характеристик выходных сигналов и их компоновку во входные сигналы элементов, т.е. схему сопряжения элементов (иногда говорят структуру связности).

*Схема сопряжения элементов системы.* Пусть сложная система  $S$  содержит элементы  $C_1, C_2, \dots, C_N$ . Предположение 1. Элементарные сигналы передаются в системе по элементарным каналам: каждый  $l$ -ый элементарный канал, подключенный к выходу элемента  $C_j$ , способен передавать только элементарные сигналы  $u_l^{(j)}$ , имеющие фиксированный индекс  $l$ . Внешнюю среду можно представить в виде фиктивного элемента  $C_0$  системы  $S$ , вход которого содержит  $m_0$  входных контактов  $X_l(0)$ , а выход  $r_0$  выходных контактов  $Y_l(0)$ . Каждый  $C_j$  (в том числе и  $C_0$ ) как элемент системы  $S$  в рамках принятых предположений о механизме обмена сигналами, достаточно характеризовать множеством входных контактов  $X_1^{(j)}, X_2^{(j)}, \dots, X_m^{(j)}$ , которые мы будем обозначать  $[X_i^{(j)}]_1^m$  и множеством выходных контактов  $Y_1^{(j)}, Y_2^{(j)}, \dots, Y_r^{(j)}$ , обозначаемым  $[Y_l^{(j)}]_1^r$ , где для простоты приняты обозначения  $m=m_j; r=r_j; j=0,1,\dots,N$ .

Другими словами, математической моделью элемента  $C_j$ , используемой для формального описания сопряжения его с прочими элементами системы

и внешней средой, является пара множеств:  $[X_i^{(j)}]_1^m$  и  $[Y_l^{(j)}]_1^r$ . Для исключения неоднозначности введем предположение 2: к входному контакту любого элемента системы подключается не более чем один элементарный канал; к выходному контакту может быть подключено любое конечное число элементарных каналов, при условии, что ко входу одного и того же элемента системы направляется не более чем один из упомянутых элементарных каналов.

Рассмотрим множество всех входных контактов всех элементов системы и внешней среды  $\bigcup_{j=0}^N [X_i^{(j)}]_1^m$ , а также всех выходных контактов  $\bigcup_{j=0}^N [Y_l^{(j)}]_1^r$ .

В силу второго предположения каждому входному контакту  $X_i^{(j)}$  соответствует не более чем один выходной контакт  $Y_l^{(k)}$ , с которым он связан элементарным каналом. Поэтому можно ввести однозначный оператор

$$Y_l^{(k)} = R(X_i^{(j)}) \quad (3.4) \text{ с областью определения во множестве } \bigcup_{j=0}^N [X_i^{(j)}]_1^m \text{ и об-}$$

ластью значений во множестве  $\bigcup_{j=0}^N [Y_l^{(j)}]_1^r$ , сопоставляющий входному кон-

такту  $X_i^{(j)}$  выходной контакт  $Y_l^{(k)}$ , связанный с ним элементарным каналом.

Если в рассматриваемой системе к данному контакту не подключен никакой элементарный канал, то оператор (3.4) не определен на этом  $X_i^{(j)}$ . Со-

вокупность множеств  $[X_i^{(j)}]_1^m$  и  $[Y_l^{(j)}]_1^r$  и оператора  $R$  будем называть схемой

сопряжения элементов в системе  $S$ , а оператор  $R$  - оператором сопряжения.

Оператор сопряжения можно задать в виде таблицы, в которой на пересечении строк с номерами элементов системы  $j$  и столбцов с номерами кон-

тактов  $i$  располагаются пары чисел  $(k,l)$ , указывающие номер элемента  $k$  и номер контакта  $l$ , с которым соединен контакт  $X_i^{(j)}$ . Другой способ задания

оператора  $R$  получим, если столбцы и строки таблицы пронумеровать двойными номерами  $(j,i)$  и  $(k,l)$  соответственно, а на пересечениях помещать 1

для контактов  $X_i^{(j)}$  и  $Y_l^{(k)}$ , соединенных элементарным каналом и 0 в противном случае. Хотя таблицы такого рода громоздки, они не редко находят применение, т.к. представляют собой матрицы смежности ориентированных графов, вершинами которых являются контакты, а ребрами – элементарными каналами. Заметим, что соответствие между  $Y_l^{(k)}$  и  $X_i^{(j)}$ , описываемое оператором  $R$ , не является взаимно однозначным (один и тот же выход может направляться на входы разных элементов). Рассмотрим сужение оператора  $R$  на множество  $[X_i^{(j)}]_1^m$ , т.е. оператор  $R_j$ , определенный для данного элемента системы  $C_j$  (строка в таблице с номером  $j$ ). Соответствие, описываемое оператором  $R_j$  для  $\forall j=0,1,..N$  является взаимно однозначным, в силу второй части второго предположения. Поэтому существует однозначный обратный оператор  $R_j^{-1}$ , сопоставляющий каждому контакту  $Y_l^{(k)}$  из множества  $R([X_i^{(j)}]_1^m)$  соответствующий контакт  $X_i^{(j)}$ . Схема сопряжения, задаваемая множествами  $\bigcup_{j=0}^N [X_i^{(j)}]_1^m$ ,  $\bigcup_{j=0}^N [Y_l^{(j)}]_1^r$  и оператором  $R$ , содержит исчерпывающие сведения о соединениях элементов системы элементарными каналами.

Математическим описанием элементов сложной системы на единой концептуальной основе и построением соответствующей схемы сопряжения элементов исчерпывается проблема построения математической модели функционирования сложной системы.

Для моделирования технических систем в настоящее время успешно применяются структурные методы, теория автоматического управления, методы электрического аналогизирования, метод графов связей, метод, использующий библиотеку моделей компонентов – элементов и связей, тензорные методы и др.

### 3.3. Вопросы для самопроверки

1. Назовите типовые математические схемы для описания элементов сложных систем.
2. Какими способами можно задать функции переходов и выходов конечного детерминированного автомата?
3. Чем вероятностный автомат отличается от детерминированного?
4. Какие составляющие содержит формализованная схема взаимодействия элементов сложной системы в рамках механизма обмена сигналами?
5. Что входит в схему сопряжения элементов системы и какими способами ее можно описать?

## ГЛАВА 4. МОДЕЛИРОВАНИЕ ПО СХЕМЕ МАРКОВСКИХ СЛУЧАЙНЫХ ПРОЦЕССОВ

Для вычисления числовых параметров, характеризующих стохастические объекты, нужно построить некоторую вероятностную модель явления, учитывающую сопровождающие его случайные факторы. Для математического описания многих явлений, развивающихся в форме случайного процесса, может быть с успехом применен математический аппарат, разработанный в теории вероятностей для так называемых марковских случайных процессов. Поясним это понятие. Пусть имеется некоторая физическая система  $S$ , состояние которой меняется с течением времени (под системой  $S$  может пониматься что угодно: техническое устройство, ремонтная мастерская, вычислительная машина и т.д.). Если состояние  $S$  меняется по времени случайным образом, говорят, что в системе  $S$  протекает случайный процесс. Примеры: процесс функционирования ЭВМ (поступление заказов на ЭВМ, вид этих заказов, случайные выходы из строя), процесс наведения на цель управляемой ракеты (случайные возмущения (помехи) в системе управления ракетой), процесс обслуживания клиентов в парикмахерской или ремонтной мастерской (случайный характер потока заявок (требований), поступивших со стороны клиентов).

Случайный процесс называется марковским процессом (или «процессом без последствия»), если для каждого момента времени  $t_0$  вероятность любого состояния системы в будущем (при  $t > t_0$ ) зависит только от её состояния в настоящем (при  $t = t_0$ ) и не зависит от того, когда и каким образом система пришла в это состояние (т.е. как развивался процесс в прошлом). Пусть  $S$  техническое устройство, которое характеризуется некоторой степенью изношенности  $S$ . Нас интересует, как оно будет работать дальше. В первом приближении характеристики работы системы в будущем (частота

отказов, потребность в ремонте) зависят от состояния устройства в настоящий момент и не зависят от того, когда и как устройство достигло своего настоящего состояния.

Теория марковских случайных процессов – обширный раздел теории вероятности с широким спектром приложений (физические явления типа диффузии или перемешивания шихты во время плавки в доменной печи, процессы образования очередей и т.д.).

#### **4.1. Классификация марковских процессов**

Марковские случайные процессы делятся на классы. Первый классификационный признак – характер спектра состояний. Случайный процесс (СП) называется процессом с дискретными состояниями, если возможные состояния системы  $S_1, S_2, S_3, \dots$  можно перечислить, а сам процесс состоит в том, что время от времени система  $S$  скачком (мгновенно) перескакивает из одного состояния в другое.

Пример. Техническое устройство состоит из двух узлов I и II, каждый из которых может выйти из строя. Состояния:  $S_1$  – оба узла работают;  $S_2$  – первый узел отказал, второй рабочий;  $S_3$  – второй узел отказал, первый рабочий;  $S_4$  – оба узла отказали.

Существуют процессы с непрерывными состояниями (плавный переход из состояния в состояние), например, изменение напряжения в осветительной сети. Будем рассматривать только СП с дискретными состояниями. В этом случае удобно пользоваться графом состояний, в котором возможные состояния системы обозначаются узлами, а возможные переходы – дугами.

Второй классификационный признак – характер функционирования во времени. СП называется процессом с дискретным временем, если перехо-

ды системы из состояния в состояние возможны только в строго определенные, заранее фиксированные моменты времени:  $t_1, t_2, \dots$ . Если переход системы из состояния в состояние возможен в любой наперед неизвестный случайный момент, то говорят о СП с непрерывным временем.

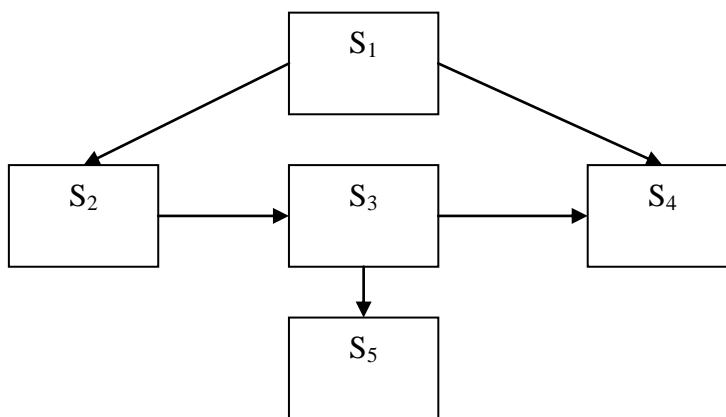


Рис. 4.1. Граф состояний системы

## 4.2. Расчет марковской цепи с дискретным временем

Пусть имеется физическая система  $S$  с дискретными состояниями  $S_1, S_2, \dots, S_n$  и дискретным временем  $t_1, t_2, \dots, t_k, \dots$  (шаги, этапы процесса, СП можно рассматривать как функцию аргумента (номера шага)). В общем случае СП состоит в том, что происходят переходы  $S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow S_1 \rightarrow \dots$  в моменты  $t_1, t_2, t_3, \dots$

Будем обозначать  $S_i^{(k)}$  событие, состоящее в том, что после  $k$  – шагов система находится в состоянии  $S_i$ . При любом  $k$  события  $S_1^k, S_2^k, \dots, S_n^k$  образуют полную группу и несовместны. СП, происходящий в системе, можно представить как последовательность событий  $S_1^{(0)}, S_2^{(1)}, S_1^{(2)}, S_2^{(3)}, S_3^{(4)}, \dots$

Такая случайная последовательность событий называется (в случае отсутствия последействия) марковской цепью. Будем описывать марковскую цепь (МЦ) с помощью вероятностей состояний. Пусть  $p_i(k)$  – вероятность

того, что после  $k$  - шагов система находится в состоянии  $S_i$ . Легко видеть, что  $\forall k \quad p_1(k) + p_2(k) + \dots + p_n(k) = 1$ . Поставим задачу: найти вероятности состояний системы для любого  $k$ .

Для любого шага (момента времени  $t_1, t_2, \dots, t_k$ ) существуют какие-то вероятности перехода системы из любого состояния в любое другое (некоторые из них равны нулю, если непосредственный переход за один шаг невозможен), а также вероятности задержки системы в одном состоянии. Будем их называть переходными вероятностями МЦ. Марковская цепь называется однородной, если переходные вероятности не зависят от номера шага, в противном случае - неоднородная МЦ. Рассмотрим однородную МЦ.

Пусть  $S = S_1, S_2, \dots, S_n$ . Обозначим переходные вероятности через  $P_{ij}$ . Пусть известна матрица

$$\|P_{ij}\| = \begin{vmatrix} P_{11} & P_{12} \dots & P_{1j} \dots & P_{1n} \\ P_{21} & P_{22} \dots & P_{2j} \dots & P_{2n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & P_{nj} & P_{nn} \end{vmatrix}.$$

Пользуясь введенными выше событиями  $S_1^k, S_2^k, \dots, S_n^k$ , переходные вероятности можно написать как условные вероятности:  $P_{ij} = P(S_j^k / S_i^{(k-1)})$ . Сумма членов в каждой строке матрицы должна быть равна 1. Вместо матрицы переходных вероятностей часто используют размеченный граф состояний (обозначают на дугах ненулевые вероятности переходов, вероятности задержки не требуются, поскольку они легко вычисляются, например  $P_{11} = 1 - \sum_2^n P_{1j}$ ). Имея в распоряжении размеченный граф состояний (или матрицу переходных вероятностей) и зная начальное состояние системы, можно найти вероятности состояний  $p_1(k), p_2(k), \dots, p_n(k) \quad \forall k$ .

Пусть начальное состояние системы  $S_m$ , тогда

$$p_1(0) = 0 \quad p_2(0) = 0 \dots p_m(0) = 1 \dots p_n(0) = 0.$$

Первый шаг:

$$p_1(1) = P_{m1}, \quad p_2(1) = P_{m2}, \dots, p_m(1) = P_{mm}, \dots, p_n(1) = P_{mn}.$$

После второго шага по формуле полной вероятности получим:

$$p_1(2) = p_1(1)P_{11} + p_2(1)P_{21} + \dots + p_n(1)P_{n1},$$

$$p_i(2) = p_1(1)P_{1i} + p_2(1)P_{2i} + \dots + p_n(1)P_{ni} \quad \text{или} \quad p_i(2) = \sum_{j=1}^n p_j(1)P_{ji}.$$

Для произвольного шага  $k$  получаем:

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

Для неоднородной МЦ переходные вероятности зависят от номера шага. Обозначим переходные вероятности для шага  $k$  через  $P_{ij}^{(k)} = P(S_j^{(k)} / S_i^{(k-1)})$ .

Тогда формула для расчета вероятностей состояний приобретает вид:

$$p_i(k) = \sum_{j=1}^n p_j(k-1)P_{ji}^{(k)}.$$

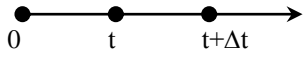
### 4.3. Марковские цепи с непрерывным временем

#### 4.3.1. Уравнения Колмогорова

На практике значительно чаще встречаются ситуации, когда переходы системы из состояния в состояние происходит в случайные моменты времени, которые заранее указать невозможно: например, выход из строя любого элемента аппаратуры, окончание ремонта (восстановление) этого элемента. Для описания таких процессов в ряде случаев может быть с успехом применена схема марковского случайного процесса с дискретными состояниями и непрерывным временем – непрерывная цепь Маркова. По-

кажем, как выражаются вероятности состояний для такого процесса. Пусть  $S = \{S_1, S_2, \dots, S_n\}$ . Обозначим через  $p_i(t)$  - вероятность того, что в момент  $t$  система  $S$  будет находиться в состоянии  $S_i (i = \overline{1, n})$ . Очевидно  $\sum_{i=1}^n p_i(t) = 1$ .

Поставим задачу – определить для любого  $t$   $p_i(t)$ . Вместо переходных вероятностей  $P_{ij}$  введем в рассмотрение плотности вероятностей перехода  $\lambda_{ij}$



$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t}$$

$$P_{ij}(\Delta t) \approx \lambda_{ij} \Delta t .$$

Если  $\lambda_{ij}$  не зависит от  $t$ , говорят об однородной цепи, иначе - о неоднородной. Пусть нам известны  $\lambda_{ij}$  для всех пар состояний (задан размеченный граф состояний). Оказывается, зная размеченный граф состояний можно определить  $p_1(t), p_2(t), \dots, p_n(t)$  как функции времени. Эти вероятности удовлетворяют определенного вида дифференциальным уравнениям, (уравнения Колмогорова).

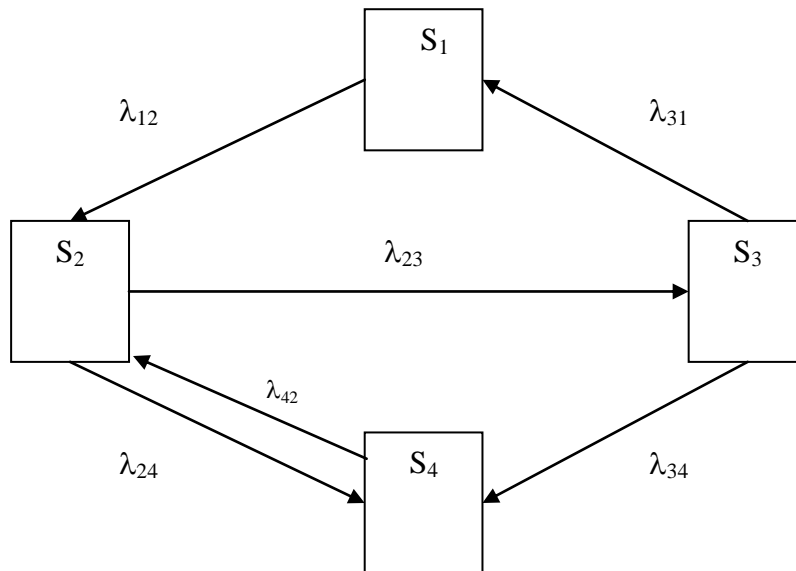


Рис. 4.2. Пример размеченного графа состояний для составления уравнений Колмогорова

Покажем на конкретном примере, как эти уравнения составляются. Пусть система  $S$  имеет четыре состояния:  $S_1, S_2, S_3, S_4$ , размеченный граф которых показан на рис. 4.2.

Рассмотрим одну из вероятностей состояний, например  $p_3(t)$ . Это - вероятность того, что в момент  $t$  система будет в состоянии  $S_3$ . Придадим  $t$  малое приращение  $\Delta t$  и найдем  $p_3(t + \Delta t)$  - вероятность того, что в момент  $t + \Delta t$  система будет в состоянии  $S_3$ . Как это может произойти? Очевидно, двумя способами: либо 1) в момент  $t$  система уже была в состоянии  $S_3$ , а за время  $\Delta t$  не вышла из него; либо 2) в момент  $t$  система была в состоянии  $S_2$ , а за время  $\Delta t$  перешла из него в  $S_3$ .

Найдем вероятность первого варианта. Вероятность того, что в момент  $t$  система была в состоянии  $S_3$ , равна  $p_3(t)$ . Эту вероятность нужно умножить на вероятность того, что, находясь в момент  $t$  в состоянии  $S_3$ , система за время  $\Delta t$  не перейдет из него ни в  $S_1$ , ни в  $S_4$ . Вероятность того, что за время  $\Delta t$  система выйдет из состояния  $S_3$ , равна  $(\lambda_{31} + \lambda_{34})\Delta t$  (так как события  $S_3 \rightarrow S_1, S_3 \rightarrow S_4$  несовместны и с точностью до бесконечно малых высших порядков); вероятность того, что не выйдет:  $1 - (\lambda_{31} + \lambda_{34})\Delta t$ . Отсюда вероятность первого варианта равна  $p_3(t) [1 - (\lambda_{31} + \lambda_{34})\Delta t]$ .

Найдем вероятность второго варианта. Она равна вероятности того, что в момент  $t$  система будет в состоянии  $S_2$ , а за время  $\Delta t$  перейдет из него в состояние  $S_3$ , т. е. она равна  $p_2(t)\lambda_{23}\Delta t$ .

Складывая вероятности обоих вариантов (по правилу сложения вероятностей), получим:

$$p_3(t + \Delta t) = p_3(t)[1 - (\lambda_{31} + \lambda_{34})\Delta t] + p_2(t)\lambda_{23}\Delta t$$

Раскроем квадратные скобки, перенесем  $p_3(t)$  в левую часть и разделим обе части на  $\Delta t$ :

$$\frac{p_3(t + \Delta t) - p_3(t)}{\Delta t} = \lambda_{23}p_2(t) - (\lambda_{31} + \lambda_{34})p_3(t).$$

Устремим, как и полагается в подобных случаях,  $\Delta t$  к нулю; слева получим в пределе производную функции  $p_3(t)$ . Таким образом, запишем дифференциальное уравнение для  $p_3(t)$ :

$$\frac{dp_3(t)}{dt} = \lambda_{23}p_2(t) - (\lambda_{31} + \lambda_{34})p_3(t),$$

или, короче, отбрасывая аргумент  $t$  у функций  $p_1, p_2$ :

$$\frac{dp_3}{dt} = \lambda_{23}p_2 - (\lambda_{31} + \lambda_{34})p_3.$$

Рассуждая аналогично для всех остальных состояний, напишем еще три дифференциальных уравнения. В результате получим систему дифференциальных уравнений для вероятностей состояний:

$$\left. \begin{aligned} \frac{dp_1}{dt} &= \lambda_{31}p_3 - \lambda_{12}p_1 \\ \frac{dp_2}{dt} &= \lambda_{12}p_1 + \lambda_{42}p_4 - (\lambda_{23} + \lambda_{24})p_2 \\ \frac{dp_3}{dt} &= \lambda_{23}p_2 - (\lambda_{31} + \lambda_{34})p_3 \\ \frac{dp_4}{dt} &= \lambda_{24}p_2 + \lambda_{34}p_3 - \lambda_{42}p_4 \end{aligned} \right\} \quad (4.1)$$

Это - система четырех линейных дифференциальных уравнений с четырьмя неизвестными функциями  $p_1, p_2, p_3, p_4$ . Заметим, что одно из них (любое) можно отбросить, пользуясь тем, что  $p_1 + p_2 + p_3 + p_4 = 1$ : выразить любую из вероятностей  $p_i$  через другие, это выражение подставить в (4.1),

а соответствующее уравнение с производной  $\frac{dp_i}{dt}$  отбросить.

Сформулируем теперь общее правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятности какого-то ( $i$ -го) состояния. В правой части - сумма произведений вероятности

стей всех состояний, из которых идут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного ( $i$ -го) состояния.

Чтобы решить уравнения Колмогорова и найти вероятности состояний, прежде всего надо задать начальные условия. Если мы точно знаем начальное состояние системы  $S_i$ , то в начальный момент (при  $t=0$ )  $p_i(0) = 1$ , а все остальные начальные вероятности равны нулю.

Как решать подобные уравнения? Вообще говоря, линейные дифференциальные уравнения с постоянными коэффициентами можно решать аналитически, но это удобно только когда число уравнений не превосходит двух (иногда - трех). Если уравнений больше, обычно их решают численно - вручную или на ЭВМ.

#### **4.3.2. Поток событий. Простейший поток и его свойства**

При рассмотрении процессов, протекающих в системе с дискретными состояниями и непрерывным временем, часто бывает удобно представить себе процесс так, как будто переходы системы из состояния в состояние происходят под действием каких-то потоков событий. Поток событий называется последовательность однородных событий, следующих одно за другим в какие-то, вообще говоря, случайные моменты времени (поток вызовов на телефонной станции; поток неисправностей (сбоев) ЭВМ; поток грузовых составов, поступающих на станцию; поток посетителей; поток выстрелов, направленных на цель). Будем изображать поток событий последовательностью точек на оси времени  $ot$ . Положение каждой точки на оси случайно. Поток событий называется регулярным, если события следуют одно за другим через строго определенные промежутки времени

(редко встречается на практике). Рассмотрим специального типа потоки, для этого введем ряд определений. 1. Поток событий называется *стационарным*, если вероятность попадания того или иного числа событий на участок времени длиной  $\tau$  зависит только от длины участка и не зависит от того, где именно на оси  $ot$  расположен этот участок (однородность по времени) – вероятностные характеристики такого потока не должны меняться от времени. В частности, так называемая интенсивность (или плотность) потока событий (среднее число событий в единицу времени) постоянна.

2. Поток событий называется *потоком без последствия*, если для любых непересекающихся участков времени число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой (или другие, если рассматривается больше двух участков). Отсутствие последствия в потоке означает, что события, образующие поток, появляются в последовательные моменты времени независимо друг от друга.

3. Поток событий называется *ординарным*, если вероятность попадания на элементарный участок двух или более событий пренебрежительно мала по сравнению с вероятностью попадания одного события (события в потоке приходят поодиночке, а не парами, тройками и т.д.).

Поток событий, обладающий всеми тремя свойствами, называется *простейшим* (или стационарным пуассоновским). Нестационарный пуассоновский поток обладает только свойствами 2 и 3. Пуассоновский поток событий (как стационарный, так и нестационарный) тесно связан с известным распределением Пуассона. А именно, число событий потока, попадающих на любой участок, распределено по закону Пуассона. Поясним это подробнее.

Рассмотрим на оси  $ot$ , где наблюдается поток событий, некоторый участок длины  $\tau$ , начинающийся в момент  $t_0$  и заканчивающийся в момент

$t_0 + \tau$ . Нетрудно доказать (доказательство дается во всех курсах теории вероятности), что вероятность попадания на этот участок ровно  $m$  событий выражается формулой:

$$P_m = \frac{a^m}{m!} e^{-a} \quad (m=0, 1 \dots),$$

где  $a$  – среднее число событий, приходящееся на участок  $\tau$ .

Для стационарного (простейшего) пуассоновского потока  $a = \lambda\tau$ , т.е. не зависит от того, где на оси  $ot$  взят участок  $\tau$ . Для нестационарного пуассоновского потока величина  $a$  выражается формулой

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

и значит, зависит от того, в какой точке  $t_0$  начинается участок  $\tau$ .

Рассмотрим на оси  $ot$  простейший поток событий с постоянной интенсивностью  $\lambda$ . Нас будет интересовать интервал времени  $T$  между событиями в этом потоке. Пусть  $\lambda$  - интенсивность (среднее число событий в единицу времени) потока. Плотность распределения  $f(t)$  случайной величины  $T$  (интервал времени между соседними событиями в потоке)  $f(t) = \lambda e^{-\lambda t}$  ( $t > 0$ ). Закон распределения с такой плотностью называется показательным (экспоненциальным). Найдем численные значения случайной величины  $T$ :

математическое ожидание (среднее значение)  $m_t = \int_0^{\infty} t f(t) dt = \frac{1}{\lambda}$  и дисперсию

$$D_t = \int_0^{\infty} t^2 f(t) dt - m_t^2 = \frac{1}{\lambda^2}.$$

Промежуток времени  $T$  между соседними событиями в простейшем потоке распределен по показательному закону; его среднее значение и среднее квадратичное отклонение равны  $\frac{1}{\lambda}$ , где  $\lambda$  - интенсивность потока.

Для такого потока вероятность появления на элементарном участке вре-

мени  $\Delta t$  ровно одного события потока выражается как  $P_1(\Delta t) \approx \lambda \Delta t$ . Эту вероятность мы будем называть «элементом вероятности появления события».

Для нестационарного пуассоновского потока закон распределения промежутка  $T$  уже не будет показательным. Вид этого закона будет зависеть, во первых, от того, где на оси от расположено первое из событий, во вторых, от вида зависимости  $\lambda(t)$ . Однако, если  $\lambda(t)$  меняется сравнительно медленно и его изменение за время между двумя событиями невелико, то закон распределения промежутка времени между событиями можно приближенно считать показательным, полагая в этой формуле величину  $\lambda$  равной среднему значению  $\lambda(t)$  на том участке, который нас интересует.

#### 4.3.3. Пуассоновские потоки событий и непрерывные марковские цепи

Рассмотрим некоторую физическую систему  $S = \{S_1, S_2, \dots, S_n\}$ , которая переходит из состояния в состояние под влиянием каких-то случайных событий (вызовы, отказы, выстрелы). Будем себе это представлять так, будто события, переводящие систему из состояния в состояние, представляют собой какие-то потоки событий.

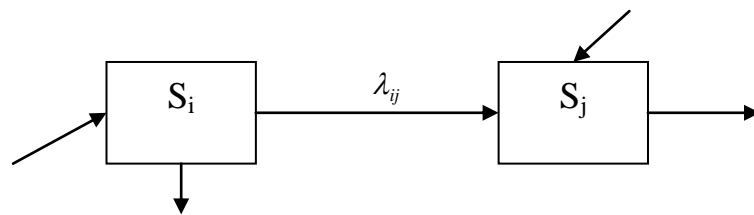


Рис. 4.3. Фрагмент графа состояний системы

Пусть система  $S$  в момент времени  $t$  находится в состоянии  $S_i$  и может перейти из него в состояние  $S_j$  под влиянием какого-то пуассоновского потока событий с интенсивностью  $\lambda_{ij}$  (рис. 4.3): как только появляется первое

событие этого потока, система мгновенно переходит из  $S_i$  в  $S_j$ . Как мы знаем, вероятность этого перехода за элементарный промежуток времени  $\Delta t$  (элемент вероятности перехода) равна  $\lambda_{ij}\Delta t$ , отсюда вытекает, что плотность вероятности перехода  $\lambda_{ij}$  в непрерывной цепи Маркова представляет собой не что иное, как интенсивность потока событий, переводящих систему по соответствующей стрелке. Если все потоки событий, переводящие систему  $S$  из состояния в состояние пуассоновские, то процесс, протекающий в системе, будет марковским.

Проставим интенсивности пуассоновских потоков (плотности вероятностей переходов) на графе состояний системы у соответствующих стрелок. Получим размеченный граф состояний. На его основе можно написать уравнения Колмогорова и вычислить вероятности состояний.

#### 4.3.4. Предельные вероятности состояний

Пусть имеется физическая система  $S = \{S_1, S_2, \dots, S_n\}$ , в которой протекает марковский случайный процесс с непрерывным временем (непрерывная цепь Маркова). Предположим, что  $\lambda_{ij} = \text{const}$ , т.е. все потоки событий простейшие (стационарные пуассоновские). Записав систему дифференциальных уравнений Колмогорова для вероятностей состояний и проинтегрировав эти уравнения при заданных начальных условиях, мы получим  $p_1(t)$ ,  $p_2(t), \dots, p_n(t)$ ,  $\sum_{i=1}^n p_i(t) = 1$  при любом  $t$ . Поставим следующий вопрос, что будет происходить с системой  $S$  при  $t \rightarrow \infty$ . Будут ли функции  $p_i(t)$  стремиться к каким-то пределам? Эти пределы, если они существуют, называются предельными (финальными) вероятностями состояний. Можно доказать теорему: если число состояний  $S$  конечно и из каждого состояния можно перейти (за то или иное число шагов) в каждое другое, то предель-

ные вероятности состояний существуют и не зависят от начального состояния системы. Предположим, что поставленное условие выполнено и предельные вероятности существуют  $\lim_{t \rightarrow \infty} p_i(t) = p_i$  ( $i=1, 2, \dots, n$ ),  $\sum_{i=1}^n p_i = 1$ .

Таким образом, при  $t \rightarrow \infty$  в системе  $S$  устанавливается некоторый предельный стационарный режим. Смысл этой вероятности: она представляет собой не что иное, как среднее относительное время пребывания системы в данном состоянии. Для вычисления  $p_i$  в системе уравнений Колмогорова, описывающих вероятности состояний, нужно положить все левые части (производные) равными 0. Систему получающихся линейных алгебраических уравнений надо решать совместно с уравнением  $\sum_{i=1}^n p_i = 1$ .

#### 4.3.5. Схема гибели и размножения

Мы знаем, что имея в распоряжении размеченный граф состояний, можно легко написать уравнения Колмогорова для вероятностей состояний, а также написать и решить алгебраические уравнения для финальных вероятностей. Для некоторых случаев удастся последние уравнения решить заранее, в буквенном виде. В частности, это удастся сделать, если граф состояний системы представляет собой так называемую «схему гибели и размножения».

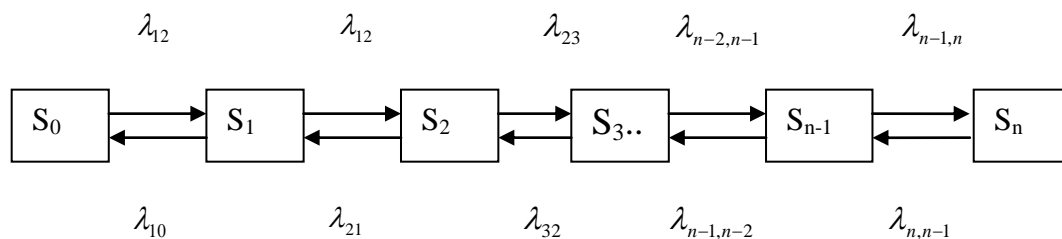


Рис. 4.4. Схема «гибели и размножения»

Граф состояний для схемы гибели и размножения имеет вид, показанный на рис. 4.4. Особенность этого графа в том, что все состояния системы можно вытянуть в одну цепочку, в которой каждое из средних состояний ( $S_1, S_2, \dots, S_{n-1}$ ) связано прямой и обратной стрелкой с каждым из соседних состояний — правым и левым, а крайние состояния ( $S_0, S_n$ ) — только с одним соседним состоянием. Термин «схема гибели и размножения» ведет начало от биологических задач, где подобной схемой описывается изменение численности популяции.

Схема гибели и размножения очень часто встречается в разных задачах практики, в частности — в теории массового обслуживания, поэтому полезно, один раз и навсегда, найти для нее финальные вероятности состояний.

Предположим, что все потоки событий, переводящие систему по стрелкам графа,— простейшие (для краткости будем называть и систему  $S$  и протекающий в ней процесс — простейшими).

Пользуясь графом рис. 4.4, составим и решим алгебраические уравнения для финальных вероятностей состояний (их существование вытекает из того, что из каждого состояния можно перейти в каждое другое, и число состояний конечно). Для первого состояния  $S_0$  имеем:

$$p_0 \lambda_{01} = \lambda_{10} p_1. \quad (4.2)$$

Для второго состояния  $S_1$ :

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

В силу (4.2) последнее равенство приводится к виду

$$p_1 \lambda_{12} = \lambda_{21} p_2$$

далее, совершенно аналогично

$$p_2 \lambda_{23} = \lambda_{32} p_3$$

и вообще

$$p_{k-1}\lambda_{k-1,k} = \lambda_{k,k-1}p_k,$$

где  $k$  принимает все значения от 0 до  $n$ . Итак, финальные вероятности  $p_0, p_1, \dots, p_n$  удовлетворяют уравнениям

$$\left. \begin{aligned} p_0\lambda_{01} &= \lambda_{10}p_1 \\ p_1\lambda_{12} &= \lambda_{21}p_2 \\ \dots\dots\dots \\ p_{k-1}\lambda_{k-1,k} &= \lambda_{k,k-1}p_k \\ \dots\dots\dots \\ p_{n-1}\lambda_{n-1,n} &= \lambda_{n,n-1}p_n \end{aligned} \right\} \quad (4.3)$$

кроме того, надо учесть нормировочное условие

$$p_0 = p_1 + p_2 + \dots + p_n = 1. \quad (4.4)$$

Решим эту систему уравнений. Из первого уравнения (4.3) выразим  $p_1$  через  $p_0$ .

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0. \quad (4.5)$$

Из второго, с учетом (4.5), получим:

$$p_2 = \frac{\lambda_{12}}{\lambda_{21}} p_1 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0 \quad (4.6)$$

из третьего, с учетом (4.6),

$$p_3 = \frac{\lambda_{23}}{\lambda_{32}} \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0 \quad (4.7)$$

и вообще, для любого  $k$  (от 1 до  $N$ ):

$$p_k = \frac{\lambda_{k-1,k} \dots \lambda_{12}\lambda_{01}}{\lambda_{k,k-1} \dots \lambda_{21}\lambda_{10}} p_0. \quad (4.8)$$

Обратим внимание на формулу (4.8). В числителе стоит произведение всех интенсивностей, стоящих у стрелок, ведущих слева направо (с начала и до данного состояния  $S_k$ ), а в знаменателе — произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево (с начала и до  $S_k$ ).

Таким образом, все вероятности состояний  $p_1, p_2, \dots, p_n$  выражены через одну из них ( $p_0$ ). Подставим эти выражения в нормировочное условие (4.4). Получим, вынося за скобку  $p_0$ :

$$p_0 \left( 1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}} \right) = 1$$

отсюда получим выражение для  $p_0$ .

$$p_0 = \left( 1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}} \right)^{-1} \quad (4.9)$$

(скобку мы возвели в степень  $-1$ , чтобы не писать двухэтажных дробей). Все остальные вероятности выражены через  $p_0$  (см. формулы (4.5) — (4.8)). Заметим, что коэффициенты при  $p_0$  в каждой из них представляют собой не что иное, как последовательные члены ряда, стоящего после единицы в формуле (4.9). Значит, вычисляя  $p_0$ , мы уже нашли все эти коэффициенты.

Полученные формулы очень полезны при решении простейших задач теории массового обслуживания.

#### 4.4. Вопросы для самопроверки

1. По каким признакам производят классификацию марковских случайных процессов?
2. Чем отличаются расчеты однородной и неоднородной марковской цепи с дискретным временем?
3. Как от размеченного графа состояний марковской цепи с дискретным временем перейти к матрице переходных вероятностей?
4. Сформулируйте правила составления уравнений Колмогорова для марковской цепи с непрерывным временем.
5. Какими свойствами должен обладать простейший поток событий?

6. В каких случаях существуют предельные вероятности состояний для непрерывных марковских цепей и каким образом эти вероятности определяются?

7. Начертите граф состояний системы для схемы гибели и размножения.

## ГЛАВА 5. ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ

В разных областях техники, в организации производства, в социальной сфере и в военном деле постоянно возникает необходимость решения вероятностных задач, связанных с работой систем массового обслуживания (СМО) разного вида требований. Термин «массовое обслуживание» предполагает многократную повторяемость ситуаций в том или ином смысле (много прибывших в систему и обслуженных заявок, большое число находящихся в эксплуатации аналогичных систем) и статистическую устойчивость картины. Выводы и рекомендации, получаемые методами теории массового обслуживания (ТМО), применимы лишь при наличии одного или обоих перечисленных факторов повторяемости.

Примерами СМО могут служить: телефонные станции, ремонтные мастерские, билетные кассы, справочные бюро, магазины, парикмахерские и т. п. Каждая СМО состоит из какого-то числа обслуживающих единиц (или «приборов»), которые мы будем называть каналами обслуживания. Каналами могут быть: линии связи, рабочие точки, кассиры, продавцы, лифты, автомашины и др. СМО могут быть одноканальными и многоканальными.

Всякая СМО предназначена для обслуживания какого-то потока заявок (или «требований»), поступающих в какие-то случайные моменты времени. Обслуживание заявки продолжается какое-то, вообще говоря, случайное время  $T_{об}$ , после чего канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времен обслуживания приводит к тому, что в какие-то периоды времени на входе СМО скапливается излишне большое число заявок (они либо становятся в очередь, либо покидают СМО необслуженными); в другие же периоды СМО будет работать с недогрузкой или вообще простаивать.

В СМО происходит какой-то СП с дискретными состояниями и непрерывным временем; состояние СМО меняется скачком в моменты появления каких-то событий (или прихода новой заявки, или окончания обслуживания, или момента, когда заявка, которой надоело ждать, покидает очередь). Чтобы дать рекомендации по рациональной организации этого процесса и предъявить разумные требования к СМО, необходимо изучить СП, описать его математически. Этим и занимается ТМО.

Предмет теории массового обслуживания — построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками — показателями эффективности СМО, описывающими, с той или другой точки зрения, ее способность справляться с потоком заявок. В качестве таких показателей (в зависимости от обстановки и целей исследования) могут применяться разные величины, например: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналов; среднее число заявок в очереди и среднее время ожидания обслуживания; вероятность того, что число заявок в очереди превысит какое-то значение, и т. д. Область применения математических методов ТМО непрерывно расширяется и все больше выходит за пределы задач, связанных с «обслуживающими организациями» в буквальном смысле слова.

Легче указать ситуации, где не может быть использована ТМО, чем перечислить все сферы ее потенциального применения. К таким сферам, в частности, относятся:

- в технике связи — проектирование телефонных станций и сетей связи;
- на транспорте — анализ процессов дорожного движения, прохождения тоннелей, очередей у светофоров, технического осмотра автомашин, формирования железнодорожных составов на сортировочных станциях, регистрация пассажиров, взлета и посадки самолетов, диспетчерской службы в аэропортах, разгрузки и погрузки судов в речных и морских портах, резервирования билетов;

- в промышленности – при планировании сборочных операций (расчет числа линий и емкости бункеров), гибких автоматизированных производств, организации и ремонте оборудования, расчете объема восстанавливаемого ЗИП, работе пунктов обслуживания (например, инструментальных складов), расчете зон обслуживания (в прядильном и ткацком производстве), определении числа резервных забоев в шахтах;

- в автоматизированных системах – для оценки своевременности обслуживания заявок на вычислительные работы и подготовку данных, обработки результатов экспериментов, управления технологическими процессами, при исследовании производительности вычислительных систем с асинхронно работающими устройствами;

- в торговле – для определения количества магазинов, продавцов, фасовщиц, узлов расчета, торговых и кассовых автоматов;

- в здравоохранении – при определении необходимого количества аптек, больничных коек, станций и бригад скорой помощи, врачей и младшего медицинского персонала, потребностей в диагностической и лечебной аппаратуре, пропускной способности специализированных санаториев;

- в службе быта – при расчете сети парикмахерских, мастерских разного назначения, количества техники и работников в них, транспортных средств обеспечения услуг населению, пунктов приема коммунальных платежей, количества ремонтников в жилищно-эксплуатационных конторах и трестах, сотрудников бюро по трудоустройству, по обмену жилой площади и т.д.;

- в органах юстиции и внутренних дел – при расчете количества судебных органов и их штатной численности, емкости исправительно-трудовых учреждений, численности сотрудников следственного аппарата и оперативников, планировании работы контрольно-пропускных пунктов и пунктов таможенного досмотра;

- в сфере науки и образования – при исследовании некоторых видов природных процессов (регистрация элементарных частиц, фильтрация, диффузия, действие катализаторов химических реакций), запуске ИСЗ и обработке спутниковой информации, расчете количества лабораторных установок, экзаменаторов (в том числе автоматических), возможностей учебно-опытных производств, бытового и медицинского обслуживания студентов, при проектировании и анализе работы крупных библиотек;

- в издательском деле – при расчете численности сотрудников, участвующих в подготовке рукописей к изданию;

- в военном деле – при проектировании систем ПВО (каждая цель может рассматриваться как «заявка» на обслуживание, то есть обстрел), организации охраны границ, расчете патрульных нарядов и во многих других случаях типа рассмотренных выше, но применительно к боевой и повседневной деятельности войск.

Задачи указанных типов приходится решать не только при проектировании вновь создаваемых систем и сетей обслуживания, но и в процессе эксплуатации имеющихся:

- при увеличении нагрузки;
- изменении трудоемкости обработки заявки;
- выходе из строя, деградации или модернизации техники;
- снижении квалификации персонала;
- пересмотре требований к оперативности обработки заявок и т.п.

Математический анализ работы СМО очень облегчается, если процесс этой работы — марковский. Для этого достаточно, чтобы все потоки событий, переводящие систему из состояния в состояние (потоки заявок, потоки «обслуживаний»), были простейшими. Если это свойство нарушается, то математическое описание процесса становится гораздо сложнее и довести его до явных, аналитических формул удастся лишь в редких случаях. Однако все же аппарат простейшей, марковской теории массового обслуживания может пригодиться для приближенного описания работы СМО даже в тех случаях, когда потоки событий — не простейшие. Во многих случаях для принятия разумного решения по организации работы СМО вовсе и не требуется точного знания всех ее характеристик — зачастую достаточно и приближенного, ориентировочного. Причем, чем сложнее СМО, чем больше в ней каналов обслуживания, тем точнее оказываются эти приближенные формулы.

### **5.1. Классификация СМО и их основные характеристики**

Системы массового обслуживания делятся на типы (или классы) по ряду признаков. Первое деление: СМО *с отказами* и СМО *с очередью*. В

СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует. Примеры СМО с отказами встречаются в телефонии: заявка на разговор, пришедшая в момент, когда все каналы связи заняты, получает отказ и покидает СМО необслуженной. В СМО с очередью заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь и ожидает возможности быть обслуженной. На практике чаще встречаются (и имеют большее значение) СМО с очередью; недаром теория массового обслуживания имеет второе название: «теория очередей».

СМО с очередью подразделяются на разные виды, в зависимости от того, как организована очередь—ограничена она или не ограничена. Ограничения могут касаться как длины очереди, так и времени ожидания (так называемые «СМО с нетерпеливыми заявками»). При анализе СМО должна учитываться также и «дисциплина обслуживания» — заявки могут обслуживаться либо в порядке поступления (раньше пришла, раньше обслуживается), либо в случайном порядке. Нередко встречается так называемое обслуживание с приоритетом — некоторые заявки обслуживаются вне очереди. Приоритет может быть как абсолютным — когда заявка с более высоким приоритетом «вытесняет» из-под обслуживания заявку с низшим (например, пришедший в парикмахерскую клиент высокого ранга прогоняет с кресла обыкновенного клиента), так и относительным — когда начатое обслуживание доводится до конца, а заявка с более высоким приоритетом имеет лишь право на лучшее место в очереди.

Существуют СМО с так называемым *многофазовым* обслуживанием, состоящим из нескольких последовательных этапов или «фаз» (например, покупатель, пришедший в магазин, должен сначала выбрать товар, затем оплатить его в кассе, затем получить на контроле).

Кроме этих признаков, СМО делятся на два класса: «открытые» и «замкнутые». В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято). В замкнутой СМО — зависят. Например, если один рабочий обслуживает группу станков, время от времени требующих наладки, то интенсивность потока «требований» со стороны станков зависит от того, сколько их уже неисправно и ждет наладки. Это — пример замкнутой СМО.

В зависимости от типа СМО при оценке её эффективности могут применяться те или иные величины (показатели эффективности). Например, для СМО с отказами одной из важнейших характеристик её продуктивности является так называемая абсолютная пропускная способность – среднее число заявок, которое может обслужить система за единицу времени. Наряду с абсолютной, часто рассматривается относительная пропускная способность – средняя доля поступивших заявок, обслуживаемая системой (отношение среднего числа обслуживаемых в единицу времени заявок к среднему числу поступающих заявок за это время). Помимо этого при анализе СМО с отказами могут интересоваться ещё среднее число занятых каналов, среднее относительное время простоя системы в целом и отдельного канала и т.д.

Характеристики СМО с ожиданиями. Для СМО с неограниченным ожиданием абсолютные и относительные пропускные способности теряют смысл. Зато важными являются: среднее число заявок в очереди, среднее число заявок в системе (в очереди и под обслуживанием), среднее время ожидания заявки в очереди, среднее время пребывания заявки в системе и другие. Для СМО с ограниченным ожиданием интерес представляют обе группы характеристик.

Для анализа процесса, протекающего в СМО, существенно знать основные параметры системы: число каналов  $n$ , интенсивность потока заявок

$\lambda$ , производительность каждого канала (среднее число заявок  $\mu$ , обслуживаемых непрерывно занятым каналом в единицу времени), условия образования очереди (ограничения, если они есть).

Условимся все потоки событий, переводящие СМО из состояния в состояние, считать пуассоновскими.

## 5.2. Одноканальная СМО с отказами

Рассмотрим простейшую из всех задач теории массового обслуживания – задачу о функционировании одноканальной СМО с отказами. Пусть система массового обслуживания состоит только из одного канала ( $n=1$ ) и на нее поступает пуассоновский поток заявок с интенсивностью  $\lambda$ , зависящей, в общем случае, от времени:

$$\lambda = \lambda(t) . \quad (5.1)$$

Заявка, заставшая канал занятым, получает отказ и покидает систему.

Обслуживание заявки продолжается в течение случайного времени  $T_{об}$ , распределенного по показательному закону с параметром  $\mu$ :

$$f(t) = \mu e^{-\mu t} (t > 0) . \quad (5.2)$$

Из этого следует, что «поток обслуживаний» - простейший, с интенсивностью  $\mu$ . Чтобы представить себе реально этот поток, вообразим один непрерывно занятый канал – он будет выдавать обслуженные заявки потоком с интенсивностью  $\mu$ .

Требуется найти:

абсолютную пропускную способность СМО ( $A$ );

относительную пропускную способность СМО ( $q$ ).

Рассмотрим единственный канал обслуживания как физическую систему  $S$ , которая может находиться в одном из двух состояний:

$S_0$  – свободен,

$S_1$  – занят.

Граф состояний системы показан на рис. 5.1.

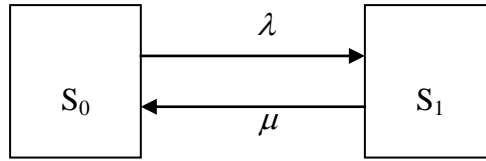


Рис. 5.1. Граф состояний одноканальной СМО с отказами

Из состояния  $S_0$  в  $S_1$  систему, очевидно, переводит поток заявок с интенсивностью  $\lambda$ ; из  $S_1$  в  $S_0$  – «поток обслуживаний» с интенсивностью  $\mu$ . Обозначим вероятности состояний  $p_0(t)$  и  $p_1(t)$ . Очевидно, для любого момента  $t$

$$p_0(t) + p_1(t) = 1. \quad (5.3)$$

Составим дифференциальные уравнения Колмогорова для вероятностей состояний согласно правилу, данному в 4.3.1. Имеем:

$$\left. \begin{aligned} \frac{dp_0}{dt} &= -\lambda p_0 + \mu p_1 \\ \frac{dp_1}{dt} &= -\mu p_1 + \lambda p_0 \end{aligned} \right\} . \quad (5.4)$$

Из двух уравнений (5.4) одно является лишним, так как  $p_0$  и  $p_1$  связаны отношением (5.3). Учитывая это, отбросим второе уравнение, а в первое подставим вместо  $p_1$  его выражение  $(1 - p_0)$ :

$$\frac{dp_0}{dt} = -\lambda p_0 + \mu(1 - p_0)$$

или

$$\frac{dp_0}{dt} = -(\mu + \lambda)p_0 + \mu \quad (5.5)$$

Это уравнение естественно решать при начальных условиях:

$p_0(0) = 1, p_1(0) = 0$  (в начальный момент канал свободен).

Линейное дифференциальное уравнение (5.5) с одной неизвестной функцией  $p_0$  легко может быть решено не только для простейшего потока

заявок ( $\lambda = \text{const}$ ), но и для случая, когда интенсивность этого потока со временем меняется ( $\lambda = \lambda(t)$ ). Не останавливаясь на последнем случае, приведем решение уравнения (5.5) только для случая  $\lambda = \text{const}$ :

$$p_0 = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \quad . \quad (5.6)$$

Зависимость величины  $p_0$  от времени имеет вид, изображенный на рис. 5.2. В начальный момент (при  $t = 0$ ) канал заведомо свободен ( $p_0(0) = 1$ ). С увеличением  $t$  вероятность  $p_0$  уменьшается и в пределе (при  $t \rightarrow \infty$ ) равна  $\frac{\mu}{\lambda + \mu}$ . Величина  $p_1(t)$ , дополняющая  $p_0(t)$  до единицы, изменяется как показано на том же рис. 5.2.

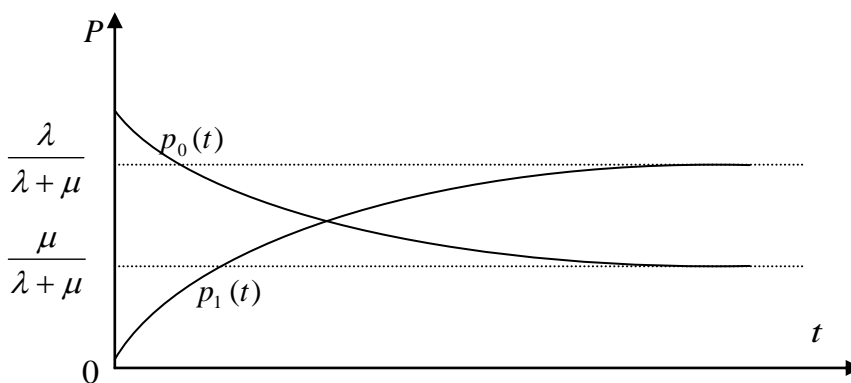


Рис. 5.2. Результаты решения системы уравнений (5.4)

Нетрудно убедиться, что для одноканальной СМО с отказами вероятность  $p_0$  есть не что иное, как относительная пропускная способность  $q$ .

Действительно,  $p_0$  есть вероятность того, что в момент  $t$  канал свободен, иначе вероятность того, что заявка, пришедшая в момент  $t$ , будет обслужена. А значит, для данного момента времени  $t$  среднее отношение числа обслуженных заявок к числу поступивших также равно  $p_0$ :  $q = p_0$ .

В пределе, при  $t \rightarrow \infty$ , когда процесс обслуживания уже установится, предельное значение относительной пропускной способности будет равно:

$$q = \frac{\mu}{\lambda + \mu} \quad . \quad (5.7)$$

Зная относительную пропускную способность  $q$ , легко найти абсолютную пропускную способность  $A$ , они связаны очевидным соотношением:

$$A = \lambda q \quad . \quad (5.8)$$

В пределе, при  $t \rightarrow \infty$ , абсолютная пропускная способность тоже установится и будет равна

$$A = \frac{\lambda \mu}{\lambda + \mu} \quad . \quad (5.9)$$

Зная относительную пропускную способность системы  $q$  (вероятность того, что пришедшая в момент  $t$  заявка будет обслужена), легко найти вероятность отказа:

$$P_{отк} = 1 - q \quad . \quad (5.10)$$

Вероятность отказа  $P_{отк}$  есть не что иное, как средняя доля необслуженных заявок среди поданных.

В пределе, при  $t \rightarrow \infty$ ,

$$P_{отк} = 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu} \quad . \quad (5.11)$$

### 5.3. Многоканальная СМО с отказами

Рассмотрим  $n$ -канальную СМО с отказами. Будем нумеровать состояния системы по числу занятых каналов (или, что в данном случае то же самое, по числу заявок, связанных с системой). Состояния будут:

$S_0$  – все каналы свободны,

$S_1$  – занят ровно один канал, остальные свободны,

.....

$S_k$  – заняты ровно  $k$  каналов, остальные свободны,

.....  
 $S_n$  – заняты все  $n$  каналов.

Граф состояний СМО представлен на рис. 5.3. Разметим граф, т.е. проставим у стрелок интенсивности соответствующих потоков событий. По стрелкам слева направо систему переводит один и тот же поток – поток заявок с интенсивностью  $\lambda$ . Если система находится в состоянии  $S_k$  (занято  $k$  каналов) и пришла новая заявка, система переходит (перескакивает) в состояние  $S_{k+1}$ .

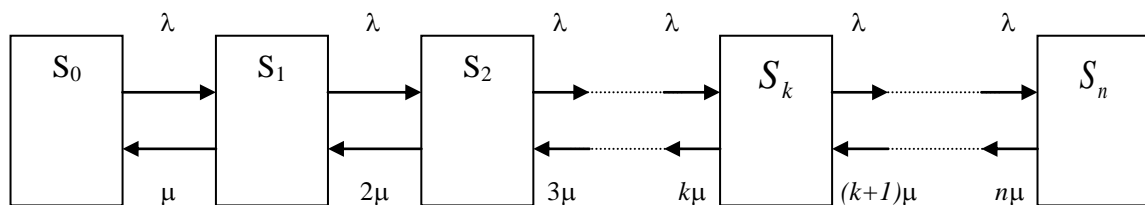


Рис. 5.3. Граф состояний многоканальной СМО с отказами

Определим интенсивности потоков событий, переводящих систему по стрелкам справа налево.

Пусть система находится в состоянии  $S_1$  (занят один канал). Тогда, как только закончится обслуживание заявки, занимающей этот канал, система перейдет в  $S_0$ ; значит, поток событий, переводящий систему по стрелке  $S_1 \rightarrow S_0$ , имеет интенсивность  $\mu$ . Очевидно, если обслуживанием занято два канала, а не один, поток обслуживаний, переводящий систему по стрелке  $S_2 \rightarrow S_1$ , будет вдвое интенсивнее ( $2\mu$ ); если занято  $k$  каналов – в  $k$  раз интенсивнее ( $k\mu$ ). Проставим соответствующие интенсивности у стрелок, ведущих справа налево.

Из рис. 5.3 видно, что процесс, протекающий в СМО, представляет собой частный случай процесса гибели и размножения.

Пользуясь общими правилами, можно составить уравнения Колмогорова для вероятностей состояний:

$$\left. \begin{aligned} \frac{dp_0}{dt} &= -\lambda p_0 + \mu p_1, \\ \frac{dp_1}{dt} &= -(\lambda + \mu)p_1 + \lambda p_0 + 2\mu p_2, \\ &\dots\dots\dots \\ \frac{dp_k}{dt} &= -(\lambda + k\mu)p_k + \lambda p_{k-1} + (k+1)\mu p_{k+1}, \\ &\dots\dots\dots \\ \frac{dp_n}{dt} &= -n\mu p_n + \lambda p_{n-1}. \end{aligned} \right\} \quad (5.12)$$

Уравнения (5.12) называются уравнениями Эрланга. Естественными начальными условиями для их решения являются:

$$p_0(0) = 1, p_1(0) = p_2(0) = \dots = p_n(0) = 0$$

(в начальный момент система свободна).

Интегрирование системы уравнений (5.12) в аналитическом виде довольно сложно; на практике такие системы дифференциальных уравнений обычно решаются численно. Такое решение дает все вероятности состояний  $p_0(t), p_1(t), \dots, p_k(t), \dots, p_n(t)$  как функции времени.

На практике больше всего интересуются предельными вероятностями состояний  $p_0, p_1, \dots, p_k, \dots, p_n$ , характеризующими установившийся режим работы СМО (при  $t \rightarrow \infty$ ). Для нахождения предельных вероятностей воспользуемся уже готовым решением задачи, полученным для схемы гибели и размножения. Согласно этому решению,

$$\left. \begin{aligned} p_k &= \frac{\lambda^k}{\mu \cdot 2\mu \dots k\mu} p_0 = \frac{(\lambda/\mu)^k}{k!} p_0 (k = 1, 2, \dots, n); \\ p_0 &= \frac{1}{1 + \frac{\lambda/\mu}{1!} + \frac{(\lambda/\mu)^2}{2!} + \dots + \frac{(\lambda/\mu)^n}{n!}} \end{aligned} \right\} \quad (5.13)$$

В этих формулах интенсивность потока заявок  $\lambda$  и интенсивность потока обслуживания (для одного канала)  $\mu$  не фигурируют по отдельности, а входят только своим отношением  $\lambda/\mu$ . Обозначим это отношение  $\lambda/\mu = \rho$  и будем называть величину  $\rho$  «приведенной интенсивностью» потока заявок. Физический смысл ее таков: величина  $\rho$  представляет собой среднее число заявок, приходящих в СМО за среднее время обслуживания одной заявки.

С учетом этого обозначения формулы (5.13) примут вид:

$$\left. \begin{aligned} p_k &= \frac{\rho^k}{k!} p_0 (k = 1, 2, \dots, n); \\ p_0 &= \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!}} = \left[ 1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} \right]^{-1}. \end{aligned} \right\} \quad (5.14)$$

Формулы (5.14) называются формулами Эрланга. Зная  $p_0, p_1, \dots, p_k, \dots, p_n$  можно найти характеристики эффективности СМО: относительную пропускную способность  $q$ , абсолютную пропускную способность  $A$  и вероятность отказа  $P_{отк}$ .

Действительно, заявка получает отказ, если приходит в момент, когда все  $n$  каналов заняты. Вероятность этого равна

$$P_{отк} = p_n = \frac{\rho^n}{n!} p_0 \quad . \quad (5.15)$$

Вероятность того, что заявка будет принята к обслуживанию (она же относительная пропускная способность  $q$ ) дополняет  $P_{отк}$  до единицы:

$$q = 1 - p_n \quad . \quad (5.16)$$

Абсолютная пропускная способность:

$$A = \lambda q = \lambda(1 - p_n) \quad . \quad (5.17)$$

Одной из важных характеристик СМО с отказами является среднее число занятых каналов (в данном случае оно совпадает со средним числом заявок, находящихся в системе). Обозначим это среднее число  $\bar{z}$ .

Величину  $\bar{z}$  можно вычислить непосредственно через вероятности  $p_0, p_1, \dots, p_n$  по формуле:

$$\bar{z} = 0 \cdot p_0 + 1 \cdot p_1 + \dots + n \cdot p_n$$

как математическое ожидание дискретной случайной величины, принимающей значения  $0, 1, \dots, n$  с вероятностями  $p_0, p_1, \dots, p_n$ . Однако значительно проще выразить среднее число занятых каналов через абсолютную пропускную способность  $A$ , которая уже известна. Действительно,  $A$  есть не что иное, как среднее число заявок, обслуживаемых в единицу времени; один занятый канал обслуживает в среднем за единицу времени  $\mu$  заявок; среднее число занятых каналов получится делением  $A$  на  $\mu$ :

$$\bar{z} = \frac{A}{\mu} = \frac{\lambda(1-p_n)}{\mu} \text{ или } \bar{z} = \rho(1-p_n).$$

#### 5.4. Одноканальная СМО с ограничением по длине очереди

Рассмотрим простейшую из всех возможных СМО с ожиданием - одноканальную систему ( $n = 1$ ), на которую поступает поток заявок с интенсивностью  $\lambda$ ; интенсивность обслуживания  $\mu$  (т. е. в среднем непрерывно занятый канал будет выдавать  $\mu$  обслуженных заявок в единицу времени). Заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания.

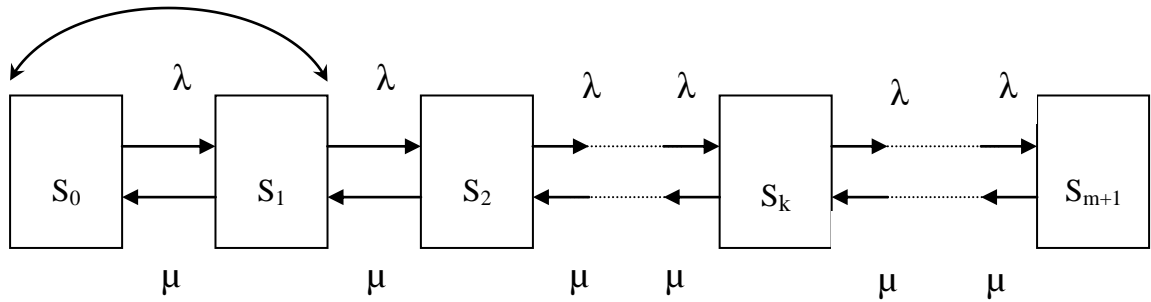


Рис. 5.4. Размеченный граф состояний одноканальной СМО с ожиданием

Предположим, что количество мест в очереди ограничено числом  $m$ , т. е. если заявка пришла в момент, когда в очереди уже стоят  $m$  заявок, она покидает систему необслуженной. В дальнейшем, устремив  $m$  к бесконечности, можно получить характеристики одноканальной СМО без ограничений по длине очереди.

Будем нумеровать состояния СМО по числу заявок, находящихся в системе (как обслуживаемых, так и ожидающих обслуживания):

- $S_0$  — канал свободен,
- $S_1$  — канал занят, очереди нет,
- $S_2$  — канал занят, одна заявка стоит в очереди,
- .....
- $S_k$  — канал занят,  $k - 1$  заявок стоит в очереди,
- .....
- $S_{m+1}$  — канал занят,  $m$  заявок стоят в очереди.

Граф состояний СМО показан на рис. 5.4. Интенсивности потоков событий, переводящих в систему по стрелкам слева направо, все равны  $\lambda$ , а справа налево —  $\mu$ . Действительно, по стрелкам слева направо систему переводит поток заявок (как только придет заявка, система переходит в следующее состояние), справа же налево — поток «освобождений» занятого канала, имеющий интенсивность  $\mu$  (как только будет обслужена очередная заявка, канал либо освободится, либо уменьшится число заявок в очереди).

Изображенная на рис. 5.4 схема представляет собой схему гибели и размножения. Пользуясь общим решением, данным для схемы гибели и размножения в 4.3.5, напишем выражения предельных вероятностей состояний:

$$\left. \begin{aligned} p_1 &= (\lambda/\mu)p_0, \\ p_2 &= (\lambda/\mu)^2 p_0, \\ \dots\dots\dots \\ p_k &= (\lambda/\mu)^k p_0, \\ \dots\dots\dots \\ p_{m+1} &= (\lambda/\mu)^{m+1} p_0, \\ p_0 &= \frac{1}{1 + (\lambda/\mu) + (\lambda/\mu)^2 + \dots + (\lambda/\mu)^{m+1}} \end{aligned} \right\} \cdot \quad (5.18)$$

Вводя обозначение  $\lambda/\mu = \rho$ , перепишем формулы (5.18) в виде:

$$\left. \begin{aligned} p_1 &= \rho p_0, \\ p_2 &= \rho^2 p_0, \\ \dots\dots\dots \\ p_k &= \rho^k p_0, \\ \dots\dots\dots \\ p_{m+1} &= \rho^{m+1} p_0, \\ p_0 &= \frac{1}{1 + \rho + \rho^2 + \dots + \rho^{m+1}} = [1 + \rho + \rho^2 + \dots + \rho^{m+1}]^{-1} \end{aligned} \right\} \cdot \quad (5.19)$$

Заметим, что в знаменателе последней формулы (5.19) стоит геометрическая прогрессия с первым членом 1 и знаменателем  $\rho$ ; суммируя эту прогрессию, находим:

$$p_0 = \frac{1}{(1 - \rho^{m+2})/(1 - \rho)} = \frac{1 - \rho}{1 - \rho^{m+2}} \cdot \quad (5.20)$$

Таким образом, формулы (5.19) окончательно примут вид:

$$\left. \begin{aligned}
 p_0 &= \frac{1-\rho}{1-\rho^{m+2}}, \\
 p_1 &= \rho p_0, \\
 p_2 &= \rho^2 p_0, \\
 \dots\dots\dots \\
 p_k &= \rho^k p_0, \\
 \dots\dots\dots \\
 p_{m+1} &= \rho^{m+1} p_0
 \end{aligned} \right\} \cdot \quad (5.21)$$

Обратим внимание на то, что формула (5.20) справедлива только при  $\rho \neq 1$  (при  $\rho = 1$  она дает неопределенность вида  $0/0$ ). Но сумму геометрической прогрессии со знаменателем  $\rho = 1$  найти еще проще, чем по формуле (5.20): она равна  $m + 2$ , и в этом случае  $p_0 = 1/(m + 2)$ . Заметим, что тот же результат можно получить более сложным способом, раскрывая неопределенность (5.20) по правилу Лопиталя.

Определим характеристики СМО: вероятность отказа  $P_{\text{отк}}$ , относительную пропускную способность  $q$ , абсолютную пропускную способность  $A$ , среднюю длину очереди  $\bar{r}$ , среднее число заявок, связанных с системой  $\bar{k}$ .

Очевидно, заявка получает отказ только в случае, когда канал занят и все  $m$  мест очереди — тоже:

$$P_{\text{отк}} = p_{m+1} = \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}}. \quad (5.22)$$

Находим относительную пропускную способность:

$$q = 1 - P_{\text{отк}} = 1 - \frac{\rho^{m+1}(1-\rho)}{1-\rho^{m+2}}. \quad (5.23)$$

Абсолютная пропускная способность:

$$A = \lambda q.$$

Найдем среднее число  $\bar{r}$  заявок, находящихся в очереди; определим эту величину как математическое ожидание дискретной случайной величины  $R$  — числа заявок, находящихся в очереди:  $\bar{r} = M[R]$ .

С вероятностью  $p_2$  в очереди стоит одна заявка, с вероятностью  $p_3$  две заявки, вообще с вероятностью  $p_k$  в очереди стоят  $k - 1$  заявок, наконец, с вероятностью  $p_{m+1}$  в очереди стоят  $m$  заявок. Среднее число заявок в очереди получим, умножая число заявок в очереди на соответствующую вероятность и складывая результаты:

$$\begin{aligned}\bar{r} &= 1 \cdot p_2 + 2 \cdot p_3 + \dots + (k-1) \cdot p_k + \dots + m \cdot p_{m+1} = \\ &= 1 \cdot \rho^2 p_0 + 2 \cdot \rho^3 p_0 + \dots + (k-1) \cdot \rho^k p_0 + \dots + m \cdot \rho^{m+1} p_0.\end{aligned}\quad (5.24)$$

Вынесем в этом выражении  $\rho^2 p_0$  за скобки:

$$\bar{r} = \rho^2 p_0 (1 + 2\rho + \dots + (k-1)\rho^{k-2} + \dots + m\rho^{m-1}).\quad (5.25)$$

Выведем формулу для суммы, стоящей в скобках. Очевидно, рассматриваемая сумма представляет собой не что иное, как производную по  $\rho$  суммы

$$\sum = \rho + \rho^2 + \dots + \rho^{k-1} + \dots + \rho^m,\quad (5.26)$$

а для этого выражения мы можем воспользоваться формулой суммы геометрической прогрессии:

$$\sum = \frac{\rho - \rho^{m+1}}{1 - \rho}.$$

Продифференцируем (5.26) по  $\rho$ :

$$\begin{aligned}\sum'_{\rho} &= \frac{[1 - (m+1)\rho^m](1 - \rho) + (\rho - \rho^{m+1})}{(1 - \rho)^2} = \\ &= \frac{1 - (m+1)\rho^m - \rho + (m+1)\rho^{m+1} + \rho - \rho^{m+1}}{(1 - \rho)^2} = \\ &= \frac{1 - (m+1)\rho^m + m\rho^{m+1}}{(1 - \rho)^2} = \frac{1 - \rho^m(m+1 - m\rho)}{(1 - \rho)^2}.\end{aligned}$$

Итак, выражение для суммы, стоящей в скобках в правой части (5.25), найдено:

$$1 + 2\rho + \dots + (k-1)\rho^{k-2} + \dots + m\rho^{m-1} = \frac{1 - \rho^m(m+1 - m\rho)}{(1 - \rho)^2}.\quad (5.27)$$

Подставляя его в (5.25), получим:

$$\bar{r} = \rho^2 p_0 \frac{1 - \rho^m (m + 1 - m\rho)}{(1 - \rho)^2}.$$

Учитывая выражение для  $p_0$  из (5.21), имеем:

$$\bar{r} = \rho^2 \frac{(1 - \rho)[1 - \rho^m (m + 1 - m\rho)]}{(1 - \rho^{m+2})(1 - \rho)^2}.$$

Или, окончательно,

$$\bar{r} = \frac{\rho^2 [1 - \rho^m (m + 1 - m\rho)]}{(1 - \rho^{m+2})(1 - \rho)}. \quad (5.28)$$

Таким образом, выведено выражение для среднего числа заявок, ожидающих обслуживания в очереди. Выведем теперь формулу для среднего числа  $\bar{k}$  заявок, связанных с системой, (как стоящих в очереди, так и находящихся под обслуживанием). Будем решать задачу следующим образом: рассмотрим общее число заявок  $K$ , связанных с системой, как сумму двух случайных величин: числа заявок, стоящих в очереди, и числа заявок, находящихся под обслуживанием:

$$K = R + \Omega,$$

где  $R$  – число заявок в очереди,  $\Omega$  – число заявок под обслуживанием.

По теореме сложения математических ожиданий

$$\bar{k} = M[K] = M[R] + M[\Omega] = \bar{r} + \bar{\omega},$$

где  $\bar{r}$  — среднее число заявок в очереди,  $\bar{\omega}$  — среднее число заявок под обслуживанием.

Величина  $\bar{r}$  выше была найдена; найдем величину  $\bar{\omega}$ . Так как канал один, то случайная величина  $\Omega$  может принимать только два значения: 0 или 1. Значение 0 она принимает, если канал свободен; вероятность этого равна

$$p_0 = \frac{1 - \rho}{1 - \rho^{m+2}}.$$

Значение 1 она принимает, если канал занят; вероятность этого равна

$$1 - p_0 = \frac{\rho - \rho^{m+2}}{1 - \rho^{m+2}}.$$

Отсюда находим математическое ожидание числа заявок, находящихся под обслуживанием:

$$\bar{\omega} = 0 \cdot p_0 + 1 \cdot (1 - p_0) = \frac{\rho - \rho^{m+2}}{1 - \rho^{m+2}}.$$

Таким образом, среднее число заявок, связанных с СМО, будет

$$\bar{k} = \bar{r} + \frac{\rho - \rho^{m+2}}{1 - \rho^{m+2}}, \quad (5.29)$$

где величина  $\bar{r}$  определяется по формуле (5.28).

Выведем выражение еще для одной существенной характеристики СМО с ожиданием: среднего времени ожидания заявки в очереди. Обозначим его  $\bar{t}_{ож}$ . Пусть заявка приходит в систему в какой-то момент времени. С вероятностью  $p_0$  канал обслуживания не будет занят, и ей не придется стоять в очереди (время ожидания равно нулю). С вероятностью  $p_1$  она придет в систему во время обслуживания какой-то заявки, но перед ней не будет очереди, и заявка будет ждать начала своего обслуживания в течение времени  $1/\mu$  (среднее время обслуживания одной заявки). С вероятностью  $p_2$  в очереди перед рассматриваемой заявкой будет стоять еще одна, и время ожидания в среднем будет равно  $2/\mu$  и т. д. Вообще, с вероятностью  $p_k$  пришедшая заявка застанет в системе  $k$  заявок и будет ждать в среднем  $k/\mu$  единиц времени; здесь  $k$  может быть любым целым числом до  $m$ . Что же касается  $k = m + 1$ , т. е. случая, когда вновь приходящая заявка застаёт канал обслуживания занятым и еще  $m$  заявок в очереди (вероятность этого  $p_{m+1}$ ), то время ожидания в этом случае также равно нулю, потому что заявка не становится в очередь (и не обслуживается). Поэтому среднее время ожидания будет:

$$\bar{t}_{ож} = p_1 \frac{1}{\mu} + p_2 \frac{2}{\mu} + \dots + p_k \frac{k}{\mu} + \dots + p_m \frac{m}{\mu}.$$

Подставляя сюда выражения для  $p_1, \dots, p_m$ , получаем:

$$\begin{aligned} \bar{t}_{ож} &= p_0 \rho \frac{1}{\mu} + p_0 \rho^2 \frac{2}{\mu} + \dots + p_0 \rho^k \frac{k}{\mu} + \dots + p_0 \rho^m \frac{m}{\mu} = \\ &= \frac{p_0 \rho}{\mu} (1 + 2\rho + \dots + k\rho^{k-1} + \dots + m\rho^{m-1}) . \end{aligned}$$

Преобразуем сумму в скобках, пользуясь формулой (5.27):

$$\bar{t}_{ож} = \frac{p_0 \rho}{\mu} \frac{1 - \rho^m (m + 1 - m\rho)}{(1 - \rho)^2} .$$

Или, выражая  $p_0$  через  $\rho$ :

$$\bar{t}_{ож} = \frac{1}{\mu} \frac{\rho(1 - \rho)}{1 - \rho^{m+2}} \frac{1 - \rho^m (m + 1 - m\rho)}{(1 - \rho)^2} = \frac{\rho [1 - \rho^m (m + 1 - m\rho)]}{\mu (1 - \rho^{m+2})(1 - \rho)} . \quad (5.30)$$

Сравнивая это выражение с формулой (5.28), замечаем, что

$$\bar{t}_{ож} = \frac{1}{\rho\mu} \bar{r} = \frac{\bar{r}}{\lambda} \quad (5.31)$$

т. е. среднее время ожидания равно среднему числу заявок в очереди, деленному на интенсивность потока заявок.

Выведем еще формулу для среднего времени пребывания заявки в системе. Обозначим  $T_{сист}$  случайную величину — время пребывания заявки в СМО. Эта случайная величина складывается из двух слагаемых (тоже случайных):

$$T_{сист} = T_{ож} + \theta ,$$

где  $T_{ож}$  — время ожидания заявки в очереди,  $\theta$  — случайная величина, равная времени обслуживания  $T_{об}$ , если заявка обслуживается, и нулю, если она не обслуживается (получает отказ). По теореме сложения математических ожиданий:

$$\bar{t}_{сист} = M[T_{сист}] = M[T_{ож}] + M[\theta] .$$

Но, в наших обозначениях,  $M[T_{ож}] = \bar{t}_{ож}$ , а  $M[\theta] = q\bar{t}_{обс} = q/\mu$ . Отсюда находим:  $\bar{t}_{сист} = \bar{t}_{ож} + q/\mu$ , или с учетом формулы (5.31),

$$\bar{t}_{сист} = \bar{r}/\lambda + q/\mu \quad (5.32)$$

Формула Литтла (первая): для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания среднее время пребывания заявки в системе равно среднему числу заявок в системе, деленному на интенсивность потока заявок.

Формула Литтла (вторая): для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания среднее время пребывания заявки в очереди равно среднему числу заявок в очереди, деленному на интенсивность потока заявок.

Получим вероятности состояний СМО с неограниченной очередью путем предельного перехода (при  $m \rightarrow \infty$ ) из формул (5.21).

Заметим, что при этом знаменатель в последней формуле (5.19) представляет собой сумму бесконечного числа членов геометрической прогрессии. Эта сумма сходится только, когда прогрессия бесконечно убывающая, т.е.  $\rho < 1$  есть условие, при котором в СМО с ожиданием существует предельный установившийся режим; при  $\rho > 1$  такого режима не существует, и очередь при  $t \rightarrow \infty$  растет до бесконечности.

Предположим, что  $\rho < 1$ , т. е. что предельный режим существует. Устремим в формулах (5.21)  $m \rightarrow \infty$  и выведем формулы для предельных вероятностей состояний в СМО без ограничений по длине очереди. Получим:

$$\left. \begin{aligned} p_0 &= 1 - \rho, \\ p_1 &= \rho(1 - \rho), \\ p_2 &= \rho^2(1 - \rho), \\ \dots\dots\dots \\ p_k &= \rho^k(1 - \rho), \\ \dots\dots\dots \end{aligned} \right\} \quad (5.33)$$

При отсутствии ограничений по длине очереди каждая заявка, пришедшая в систему, будет обслужена, поэтому  $q = 1$ ,  $A = \lambda q = \lambda$ . Среднее число заявок в очереди получим из (5.28) при  $m \rightarrow \infty$ :

$$\bar{r} = \frac{\rho^2}{1 - \rho} . \quad (5.34)$$

Среднее число заявок в системе по формуле (5.29) при  $m \rightarrow \infty$  будет равно

$$\bar{k} = \bar{r} + \rho = \frac{\rho^2}{1 - \rho} + \rho = \frac{\rho}{1 - \rho} . \quad (5.35)$$

Среднее время ожидания  $\bar{t}_{ож}$  также получим из формулы (5.31) при  $m \rightarrow \infty$ :

$$\bar{t}_{ож} = \frac{1}{\mu} \frac{\rho}{1 - \rho} . \quad (5.36)$$

Или, в другой форме:

$$\bar{t}_{ож} = \frac{\rho^2}{\lambda(1 - \rho)} . \quad (5.37)$$

Среднее время пребывания заявки в СМО равно среднему времени ожидания плюс среднее время обслуживания  $\bar{t}_{обсл} = \frac{1}{\mu}$ :

$$\bar{t}_{сист} = \frac{1}{\mu} \cdot \frac{\rho}{1 - \rho} + \frac{1}{\mu} = \frac{1}{\mu} \cdot \frac{1}{1 - \rho} . \quad (5.38)$$

## 5.5. Многоканальная СМО с ограничением по длине очереди

Рассмотрим  $n$ -канальную СМО с ожиданием, на которую поступает поток заявок с интенсивностью  $\lambda$ ; интенсивность обслуживания (для одного канала)  $\mu$ ; число мест в очереди  $m$ . Состояния системы будем нумеровать по числу заявок, связанных с системой:

$S_0$  – все каналы свободны,

$S_1$  – занят один канал, остальные свободны,  
 .....  
 $S_k$  – заняты  $k$  каналов, остальные свободны,  
 .....  
 $S_n$  – 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,  
 .....  
 $S_{n+m}$  – заняты все  $n$  каналов,  $m$  заявок стоят в очереди.

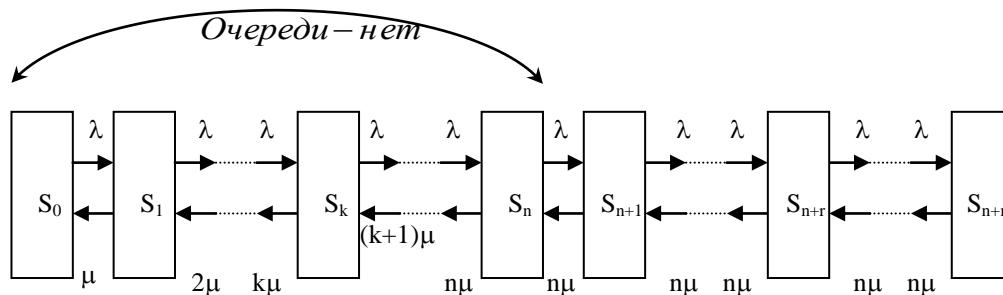


Рис. 5.5. Размеченный граф состояний многоканальной СМО с ожиданием

Граф состояний приведен на рис. 5.5. У каждой стрелки поставлены соответствующие интенсивности потоков событий. Действительно, по стрелкам слева направо систему переводит всегда один и тот же поток заявок с интенсивностью  $\lambda$ ; по стрелкам справа налево систему переводит поток обслуживаний, интенсивность которого равна  $\mu$ , умноженному на число занятых каналов.

Граф на рис. 5.5 представляет собой схему гибели и размножения, для которой решение в общем виде уже получено. Можно написать выражения для предельных вероятностей состояний.

Найдем некоторые характеристики эффективности обслуживания. Поступившая заявка получает отказ, если заняты все  $n$  каналов и все  $m$  мест в очереди:

$$P_{отк} = p_{n+m}.$$

Относительная пропускная способность, как всегда, дополняет вероятность отказа до единицы:

$$q = 1 - P_{отк}.$$

Абсолютная пропускная способность СМО будет равна:

$$A = \lambda q.$$

Найдем среднее число занятых каналов:

$$\bar{z} = \frac{A}{\mu}.$$

Среднее число заявок в очереди можно вычислить непосредственно, как математическое ожидание дискретной случайной величины, умножая любое возможное число заявок на вероятность того, что именно это число заявок будет в очереди, и складывая результаты:

$$\bar{r} = 1 \cdot p_{n+1} + 2 \cdot p_{n+2} + \dots + m \cdot p_{n+m}.$$

Складывая среднее число заявок в очереди  $\bar{r}$  и среднее число занятых каналов  $\bar{z}$ , получим среднее число заявок, связанных с системой:

$$\bar{k} = \bar{z} + \bar{r}.$$

Теперь найдем среднее время ожидания заявки в очереди:  $\bar{t}_{ож}$ . Для этого используем вторую формулу Литтла:

$$\bar{t}_{ож} = \frac{\bar{r}}{\lambda}.$$

Среднее время пребывания заявки в системе, также как и для одноканальной СМО, отличается от среднего времени ожидания на среднее время обслуживания, умноженное на относительную пропускную способность:

$$\bar{t}_{сист} = M[T_{ож}] + M[\Theta] = \bar{t}_{ож} + q / \mu.$$

Выше мы рассмотрели  $n$ -канальную СМО с ожиданием, когда в очереди одновременно могут находиться не более  $m$  заявок. Так же, как и в предыдущем разделе, можно получить выражения для неограниченной очереди, совершив предельный переход  $m \rightarrow \infty$ . Граф состояний в этом случае – бесконечный.

### 5.6. Многоканальная СМО с ограниченным временем ожидания

До сих пор мы рассматривали СМО с ожиданием, ограниченным только длиной очереди (числом  $m$  заявок, одновременно находящихся в очереди). В такой СМО заявка, ставшая в очередь, уже не покидает ее и «терпеливо» дожидается обслуживания. На практике нередко встречаются СМО другого типа, в которых заявка, подождав некоторое время, может уйти из очереди (так называемые «нетерпеливые заявки»).

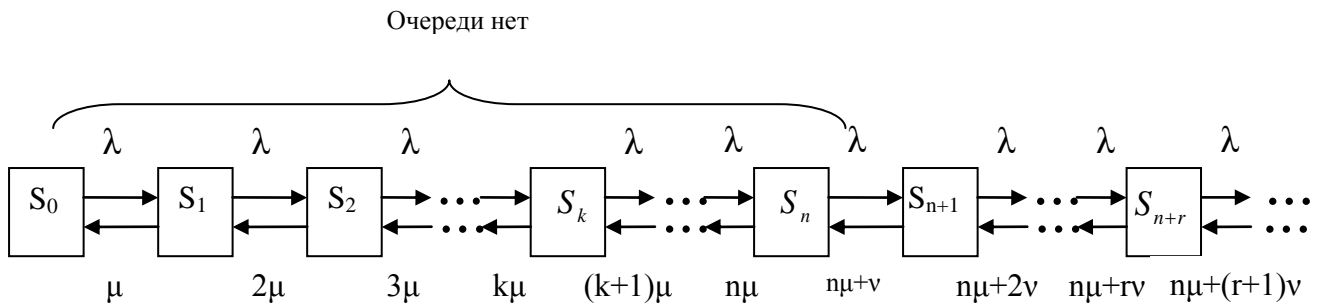


Рис. 5.6. Размеченный граф состояний многоканальной СМО с ограниченным временем ожидания

Рассмотрим СМО подобного типа, оставаясь в рамках марковской схемы. Предположим, что имеется  $n$ -канальная СМО с ожиданием, в которой число мест в очереди не ограничено, но время пребывания заявки в очереди ограничено некоторым случайным сроком  $T_{оч}$  со средним значением  $\overline{t_{оч}}$ , таким образом, на каждую заявку, стоящую в очереди, действует как бы «поток уходов» с интенсивностью

$$v = \frac{1}{t_{оч}}$$

Если этот поток пуассоновский, то процесс, протекающий в СМО, будет марковским. Найдем для него вероятности состояний. Будем снова нумеровать состояния системы по числу заявок, связанных с системой как обслуживаемых, так и стоящих в очереди:

$S_0$  – все каналы свободны,

$S_1$  – занят один канал,

$S_k$  – заняты два канала,

$S_n$  – 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,

и т.д. Граф состояний показан на рис 5.6.

Разметим этот граф, т.е. проставим у стрелок соответствующие интенсивности. Снова, как и раньше, у всех стрелок, ведущих слева направо, будет стоять интенсивность потока заявок  $\lambda$ . Для состояний без очереди у стрелок, ведущих из них справа налево, будет, как и раньше, стоять суммарная интенсивность потока обслуживаний всех занятых каналов. Что касается состояний с очередью, то у стрелок, ведущих из них справа налево, будет стоять суммарная интенсивность потока обслуживаний всех  $n$  каналов  $n\mu$ , плюс соответствующая интенсивность потока уходов из очереди. Если в очереди стоят  $r$  заявок, то суммарная интенсивность потока уходов будет равна  $rv$ .

Граф на рис. 5.6 представляет собой схему гибели и размножения, для которой решение в общем виде уже получено. Можно написать выражения для предельных вероятностей состояний.

Для СМО с «нетерпеливыми» заявками понятие «вероятность отказа» не имеет смысла — каждая заявка становится в очередь, но может и не дожидаться обслуживания, уйдя раньше времени.

Относительную пропускную способность  $q$  такой СМО можно подсчитать следующим образом. Очевидно, обслужены будут все заявки, кроме тех, которые уйдут из очереди досрочно. Подсчитаем, какое в среднем число заявок покидает очередь досрочно. Для этого вычислим среднее число заявок в очереди:  $\bar{r} = 1 \cdot p_{n+1} + 2 \cdot p_{n+2} + \dots + r \cdot p_{n+r} + \dots$ .

На каждую из этих заявок действует «поток уходов» с интенсивностью  $v$ . Значит, из среднего числа  $\bar{r}$  заявок в очереди в среднем будет уходить, не дождавсь обслуживания,  $v\bar{r}$  заявок в единицу времени; всего в единицу времени в среднем будет обслужено

$$A = \lambda - v\bar{r}$$

заявок. Относительная пропускная способность СМО будет

$$q = \frac{A}{\lambda} = \frac{\lambda - v\bar{r}}{\lambda} = 1 - \frac{v}{\lambda} \bar{r}.$$

Введем обозначение  $\frac{v}{\mu} = \beta$ . Среднее число занятых каналов  $\bar{z}$  по-прежнему получим, деля абсолютную пропускную способность на  $\mu$

$$\bar{z} = \frac{A}{\mu} = \frac{\lambda - v\bar{r}}{\mu} = \rho - \beta\bar{r}.$$

Складывая среднее число заявок в очереди  $\bar{r}$  и среднее число занятых каналов  $\bar{z}$ , получим среднее число заявок, связанных с системой:

$$\bar{k} = \bar{z} + \bar{r}.$$

Теперь найдем среднее время ожидания заявки в очереди:  $\bar{t}_{ож}$ . Для этого используем вторую формулу Литтла:

$$\bar{t}_{ож} = \frac{\bar{r}}{\lambda}.$$

Среднее время пребывания заявки в системе, также как и в предыдущем случае отличается от среднего времени ожидания на среднее время обслуживания, умноженное на относительную пропускную способность:

$$\bar{t}_{сист} = M[T_{ож}] + M[\Theta] = \bar{t}_{ож} + q / \mu.$$

Для нахождения среднего времени пребывания заявки в системе можно использовать также первую формулу Литтла.

### 5.7. Замкнутые системы массового обслуживания

До сих пор рассматривались такие системы массового обслуживания, где заявки приходили откуда-то извне и интенсивность потока заявок не зависела от состояния самой системы. В настоящем параграфе мы рассмотрим системы массового обслуживания другого типа — такие, в которых интенсивность потока поступающих заявок зависит от состояния самой СМО. Такие системы массового обслуживания называются замкнутыми.

В качестве примера замкнутой СМО рассмотрим следующую систему. Рабочий-наладчик обслуживает  $n$  станков. Каждый станок может в любой момент выйти из строя и потребовать обслуживания со стороны наладчика. Интенсивность потока неисправностей каждого станка равна  $\lambda$ . Вышедший из строя станок останавливается. Если в этот момент рабочий свободен, он берется за наладку станка; на это он тратит среднее время

$$\bar{t}_{об} = \frac{1}{\mu},$$

где  $\mu$  — интенсивность потока обслуживаний (наладок).

Если в момент выхода станка из строя рабочий занят, станок становится в очередь на обслуживание и ждет, пока рабочий не освободится.

Требуется найти вероятности состояний данной системы и ее характеристики:

вероятность того, что рабочий не будет занят,

вероятность наличия очереди,

среднее число станков, ожидающих очереди на ремонт и т. д.

Перед нами — своеобразная система массового обслуживания, где источниками заявок являются станки, имеющиеся в ограниченном количестве и подающие или не подающие заявки в зависимости от своего состояния: при выходе станка из строя он перестает быть источником новых заявок. Следовательно, интенсивность общего потока заявок, с которым приходится иметь дело рабочему, зависит от того, сколько имеется неисправных станков, т. е. сколько заявок связано с процессом обслуживания (непосредственно обслуживается или стоит в очереди).

Характерным для замкнутой системы массового обслуживания является наличие ограниченного числа источников заявок.

В сущности, любая СМО имеет дело только с ограниченным числом источников заявок, но в ряде случаев число этих источников так велико, что можно пренебречь влиянием состояния самой СМО на поток заявок. Например, поток вызовов на АТС крупного города исходит от ограниченного числа абонентов, но это число так велико, что практически можно считать интенсивность потока заявок независимой от состояний самой АТС (сколько каналов занято в данный момент). В замкнутой же системе массового обслуживания источники заявок, наряду с каналами обслуживания, рассматриваются как элементы СМО.

Рассмотрим сформулированную выше задачу о рабочем-наладчике в рамках общей схемы марковских процессов.

Система, включающая рабочего и  $n$  станков, имеет ряд состояний, которые мы будем нумеровать по числу неисправных станков (станков, связанных с обслуживанием):

$S_0$  – все станки исправны (рабочий свободен),

$S_1$  – один станок неисправен, рабочий занят его наладкой,

$S_2$  – два станка неисправны, один налаживается, другой ожидает очереди,

...

$S_n$  – все  $n$  станков неисправны, один налаживается,  $n-1$  стоят в очереди.

Граф состояний приведен на рис. 5.9. Интенсивности потоков событий, переводящих систему из состояния в состояние, проставлены у стрелок. Из состояния  $S_0$  в  $S_1$  систему переводит поток неисправностей всех работающих станков; его интенсивность равна  $n\lambda$ . Из состояния  $S_1$  в  $S_2$  систему переводит поток неисправностей уже не  $n$ , а  $n-1$  станков (работают всего  $n-1$ ) и т. д. Что касается интенсивностей потоков событий, переводящих систему по стрелкам справа налево, то они все одинаковы — работает все время один рабочий с интенсивностью обслуживания  $\mu$ .

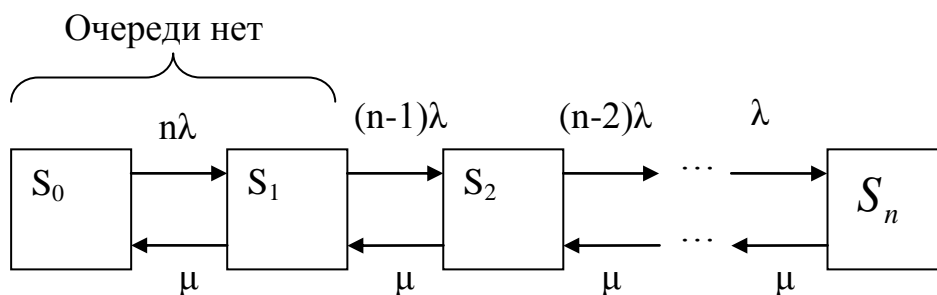


Рис. 5.9. Размеченный граф состояний замкнутой одноканальной СМО.

Пользуясь, как обычно, общим решением задачи о предельных вероятностях состояний для схемы гибели и размножения можно найти вероятности всех состояний СМО.

В силу своеобразия замкнутой СМО, характеристики ее эффективности будут отличны от тех, которые мы применяли ранее для СМО с неограниченным количеством источников заявок.

Роль «абсолютной пропускной способности» в данном случае будет играть среднее количество неисправностей, устраняемых рабочим в единицу времени. Вычислим эту характеристику. Рабочий занят наладкой станка с вероятностью

$$P_{зан} = 1 - p_0.$$

Если он занят, он обслуживает  $\mu$  станков (ликвидирует  $\mu$  неисправностей) в единицу времени; значит, абсолютная пропускная способность системы

$$A = (1 - p_0)\mu.$$

Относительную пропускную способность для замкнутой СМО мы не вычисляем, так как каждая заявка, в конце концов, будет обслужена:  $q = 1$ .

Вероятность того, что рабочий не будет занят:

$$P_{своб} = 1 - P_{зан} = p_0.$$

Вычислим среднее число неисправных станков, иначе — среднее число станков, связанных с процессом обслуживания. Обозначим это среднее число  $\bar{w}$ . Вообще говоря, величину  $\bar{w}$  можно вычислить непосредственно, по формуле

$$\bar{w} = 1 \cdot p_1 + 2 \cdot p_2 + \dots + n \cdot p_n,$$

но проще будет найти ее через абсолютную пропускную способность  $A$ . Действительно, каждый работающий станок порождает поток неисправностей с интенсивностью  $\lambda$ ; в нашей СМО в среднем работает  $n - \bar{w}$  станков; порождаемый ими средний поток неисправностей будет иметь

среднюю интенсивность  $(n - \bar{w})\lambda$ ; все эти неисправности устраняются рабочим, следовательно,

$$(n - \bar{w})\lambda = (1 - p_0)\mu ,$$

откуда

$$\bar{w} = n - \frac{\mu}{\lambda}(1 - p_0).$$

Определим теперь среднее число станков  $\bar{r}$ , ожидающих наладки в очереди. Учтем, что общее число станков  $W$ , связанных с обслуживанием, складывается из числа станков  $R$ , стоящих в очереди, плюс число станков  $\Omega$ , непосредственно находящихся под обслуживанием:

$$W = R + \Omega.$$

Число станков  $\Omega$ , находящихся под обслуживанием, равно единице, если рабочий занят, и нулю, если он свободен, т. е. среднее значение  $\Omega$  равно вероятности того, что рабочий занят:

$$\bar{w} = 1 - p_0.$$

Вычитая эту величину из среднего числа  $\bar{w}$  станков, связанных с обслуживанием (неисправных), получим среднее число станков, ожидающих обслуживания в очереди:

$$\bar{r} = n - \frac{1 - p_0}{\rho} - (1 - p_0) = n - (1 - p_0)\left(1 + \frac{1}{\rho}\right).$$

Рассмотрим теперь более общий пример замкнутой СМО: бригада из  $m$  рабочих обслуживает  $n$  станков ( $m < n$ ). Перечислим состояния системы:

- $S_0$  — все станки работают, рабочие не заняты,
- $S_1$  — один станок остановился, один рабочий занят,
- $S_2$  — два станка остановились, два рабочих заняты,
- .....
- $S_m$  —  $m$  станков остановились, все рабочие заняты,
- $S_{m+1}$  —  $m + 1$  станков остановились,  $m$  из них налаживаются, один ждет

очереди,

.....

$S_n$  – все  $n$  станков остановились,  $m$  из них налаживаются,  $n - m$  ждут в очереди.

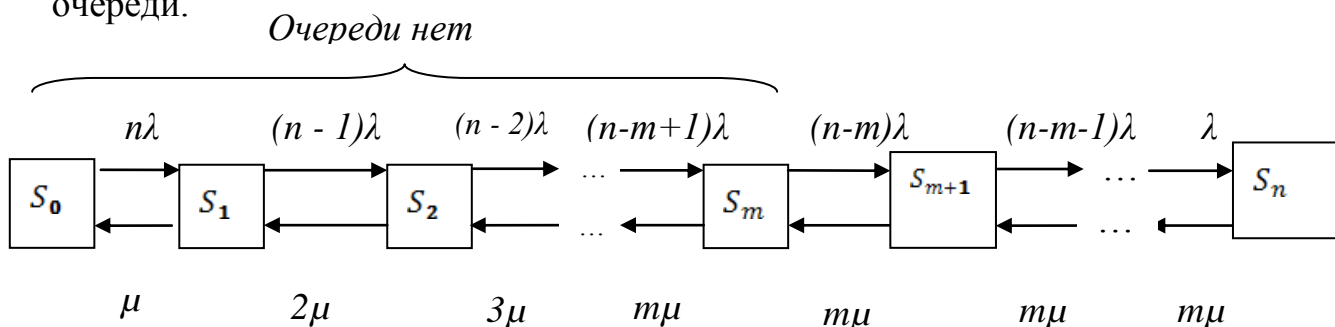


Рис. 5.10. Размеченный граф состояний замкнутой многоканальной СМО

Граф состояний системы показан на рис. 5.10. Применяя общее решение для схемы гибели и размножения, можно найти предельные вероятности всех состояний.

Через эти вероятности выражается среднее число  $\bar{z}$  занятых рабочих:

$$\begin{aligned} \bar{z} &= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \dots + m \cdot (p_m + p_{m+1} + \dots + p_n) = \\ &= p_1 + 2p_2 + \dots + (m-1)p_{m-1} + m(1 - p_0 - p_1 - \dots - p_{m-1}) \end{aligned}$$

Через  $\bar{z}$  выражается, в свою очередь, среднее число станков, обслуживаемых бригадой в единицу времени (абсолютная пропускная способность):  $A = \bar{z}\mu$ , а так же среднее число неисправных станков:

$$\bar{w} = n - \frac{\bar{z} \cdot \mu}{\lambda} = n - \frac{\bar{z}}{\rho} .$$

Аналогично вышеописанному можно рассмотреть другие задачи теории массового обслуживания, например, СМО с «взаимопомощью» между каналами (незанятые каналы «помогают» занятому в обслуживании), СМО с ошибками и другие.

### **5.8. Вопросы для самопроверки**

1. По каким признакам классифицируются системы массового обслуживания?
2. Какие показатели используются для описания эффективности функционирования систем массового обслуживания?
3. Постройте графы состояний для многоканальной СМО с отказами, многоканальной СМО с ограниченным временем ожидания в очереди и замкнутой СМО.
4. Какой режим работы СМО характеризуют предельные вероятности состояний и при каких условиях эти вероятности существуют?
5. Сформулируйте формулы Литтла для описания функционирования произвольной СМО.

## ГЛАВА 6. АГРЕГАТИВНОЕ ОПИСАНИЕ СИСТЕМ

### 6.1. Понятие «агрегат» в теории систем

Пусть  $T$  – фиксированное подмножество действительных чисел (множество рассматриваемых моментов времени),  $X, U, Y, Z$  – множества любой природы. Элементы указанных множеств назовем:  $t \in T$  – моментом времени;  $x \in X$ ; входным сигналом;  $u \in U$  – управляющим сигналом;  $y \in Y$  – выходным сигналом;  $z \in Z$  состоянием. Состояния, входные, управляющие и выходные сигналы, рассматриваемые как функции времени, обозначим  $z(t)$ ,  $x(t)$ ,  $u(t)$  и  $y(t)$ .

Под агрегатом будем понимать объект

$$\langle T, X, U, Y, Z, H, G \rangle, \quad (6.1)$$

где  $H, G$  – операторы (вообще говоря, случайные). Операторы переходов и выходов  $H$  и  $G$  реализуют функции  $z(t)$  и  $y(t)$ . Структура этих операторов собственно и выделяет агрегаты среди прочих систем. Рассмотрим подробнее вид этих операторов. При этом будем считать справедливым следующее предположение.

*Предположение 1.* За конечный интервал времени в агрегат поступает конечное число входных и управляющих сигналов и вырабатывается конечное число выходных сигналов.

*Операторы переходов.* Наряду с состоянием  $z(t)$  будем рассматривать также точки  $z(t+0)$ . Договоримся считать, что для любого  $t_1 > t$  момент  $(t+0) \in (t, t_1]$ . Вид оператора переходов  $H$  зависит от того, содержит ли рассматриваемый интервал времени моменты т.н. особых состояний агрегата или не содержит. Под особыми состояниями будем понимать его состояния в момент получения входного либо управляющего сигналов или выда-

чи выходного сигнала. Все остальные состояния агрегата будем называть неособыми.

*Предположение 2.* Из особых состояний агрегат может переходить в новое состояние скачком.

Пусть  $z(t^*)$  – некоторое особое состояние агрегата, а  $u_s$  – последний управляющий сигнал  $u_s \in U$ . Примем следующие обозначения для операторов, являющихся частными видами оператора  $H$  и определяющих состояние агрегата в момент  $t^*+0$ . Если  $t^*$  – момент поступления входного сигнала  $x$ , то

$$z(t^* + 0) = V'[z(t^*), x, u_s] . \quad (6.2)$$

Аналогично, если  $t^*$  – момент поступления управляющего сигнала  $u$ , то

$$z(t^* + 0) = V''[z(t^*), u] . \quad (6.3)$$

При одновременном поступлении  $x$  и  $u$

$$z(t^* + 0) = V[z(t^*), x, u] . \quad (6.4)$$

Наконец, если  $t^*$  – момент выдачи выходного сигнала  $y$ , то

$$z(t^* + 0) = W[z(t^*), u_s] . \quad (6.5)$$

В интервале между особыми состояниями, значение  $z(t)$  определяется при помощи операторов  $U$ , вид которых в общем случае зависит от особого состояния, являющегося для данного интервала времени начальным состоянием:

$$z(t) = U_{t_{oc}} [t, z(t_{oc} + 0), u_s] . \quad (6.6)$$

Здесь  $t_{oc}$  – момент особого состояния, являющегося исходным для данного интервала времени. Заметит, что все операторы  $U, V', V'', V$  и  $W$  в общем случае являются случайными.

*Оператор выходов.* Во множестве  $Z$  состояний  $z(t)$  агрегата выделим класс подмножеств  $\{Z_y\}$ , обладающих следующими свойствами. Выход-

ной сигнал  $y$  выдается в момент  $t'$  в тех случаях, когда: 1)  $z(t') \in Z_y$ ;  $z(t'-0) \notin Z_y$  и 2)  $z(t'+0) \in Z_y$ , но  $z(t') \notin Z_y$ . Тогда, оператор  $G$  можно представить в виде совокупности двух операторов:  $G'$ , вырабатывающего выходной сигнал

$$y = G'[z(t'), u_s] \quad (6.7)$$

и  $G''$ , проверяющего для каждого  $t$  принадлежность  $z(t)$  к одному из подмножеств  $Z_y$ . Заметим, что в общем случае, оператор  $G'$  является случайным оператором. Это значит, что данным  $t, z(t), u$  ставится в соответствие не одно определенное значение выходного сигнала, а некоторое множество значений  $y$  с соответствующим распределением вероятностей, задаваемых оператором  $G'$ .

В некоторых случаях в качестве одной составляющих  $z(t)$ , например  $z_1(t)$ , можно рассматривать время, оставшееся до выдачи выходного сигнала. Тогда оператор  $G''$  проверяет неравенство  $z_1(t) > 0$ .

## 6.2. Процесс функционирования агрегата

### 6.2.1. Функционирование агрегата общего вида

Агрегат функционирует следующим образом. В начальный момент времени  $t_0$  заданы начальное состояние агрегата  $z_0$  и начальное значение управляющего сигнала  $u_0$ .

Пусть  $t_1$  и  $t_2$  – моменты поступления первого  $x_1$  и второго  $x_2$  входных сигналов,  $\tau_1$  – момент поступления первого управляющего сигнала  $u_1$  и, для определенности  $t_1 < \tau_1 < t_2$ . Рассмотрим полуинтервал  $(t_0, t_1]$ . Состояния агрегата изменяются с течением времени по закону

$$z(t) = U_{t_0}[t, z_0, u_0] (t_0 < t \leq t_1) \quad (6.8)$$

до тех пор (оператор  $G''$ ), пока в момент  $t'$  (пусть  $t' < t_1$ ) состояние  $z(t')$  не окажется принадлежащим подмножеству  $Z'_y$ , хотя состояние  $z(t'-0)$  не принадлежало подмножеству  $Z'_y$ . В этом случае в момент  $t'$  выдается выходной сигнал  $y^{(1)}$ , вырабатываемый оператором  $G'$ . Вместе с тем закон изменения состояний (6.6) нарушается и

$$z(t'+0) = W[z(t'), u_0] . \quad (6.9)$$

Прежде чем рассматривать дальнейшие изменения состояний агрегата во времени, необходимо проверить (оператор  $G''$ ), не удовлетворяет ли состояние  $z(t'+0)$  условиям выдачи выходного сигнала, или, другими словами, не принадлежит ли (в смысле условий, упомянутых выше) состояние  $z(t'+0)$  некоторому новому подмножеству  $Z''_y$ . Если состояние  $z(t'+0)$  удовлетворяет условиям выдачи выходного сигнала (принадлежит подмножеству  $Z''_y$ ), то в момент  $t''$  выдается второй выходной сигнал  $y^{(2)}$  (оператор  $G'$ ), а состояние агрегата описывается соотношением

$$z(t'+0+0) = W[z(t'+0), u_0] = W\{W[z(t'), u_0], u_0\} \quad (6.10)$$

и т.д. В силу принятого соглашения в любой интервал времени может быть выдано лишь конечное множество выходных сигналов. Это свойство агрегата является ограничением, накладываемым на структуру подмножеств  $Z_y$  и оператор  $W$ . Предположим теперь, что  $z(t'+0)$  не принадлежит никакому из подмножеств  $Z_y$ . Поэтому далее состояние агрегата изменяется в соответствии с законом

$$z(t) = U_{t'}[t, z(t'+0), u_0] = U_{t'}\{t, W[z(t'), u_0], u_0\} . \quad (6.11)$$

Пусть теперь в момент  $t_1$  поступает входной сигнал  $x_1$ . Проследим поведение агрегата в момент  $t_1$  при различных вариантах возможных событий.

Если при достаточно малых  $\varepsilon > 0$  в момент  $t_1 - \varepsilon$  состояние агрегата не принадлежало подмножеству  $Z^*_y$ , а в момент  $t_1$   $z(t_1)$  принадлежит  $Z^*_y$ , то условимся, что в момент  $t_1$  выдается выходной сигнал  $y^*$ , а состояние агрегата есть

$$z(t_1 + 0) = W[z(t_1), u_0] . \quad (6.12)$$

Вместе с тем действие входного сигнала  $x_1$  приводит к тому, что

$$z(t_1 + 0 + 0) = V'[z(t_1 + 0), x_1, u_0] = V'\{W[z(t_1), u_0], x_1, u_1\} . \quad (6.13)$$

Очевидно, что состояние  $z(t_1 + 0 + 0)$  должно быть проверено (оператором  $G''$ ) по отношению к условиям выдачи выходного сигнала. Предположим теперь, что в момент  $t_1$  не было оснований для выдачи выходного сигнала  $y^*$ . Тогда вместо (6.12 и 6.13) в силу действия входного сигнала  $x_1$  состояние агрегата имеет вид

$$z(t_1 + 0) = V'[z(t_1), x_1, u_0] , \quad (6.14)$$

а в дальнейшем, если состояние (6.14) не соответствует выдаче выходного сигнала:

$$z(t) = U_{t_1}\{t, V'[z(t_1), x_1 u_0], u_0\} , \quad t_1 < t \leq \tau (t_1, \tau_1] . \quad (6.15)$$

Пусть в момент  $\tau_1$  в агрегат поступает управляющий сигнал  $u_1$ . Тогда состояние агрегата имеет вид

$$z(\tau_1 + 0) = V''[z(\tau_1), u_1] , \quad (6.16)$$

если в момент  $\tau_1$  не происходит выдача выходного сигнала, или

$$z(\tau_1 + 0 + 0) = V''\{[W[z(\tau_1), u_0]]u_1, \beta\} , \quad (6.17)$$

если в момент  $\tau_1$  выдается выходной сигнал.

Необходимо отметить, что управляющий сигнал  $u$  в общем случае является параметром, определяющим операторы  $V', V'', W, U, G', G''$ . Поэтому в дальнейшем вместо начального значения управляющего сигнала  $u_0$  в этих операторах должно использоваться значение  $u_1$  до тех пор, пока не посту-

пит следующий управляющий сигнал  $u_2$ . Например, в полуинтервале  $(\tau_1, t_2]$ , если нет оснований для выдачи выходного сигнала

$$z(t) = U_{\tau_1} \{t, z(\tau_1 + 0), u_1\}, \tau_1 < t \leq t_2. \quad (6.18)$$

В частном случае, операторы  $H$  и  $G$  могут оставаться неизменными при поступлении очередного управляющего сигнала. Аналогично, оператор  $U$  может быть одним и тем же при любых выходных сигналах (при попадании  $z(t)$  в любые подмножества  $Z_y$ ).

Агрегат представляет собой математическую схему весьма общего вида, частными случаями которой являются функции алгебры логики, релейно-контактные схемы, конечные автоматы, всевозможные классы систем массового обслуживания, динамические системы, описываемые обыкновенными дифференциальными уравнениями и некоторые другие объекты. С точки зрения моделирования агрегат выступает как достаточно универсальный переработчик информации – он воспринимает входные и управляющие сигналы и выдает выходные сигналы.

### 6.2.2. Функционирование СМО как агрегата

В качестве примера представления систем в виде агрегатов рассмотрим одноканальную СМО следующего вида. В моменты времени  $t_j$ , представляющие собой случайный поток однородных событий с заданным законом распределения, в систему поступают заявки, каждая из которых характеризуется параметром  $\alpha_j$ . Величина  $\alpha_j$  является случайной величиной с условным законом распределения  $f(\alpha/t)$  при условии, что  $t_j=t$ . Если линия обслуживания занята, заявка попадает в очередь и может находиться там (ждать) не более, чем

$$\tau_k^{(жс)} = \varphi(\alpha_j, \beta_i). \quad (6.19)$$

Величина  $\beta_i$  является случайной величиной с условным законом распределения  $f(\beta/t)$  при условии, что  $\tau_k=t$ . Величины  $\tau_k$  образуют случайный поток однородных событий с заданным законом распределения, не зависящим от  $t_j$ . Заявка должна быть принята к обслуживанию не позднее, чем в момент времени  $t_j + \tau_k^{(ж)}$ , в противном случае она получит отказ.

Время обслуживания заявки или время занятости линии

$$\tau_k^{(3)} = \psi(\alpha_j, \beta_i). \quad (6.20)$$

Представим эту СМО в виде агрегата. Состояние агрегата будем характеризовать вектором  $z(t) \in Z$  со следующими компонентами:

$z_1(t)$  – время, оставшееся до конца обслуживания заявки (для заявки, которая находится на обслуживании);

$z_2(t)$  – значение величины  $\beta_i$ ;

$z_3(t)$  – количество заявок в очереди;

$z_4(t)$  – оставшееся время ожидания первой заявки из очереди (время до момента отказа);

$z_5(t)$  – оставшееся время ожидания второй заявки из очереди (время до момента отказа);

$z_k(t)$  – оставшееся время ожидания последней заявки из очереди (время до момента отказа).

Входной сигнал  $x_j$  поступает в момент  $t_j$  и несет с собой информацию об  $\alpha_j$ , т.е.  $x_j = a_j$ . Управляющий сигнал  $u_i$  поступает в момент  $t_i$  и несет с собой информацию о величине  $\beta_{i+1}$ , т.е.  $u_i = \beta_{i+1}$ .

Рассмотрим подмножества  $Z_y$  и соответствующие им выходные сигналы  $y$ , операторы  $W$  и  $U$ .

Подмножество  $Z_y^{(1)}$ . Пусть в момент времени  $t_1$ :

$$z_1(t_1) = 0; \quad z_l(t) > 0; \quad l = 3, 4, \dots, k.$$





$$\begin{aligned}
z_3(t_j + 0) &= \tau_j^{(3)} = \psi(\alpha_j, \beta_i) \\
z_3(t_j + 0) &= 0 \\
&\dots\dots\dots
\end{aligned} \tag{6.27}$$

Величины  $z_l(t_j+0)$  для  $l>3$  не определяются. Если входной сигнал поступил после выходных сигналов  $y_1$  и  $y_2$ , то заявка попадает в очередь, т.е.

$$\left. \begin{aligned}
z_1(t_j + 0) &= z_1(t_j) \\
z_2(t_j + 0) &= z_2(t_j) \\
z_3(t_j + 0) &= z_3(t_j) + 1 \\
&\dots\dots\dots \\
z_l(t_j + 0) &= z_l(t_j); l = 4, 5, \dots, k \\
&\dots\dots\dots \\
z_{k+1}(t_j + 0) &= \tau_l^{(k)} = \varphi(\alpha_j, \beta_i)
\end{aligned} \right\} \tag{6.28}$$

В дальнейшем (оператор  $U_{(t_j)}$ ) состояния  $z(t)$  определяются аналогично (6.24).

Пусть теперь в момент  $\tau_i$  поступает управляющий сигнал  $u_i = \beta_{i+1}$ . При этом изменится только значение  $z_2(t)$ : вместо прежнего значения  $\beta_i$  должно быть  $z_2(t) = \beta_{i+1}$ . остальные  $z_l(t)$  не зависят от  $\beta_{i+1}$ . Из этого легко усмотреть содержание операторов  $W$  и  $U$ .

На этом рассмотрение примера можно закончить. Обзор процесса функционирования такой системы массового обслуживания как агрегата дальше можно сделать самостоятельно.

**6.3. Алгоритм процесса функционирования агрегата общего вида**

Целью моделирования в любом случае является получение характеристик, определяемых состояниями системы. Для этого необходимо фиксировать в процессе моделирования достаточно полную информацию о состояниях системы в соответствующие моменты модельного времени. При-

менительно к агрегату это означает, что необходимо получать значения состояний  $z(t)$  для некоторых моментов времени интервала исследования  $(0, T)$ . Выше было показано, что вид оператора  $H$ , решающего эту задачу, зависит от того, поступают или нет входные и управляющие сигналы в течение рассматриваемого интервала времени. Моменты поступления  $x$  и  $u$  играют значительную роль с точки зрения построения моделирующего алгоритма. В частности, мы будем рассматривать моделирование как последовательную цепь переходов из одного особого состояния агрегата в другое, причисляя условно к особым состояниям и  $z(0)$ . Вид моделирующего алгоритма существенно зависит от того, известны ли заранее моменты поступления  $x$  и  $u$  и вообще моменты последующих особых состояний. С этой точки зрения следует иметь в виду два случая. Первый связан с рассмотрением агрегата, для которого законы поступления входных и управляющих сигналов заданы. В этом случае, очевидно, для моделирования агрегата, кроме описания самого агрегата, необходимо иметь исчерпывающее описание входных и управляющих сигналов, как воздействий внешней среды. Второй случай - рассмотрение агрегата, для которого входные и управляющие сигналы вырабатываются в процессе моделирования как выходные сигналы других агрегатов. Рассмотрим более подробно первый случай.

Итак, изучаем процесс функционирования агрегата в интервале времени  $(0, T)$ . Совокупности входных и управляющих сигналов, поступающих в агрегат в течение интервала времени  $(0, T)$ , считаем известными. Весьма важным является вопрос о способах ввода в модель входных и управляющих сигналов. Возможные варианты:

1. Входные и управляющие сигналы записываются в памяти машины в виде таблицы, содержащей для всех моментов значения параметров сигнала-

лов. Данные из таблицы автоматически выбираются по ходу моделирования (если не очень много сигналов и их параметров).

2. Генерирование сигналов при помощи случайных чисел. Моменты поступления сигналов в агрегат представляются в виде потока однородных событий, задаваемого соответствующим законом распределения интервалов времени между последовательными моментами. Другие параметры сигнала следует при этом задавать условным законом распределения при условии, что момент поступления сигнала задан. Тогда моделирование совокупности сигналов сводится к моделированию потока однородных событий и реализаций случайного вектора, заданного условным законом распределения.

Существенные особенности возникают лишь в том случае, когда входной (или управляющий) сигнал приходит от нескольких источников. Чтобы учесть такого рода обстоятельства, необходимо к подалгоритмам модели, формирующей входные и управляющие сигналы, построить дополнительный подалгоритм – сортировщик, который будет располагать сигналы, вырабатываемые различными источниками, в единую последовательность по времени.

Блок-схема моделирующего алгоритма приведена на рис. 6.1 – 6.4. В состав операторов схемы входят вычислительные операторы, операторы формирования реализаций случайных процессов, операторы формирования неслучайных величин и логические операторы.

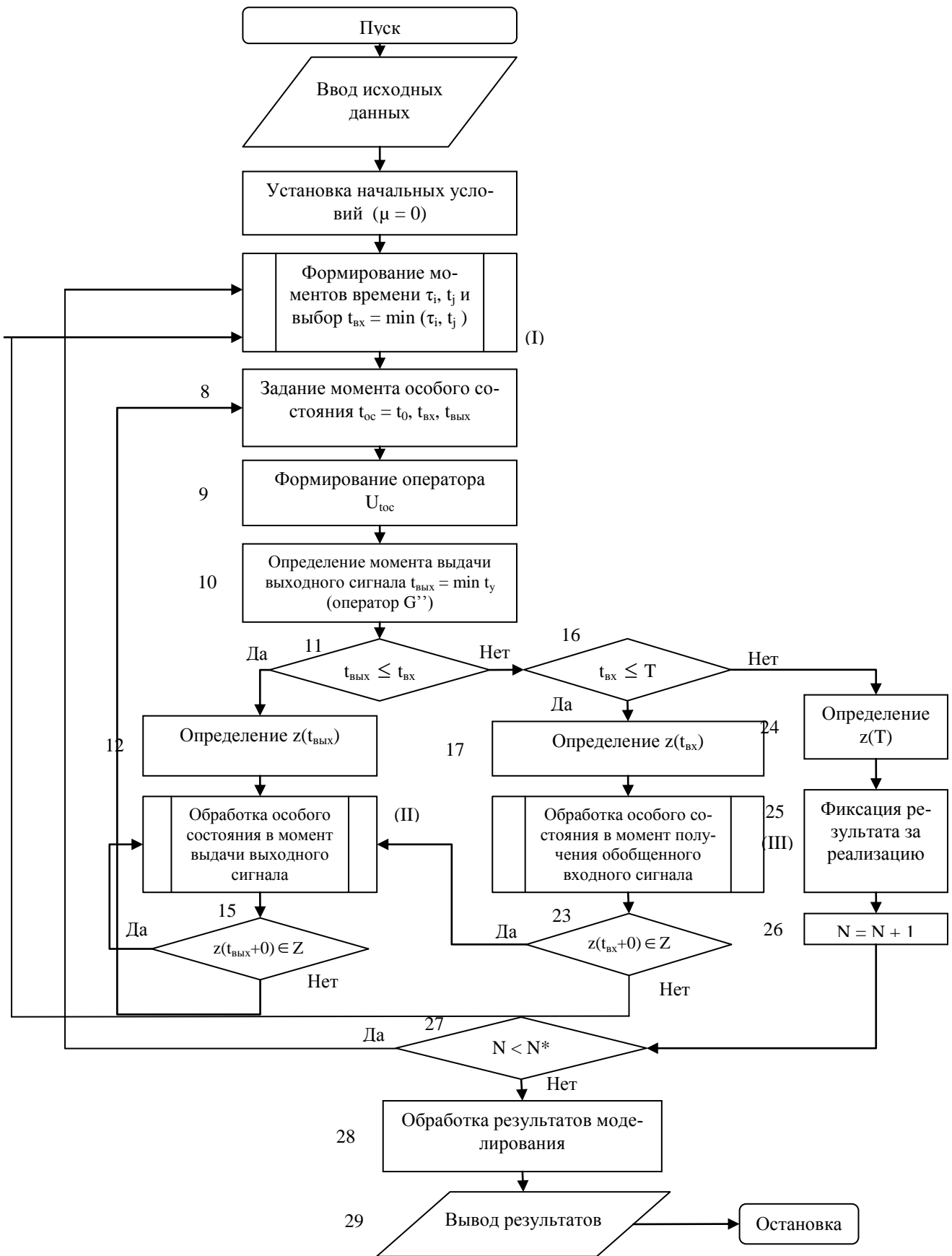


Рис. 6.1. Схема алгоритма функционирования агрегата общего вида

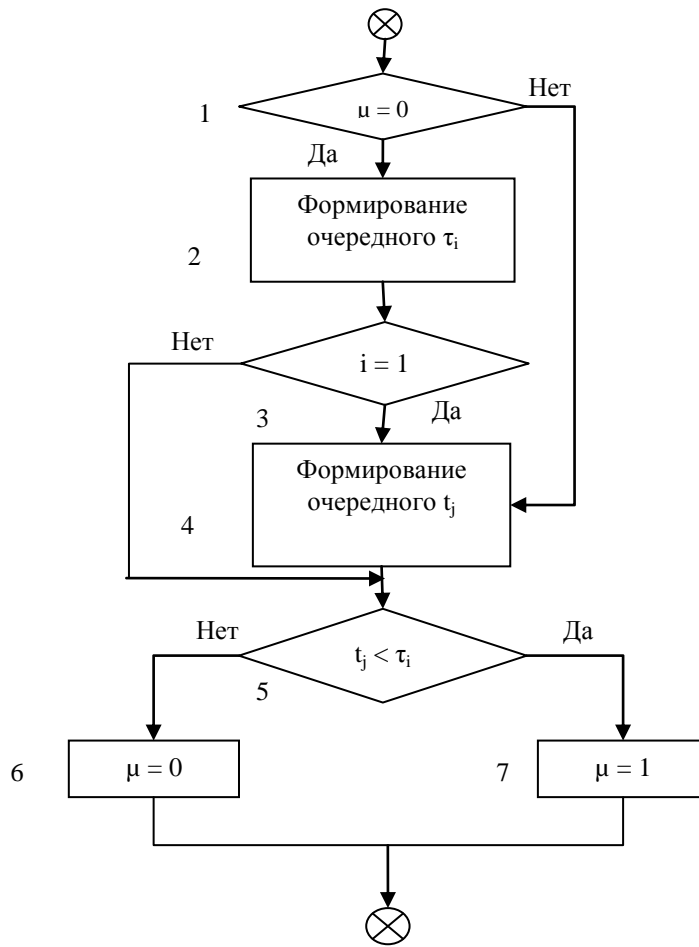


Рис 6.2 Формирование моментов времени прихода управляющего  $\tau_i$  и входного  $t_j$  сигналов, выбор ближайшего из них (сортировка).

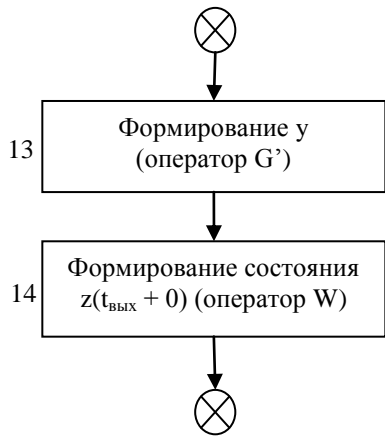


Рис 6.3 Обработка особого состояния в момент выдачи выходного сигнала

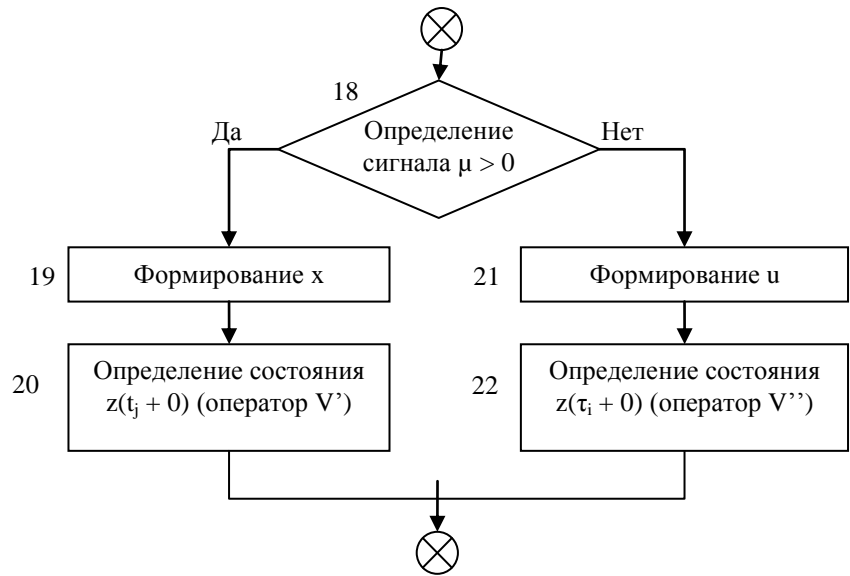


Рис 6.4 Обработка особого состояния в момент получения обобщенного входного сигнала

Работа моделирующего алгоритма протекает следующим образом. Группа операторов 1-7 (рис. 6.2) предназначена для формирования моментов поступления входных и управляющих сигналов и определения момента поступления ближайшего сигнала. Операторы 2 и 4 вырабатывают моменты поступления в агрегат управляющего сигнала  $\tau_i$  и входного сигнала  $t_j$  соответственно. Оператор 5 сравнивает  $\tau_i$  и  $t_j$  и определяет ближайший момент  $t_{\text{вх}}$  поступления в агрегат внешнего сигнала. Результаты сравнения фиксируются операторами 6 и 7; если  $\mu = 0$  – ближайшим будет управляющий сигнал, если  $\mu = 1$  – входной сигнал. Группа операторов 9-15 (рис. 6.1) моделирует процесс функционирования агрегата в интервале времени между последовательными моментами поступления внешнего сигнала. В этом интервале состояния агрегата определяются оператором  $U_{10}$  (оператор 9). Единственным видом особых состояний агрегата в этом случае являются состояния выдачи выходных сигналов или, другими словами, состояния, принадлежащие подмножествам  $Z_y$ . Задача оператора 10 состоит в

том, чтобы путем совместного моделирования состояний  $z(t)$  и условий, определяющих подмножества  $Z_y$ , определить моменты  $t_y$  выдачи выходных сигналов (оператор  $G''$  агрегата) и найти наименьший из них. Пути решения этой задачи (реализация оператора 10) зависят от свойств агрегата и способа задания оператора  $G''$ . В наиболее простых случаях задача сводится к проверке неравенств вида  $t_{\text{вых}} - t > 0$  или к совместному решению уравнений (неравенств), описывающих состояния  $z(t)$  агрегата и подмножества  $Z_y$ .

Могут быть и более сложные подходы. В худшем случае приходится «прощупывать» состояния агрегата через малые интервалы времени  $\Delta t$  и проверять принадлежность их подмножествам  $Z_y$  на каждом шаге. Выбор минимального  $t_y$ , т.е.  $t_{\text{вых}} = \min t_y$ , обычно не представляет собой проблемы.

Если момент  $t_{\text{вых}}$  оказывается внутри интервала между внешними сигналами (оператор 11), то выходной сигнал  $u$  должен быть фактически выдан. Поэтому необходимо найти  $z(t_{\text{вых}})$  (оператор 12) и сформировать сам сигнал  $u$  (оператор 13 (рис. 6.3)). Обычно при моделировании оператора  $G'$  его расчленяют на две части:

- формирование для заданного  $Z_y$  закона распределения случайного выходного сигнала  $u$ ;
- выбор самого сигнала по жребию в соответствии с этим законом распределения.

Затем, поскольку агрегат находится в особом состоянии  $z(t_{\text{вых}})$ , необходимо определить его состояние  $z(t_{\text{вых}} + 0)$  (оператор 14) и проверить (оператор 15), не принадлежит ли  $z(t_{\text{вых}} + 0)$  одному из подмножеств  $Z_y$ . Если это условие выполнено, то управление передается опять оператору 13 для формирования второго выходного сигнала, выдаваемого в момент  $t_{\text{вых}}$  и т.д. Если же  $z(t_{\text{вых}} + 0) \notin Z_y$ , то через оператор 8 управление передается операторам 9 и 10 для формирования новых моментов  $t_{\text{вых}}$ . Возвратимся к

оператору 11. Если условие, проверяемое оператором 11, не выполнено, т.е.  $t_{\text{вх}} \leq t_{\text{вых}}$ , то переходим к рассмотрению поведения агрегата под воздействием внешнего сигнала. Будем считать, что  $t_{\text{вх}} < T$  (оператор 16). В противном случае наступает конец интервала моделирования и мы переходим к определению состояния  $z(T)$  (оператор 24). Оператор 17 вычисляет состояние агрегата  $z(t_{\text{вх}})$ , а оператор 18 (рис. 6.4) определяет тип внешнего сигнала (входной, управляющий). Операторы 19,20 моделируют прием входного сигнала в момент  $t_j$ , а операторы 21,22 - прием управляющего сигнала в момент  $\tau_i$ . Работа этих операторов аналогична работе рассмотренной выше группы операторов 13,14. Заметим только, что  $z(t_{\text{вх}} + 0)$  сразу может принадлежать одному из подмножеств  $Z_y$ . Это условие проверяется оператором 23. Если условие справедливо, то управление передается оператору 13 для формирования выходного сигнала. Если нет, управление передается оператору 1 для формирования очередного момента поступления управляющего (входного) сигнала.

Сложность компьютерной реализации рассмотренного моделирующего алгоритма зависит, в основном, от сложности реализации операторов, определяющих функционирование агрегата.

Работа некоторых операторов будет подробнее рассмотрена в дальнейшем в разделе «имитационное моделирование».

#### **6.4. Кусочно-линейные агрегаты**

Как показывает анализ моделирующих алгоритмов, можно добиться их существенных упрощений, если рассматривать объекты чуть более частные, чем агрегат общего вида, но сохраняющие такую важную его особенность, как возможность описания достаточно широкого класса реальных систем. Практически удобным для формализации чрезвычайно широкой

совокупности разнообразных процессов и явлений материального мира являются так называемые кусочно-линейные агрегаты (КЛА), к описанию которых мы и приступим.

#### **6.4.1. Понятие о кусочно-линейном агрегате.**

Универсальная модель элемента была предложена Н.П. Бусленко и И.Н. Коваленко в форме кусочно-линейного агрегата. Координаты (признаки) агрегата делятся на целочисленные и вещественные. Целочисленные переменные между событиями остаются неизменными, а вещественные меняются линейно до тех пор, пока хотя бы одна из них не станет равной нулю. Этот момент и считается наступлением события. В такой момент сигналы с «выходных клемм» сработавшего агрегата мгновенно передаются на входные клеммы остальных агрегатов, изменяя координаты последних в соответствии со схемой взаимодействия. Подобными формализмами описывается весьма широкий класс систем: производственные, массового обслуживания, управления запасами, передачи данных, дорожного движения и т.д. Теория агрегативных систем интенсивно развивается. На ее базе во ВНИИ системных исследований была создана агрегативная имитационная система (АИС).

Для поставленных здесь задач достаточно считать, что на агрегат не поступают управляющие сигналы  $u$ , а поступают лишь входные сигналы  $x$  (не ограничивает общности, в качестве  $x$  можно рассматривать входной сигнал в широком смысле ( $\bar{x}$ )). Итак, мы рассматриваем агрегат как объект, который в каждый момент времени  $t$  характеризуется внутренним состоянием  $z(t)$ , имеет вход и выход. На вход агрегат в изолированные моменты времени могут поступать сигналы, с выхода могут сниматься выходные сигналы. Класс кусочно-линейных агрегатов выделяется с помощью конкретизации структуры множеств  $Z, X, Y$  а также операторов  $H$  и  $G$ . Опишем данную конкретизацию.

Рассмотрим некоторое конечное или счетное множество  $I$ . Для определенности предположим, что  $I = \{0, 1, 2, \dots\}$ , хотя в конкретных задачах  $I$  мо-

жет иметь и другой вид. Назовем  $I$ -множеством основных состояний, а элементы  $v \in I$  – основными состояниями. Каждому основному состоянию  $v \in I$  поставим в соответствие некоторое целое неотрицательное число  $\|v\|$ , которое назовем рангом основного состояния. Кроме того, каждому  $v \in I$  поставим в соответствие выпуклый многогранник  $Z^{(v)}$  в евклидовом пространстве размерности  $\|v\|$ . Будем считать, что  $Z = \cup Z^{(v)}$ , т.е. пространство состояний  $Z$  можно представить состоящим из всевозможных пар вида  $(v, z^{(v)})$ , где  $v \in I$ , а  $z^{(v)}$  является вектором размерности  $\|v\|$  и принимает значения из многогранника  $Z^{(v)}$ . Вектор  $z^{(v)}$  будем называть вектором дополнительных координат. Если  $\|v\|=0$  для некоторого  $v \in I$ , то это означает, что в данном состоянии  $v$  дополнительные координаты не определяются.

#### 6.4.2. Процесс функционирования КЛА.

Опишем сначала динамику КЛА, т.е. процесс изменения внутренних состояний во времени, в предположении отсутствия поступления  $x$ . В предыдущей терминологии, определим действия оператора  $U$ . Пусть в начальный момент времени  $t_0$  агрегат находится в состоянии  $z(t_0) = (v, z^{(v)}(0))$ , где  $z^{(v)}(0)$  – внутренняя точка многогранника  $Z^{(v)}$ . Тогда при  $t > t_0$  точка  $z^{(v)}(t)$  перемещается внутри многогранника  $Z^{(v)}$  до тех пор, пока не достигнет его границы. Пусть это произойдет в момент  $t_1$ , который назовем «опорным». Тогда при  $t_0 < t \leq t_1$ ,  $\Delta t = t - t_0$  «движение» агрегата описывается следующими законами:

$$v(t) = v = \text{const}, \quad (6.29)$$

данному значению  $v$  соответствует вектор  $\alpha^{(v)}$  размерности  $\|v\|$  и

$$z^{(v)}(t) = z^{(v)}(0) + \Delta t * \alpha^{(v)}. \quad (6.30)$$

Значение опорного момента  $t_1$  определяется траекторией  $z(t)$ , вернее её некоторыми параметрами и может быть найдено из соотношения

$$t_1 = \inf\{t : z^{(v)}(0) + (t - t_0)\alpha^{(v)} \notin Z^{(v)}, t > t_0\}, \quad (6.31)$$

поскольку  $Z^{(v)}$  многогранник, то нахождение  $t_1$  по (6.31) сводится к следующему. Пусть  $Z_j^{(v)}$  -  $j$ -тая грань многогранника  $Z^{(v)}$  (предположим, что  $Z^{(v)}$  содержит  $m(v)$  граней). Эти грани могут быть заданы линейными уравнениями:

$$\sum_{i=1}^{\|v\|} d_{ji}^{(v)} z_i^{(v)} + d_{j0}^{(v)} = 0 \quad j=1, \dots, m(v), \quad (6.32)$$

где  $z_i^{(v)}$  - компоненты вектора  $z^{(v)}$ ,  $i=1.. \|v\|$ . Легко понять, что (6.32) может быть записано в виде

$$t_1 = \min_j \{t : z^{(v)}(0) + (t - t_0)\alpha^{(v)} \in \bigcup_{j=1}^{m(v)} Z_j^{(v)}, t > t_0\} \text{ или}$$

$$t_1 = \min_j \{t : t > t_0, \sum_{i=1}^{\|v\|} d_{ji}^{(v)} [z_i^{(v)}(0) + (t - t_0)\alpha_i^{(v)}] + d_{j0}^{(v)} = 0\}. \quad (6.33)$$

$$\text{Обозначим } \tau_j = -\frac{d_{j0}^{(v)} + \sum_{i=1}^{\|v\|} d_{ji}^{(v)} z_i^{(v)}(0)}{\sum_{i=1}^{\|v\|} d_{ji}^{(v)} \alpha_i^{(v)}}, \quad j=1, \dots, m(v). \quad (6.34)$$

$$\text{Пусть } \tau = \min\{\tau_j; \tau_j > 0\}. \quad (6.35)$$

$$\text{Тогда из (6.33-6.35) следует, что } t_1 = t_0 + \tau. \quad (6.36)$$

В момент  $t_1$  состояние рассматриваемого кусочно-линейного агрегата изменяется скачкообразно. Значение  $z(t_1+0)$  является случайным, задаваемым распределением  $P_1$ , которое зависит лишь от состояния  $z(t_1)$ . В момент  $t_1$  может выдаваться выходной сигнал (см. оператор  $G$ ). Содержание (и необходимость выдачи)  $y$  зависит от состояния  $z(t_1)$ . Подмножество  $Z_y$ , введенное в общем определении агрегата, в данном случае совпадает с

$\bigcup_{v=0}^{\infty} \bigcup_{j=1}^{m(v)} Z_j^{(v)}$ . Для нас важно, указать, что множество  $Y$  имеет структуру, ана-

логичную  $Z$ , т.е. выходные сигналы  $y$  представляются  $y=(\lambda, y(\lambda))$ , где  $\lambda$ -элемент некоторого не более чем счетного множества,  $y(\lambda)$  – вектор, принимающий значения из евклидова пространства размером, зависящим от  $\lambda$ . При  $t > t_1$  движение агрегата вновь происходит в соответствии с формулами (6.29) и (6.30) до очередного «особого» момента  $t_2$ , где под  $\Delta t$  нужно понимать теперь  $t-t_1$  и т.д. Обратимся теперь к случаю поступления входного сигнала. Подчеркнем, что для КЛА множество  $X$  структурно аналогично множествам  $Z$  и  $Y$ , т.е.  $x=(\mu, x^{(\mu)})$ , где  $\mu$ -элемент конечного или счетного множества, а  $x^{(\mu)}$ - действительный вектор, размерность которого зависит от  $\mu$ . Следующее описание поведения КЛА можно рассматривать как раскрытие действия оператора  $V$ .

Пусть в рассматриваемый момент  $t$  состояние агрегата  $z(t)=(v, z^{(v)})$  и пусть в этот момент поступает входной сигнал  $x=(\mu, x^{(\mu)})$ . При этом состояние агрегата меняется скачкообразно. Значение  $z(t+0)$  является случайным, задаваемым распределением  $P_2$ , которое, вообще говоря, зависит от  $z(t)$  и  $x$ . Будем считать, что в рассматриваемый момент может выдаваться выходной сигнал, содержание и необходимость выдачи которого зависит не только от состояния  $z(t)$  (и, быть может,  $z(t+0)$ ), но и от содержания поступившего входного сигнала  $x$ . После рассматриваемого момента времени  $t$  движение агрегата происходит в соответствии с формулами (6.29) и (6.30) до следующего момента поступления входного сигнала или выхода вектора состояния на границу допустимых значений.

Динамика КЛА описана полностью. В виде КЛА могут быть формализованы многие реальные процессы: процессы передачи и обмена данными в сетях связи, системы массового обслуживания и материально-технического снабжения, процессы автомобильного движения на дорогах, разнообразные дискретные производственные процессы, вычислительные системы и т.д. При этом всюду основные состояния агрегата указывают на

качественно различные состояния моделируемых объектов. Дополнительные же координаты характеризуют происходящие количественные изменения и часто носят сугубо вспомогательный характер, «вбирая» в себя необходимую информацию о предыстории модели. Следует отметить, что представление реальных систем в форме КЛА неоднозначно, поскольку неоднозначно могут быть выбраны состояния агрегатов. Выбор же состояний определяется как целями исследования, так и стремлением уменьшить размерность задачи. При этом всегда приходится идти на компромисс между точностью описания и полнотой получаемой информации с одной стороны и простотой модели – с другой.

### 6.4.3. Примеры представления систем в виде КЛА.

Пример 1. Вероятностный автомат. Рассмотрим конечный асинхронный вероятностный автомат Мура, т.е. такой, который не имеет «жесткой» тактности, а изменяет свое состояние всякий раз, когда поступает входной сигнал. Синхронный автомат получается, если потребовать, чтобы входные сигналы поступали в моменты  $t=1,2,\dots$ . Пусть  $Z_a$  – конечное множество внутренних состояний автомата,  $X_a$  – его входной (конечный) алфавит,  $Y_a$  – его выходной (конечный) алфавит. Для определенности будем считать, что

$$Z_a = \{1,2,\dots,N\}, X_a = \{1,2,\dots,K\}, Y_a = \{1,2,\dots,M\}.$$

Динамика автомата описывается следующим образом. Если в момент  $t$  состояние автомата  $z_a(t)=i$  и поступил входной сигнал,  $x_a(t)=k$ , то состояние  $z_a(t+0)=j$  выбирается случайно с вероятностью  $p_{ij}^{(k)}$ ,  $p_{ij}^{(k)} \geq 0$ ,  $\sum_{j=1}^N p_{ij}^{(k)} = 1$  при любом  $k$ ,  $1 \leq k \leq K$ . Выходной сигнал  $y_a \in Y_a$ , выдаваемый в этот момент, является однозначной функцией «нового» состояния  $j$ :  $y_a = m = \Phi(j)$ , где  $\Phi$  – некоторая детерминированная функция с областью определения  $Z_a$  и множе-

ством значений  $Y_a$ . Представим этот автомат в виде кусочно-линейного агрегата. В качестве множества входных сигналов  $X$  КЛА возьмем  $X_a$ , а в качестве выходных сигналов  $Y$  – множество  $Y_a$ . За множество  $I$  основных состояний КЛА выберем  $Z_a$  и будем считать  $\|v\|=0$  при всех  $v \in I$ , т.е. вектор дополнительных координат  $z(v)$  не определяется. При таком выборе состояния КЛА не определяются многогранники  $Z(v)$ , отпадают вопросы о движении внутри многогранника, выходе на границу и связанного с ним распределения  $P_1$ .

Все движение рассматриваемого КЛА состоит из скачков состояния при поступлении входных сигналов, причем ввиду отсутствия вектора дополнительных координат речь идет лишь о скачках основного состояния  $v$ . Т.о., требуется задать лишь распределение  $P_2$ . Оно совпадает с рассмотренным выше распределением  $p_{ij}^{(k)}$ . Содержание же выходного сигнала, выдаваемого в момент поступления входного, определяется только функцией  $\Phi$ . Вообще, если предположить, что  $\|v\|=0, \|\lambda\|=0, \|\mu\|=0 \forall v, \lambda, \mu$ , то легко видеть, что КЛА превращается в вероятностный автомат весьма общего вида.

Пример 2. Система массового обслуживания. Пусть на обслуживающий прибор поступает ординарный поток требований, причем  $i$ -е по счету требование характеризуется параметром  $\theta_i$ , который представляет собой предельно допустимое время ожидания  $i$ -м требованием начала обслуживания. Время обслуживания  $i$ -того требования равно  $\zeta_i$ , причем  $\{\zeta_i\}$  – последовательность независимых одинаково распределенных случайных величин с общей функцией распределения  $B(x)$ . Искомыми являются вероятностные характеристики длины очереди и времени ожидания.

Представим данный процесс в виде КЛА. За состояние агрегата выберем вектор  $z=(v, z^{(v)})$ , где  $v$ - число требований, находящихся в системе в те-

кущий момент времени, которые дождутся (или дождались) начала обслуживания,  $\|v\|=v$ , а  $z^{(v)}$  – вектор, координаты которого определяются только при  $v>0$  и имеют следующий смысл:  $z_1^{(v)}$  – время, оставшееся до окончания обслуживания требования, находящегося на приборе, а  $z_i^{(v)}$  ( $1<i\leq v$ ) – длительности обслуживания требований, которые стоят в очереди и будут впоследствии обслужены. В соответствие с этим вектор  $\alpha^{(v)}$  скоростей изменения дополнительных координат имеет компоненты  $\alpha_1^{(v)}=-1$ ,  $\alpha_i^{(v)}=0$  ( $1<i\leq v$ ), многогранник  $Z^{(v)}$  при каждом  $v>0$  совпадает с первым октантом евклидова пространства размерности  $v$ :

$Z^{(v)} = \{z^{(v)} : z_i^{(v)} \geq 0, i = 1, 2, \dots, v\}$  и имеет  $m(v)=v$  ограничивающих координатных гиперплоскостей

$$Z_j^{(v)} = \{z^{(v)} : z_j^{(v)} = 0, j = 1, \dots, v\}.$$

В качестве входного сигнала рассмотрим пару  $(1, \theta)$ , где символ 1 просто указывает на факт поступления требования (и, таким образом, дискретная компонента  $\mu$  принимает лишь одно значение 1), а величина  $\theta$  равна допустимому времени ожидания поступающего требования (т.е.  $\|\mu\|=1$ ).

Рассмотрим динамику данного агрегата. Поскольку между моментами поступления входных сигналов изменяется лишь первая дополнительная координата (если  $v\geq 1$ ), то нет необходимости рассматривать случаи выхода дополнительных координат на гиперплоскости  $Z_j^{(v)}$ ,  $j>1$  (мы считаем, что  $\zeta_i>0$  с вероятностью 1). Пусть в момент  $t^*$  выполнено  $z(t^*)\in Z_1^{(v)}$ . Тогда с необходимостью

$$z(t^*) = (v, 0, z_2^{(v)}, \dots, z_v^{(v)}), \text{ где } v>0. \quad (6.37)$$

Физически это означает, что момент  $t^*$  является моментом окончания обслуживания и, следовательно, обслуженное требование должно покинуть систему, а его место занимает требование стоящее первым в очереди

(если очередь не пуста). Следовательно, из состояния (6.37) агрегат перейдет в состояние

$$z(t^{*+0}) = \begin{cases} 0, & \text{если } v = 1 \\ (v-1, z_2^{(v)}, \dots, z_v^{(v)}), & \text{если } v > 1 \end{cases} \quad (6.38)$$

Т.о. распределение  $P_1$  является вырожденным и приписывает вероятность 1 состоянию вида (6.38) (если предыдущее состояние имело вид (6.37) и вероятность 0 любому множеству состояний, не содержащему (6.38). Будем считать, что в рассматриваемый момент  $t^*$  выходной сигнал не выдается.

Рассмотрим теперь случай поступления входного сигнала  $(1, \theta)$  в момент  $t^{**}$ . Пусть при этом состояние агрегата было  $z(t^{**}) = (v, z^{(v)})$ , где  $v \geq 0$ . Физически поступление рассматриваемого входного сигнала означает приход требования, обладающего предельным временем ожидания  $\theta$ , в систему обслуживания. Рассмотрим 2 случая:  $v=0$  и  $v>0$ . В первом из них требование поступает в пустую систему, а во втором – в занятую обслуживанием. При  $v=0$  состояние  $z(t^{**+0})$  должно отражать тот факт, что поступившее требование сразу принято к обслуживанию и ему случайным образом назначено время обслуживания (с распределением  $V$ ), т.е.

$$z(t^{**+0}) = (1, \zeta), \quad (6.39)$$

где  $\zeta$  - случайная величина, имеющая распределение  $V(x)$ . При  $v>0$  состояние  $z(t^{**+0})$  должно отражать тот факт, что требование будет принято к обслуживанию тогда и только тогда, когда его предельное время ожидания  $\theta$  превосходит реальное (т.е.  $\theta > \sum_{i=1}^v z_i^{(v)}$ ) и в случае, если оно принято, ему назначается случайное время обслуживания с распределением  $V(x)$ .

$$z(t^{**}+0) = \begin{cases} z(t^{**}) = (v, z^{(v)}), & \text{если } \theta \leq \sum_{i=1}^v z_i^{(v)} \\ (v+1, z_1^{(v)}, \dots, z_v^{(v)}, \zeta), & \text{если } \theta > \sum_{i=1}^v z_i^{(v)} \end{cases}, \quad (6.40)$$

где  $\zeta(x)$  – случайная величина с распределением  $B(x)$ . Соотношения (6.39) и (6.40) конструктивно задают распределение  $P_2$ .

Будем считать, что в момент  $t^{**}$  выдается выходной сигнал, фиксирующий имеющуюся длину очереди, время ожидания и факт принятия или непринятия поступающего требования. В соответствии с этим предположим, что  $y = (\lambda, y^{(\lambda)})$ , где  $\lambda = (\lambda_1, \lambda_2)$ ,  $\lambda_1$  – число требований, находящихся в системе в момент  $t^{**}$ .  $\lambda_2 = 0$  или  $1$ , если поступающее требование не принимается или принимается соответственно к обслуживанию,  $\|\lambda\| = \lambda_2$ ,  $y^{(\lambda)}$  – время ожидания принятым требованиям начала обслуживания.

В данном случае  $\lambda$  представляет собой вектор размерности 2 с целочисленными компонентами. Из сказанного следует, что имеют место зависимости:

$$\lambda_1 = v,$$

$$\lambda_2 = \begin{cases} 0, & \text{если } \theta \leq \sum_{i=1}^v z_i^{(v)}, \\ 1, & \text{в противном случае} \end{cases}$$

$$y^{(\lambda)} = \sum_{i=1}^v z_i^{(v)} \quad (\text{координата определяется лишь при } \lambda_2=1).$$

На этом описание данной системы массового обслуживания в виде КЛА закончено.

## 6.5. Вопросы для самопроверки

1. Приведите определение понятия «агрегат» как типовой математической схемы описания элементов сложной системы.
2. Какие частные виды оператора переходов вводятся для агрегата общего вида?
3. Назовите возможные события, происходящие в процессе функционирования агрегата общего вида.
4. Какие принципы положены в основу алгоритма функционирования агрегата?

## ГЛАВА 7. СЕТИ ПЕТРИ

### 7.1. Основные определения

Сети Петри (СП) и их многочисленные модификации являются одним из классов моделей, неоспоримым достоинством которых является возможность адекватного представления не только структуры сложных организационно-технологических систем и комплексов, но также и логико-временных особенностей процессов их функционирования. Сети Петри представляют собой математическую модель для представления структуры и анализа динамики функционирования систем в терминах «условие-событие». Эта модель может быть успешно использована для описания так называемых динамических дискретных систем различных классов, таких как: вычислительные процессы и программы, технологические процессы, информационные, экономические, биологические, социальные и технические системы.

Модели сетей Петри позволяют исследовать работоспособность моделируемых систем, оптимальность их структуры, эффективность процесса их функционирования, а также возможность достижения в процессе функционирования определенных состояний. Сети Петри и их обобщения являются удобным и мощным средством моделирования асинхронных, параллельных, распределенных и недетерминированных процессов, позволяют наглядно представить динамику функционирования систем и составляющих их элементов. Свойство иерархического вложения сетей Петри позволяют рассматривать модели различной степени детализации, обеспечивая тем самым необходимую композицию сложных систем и процессов.

*Структура сети Петри.* Сеть Петри  $S$  является четверкой:  $S = (P, T, I, O)$ .  $P = \{p_1, p_2, \dots, p_n\}$  — конечное множество позиций,  $n \geq 0$ .  $T = \{t_1, t_2, \dots, t_m\}$  — конечное множество переходов,  $m \geq 0$ . Множество позиций и множество переходов не пересекаются,  $P \cap T = \emptyset$ .  $I: T \rightarrow P^\infty$  является входной функцией — отображением из переходов в комплекты позиций.  $O: T \rightarrow P^\infty$  есть выходная функция — отображение из переходов в комплекты позиций.

Позиция  $p_i$  является входной позицией перехода  $t_j$  в том случае, если  $p_i \in I(t_j)$ ;  $p_i$  является выходной позицией, если  $p_i \in O(t_j)$ . Входы и выходы переходов представляют собой комплекты позиций. Комплект является обобщением множества, в которое включены многократно повторяющиеся элементы — тиражированные элементы. Использование комплектов, а не множеств для входов и выходов перехода позволяет позиции быть кратным входом либо кратным выходом перехода. Кратность входной позиции  $p_i$  для перехода  $t_j$  есть число появлений позиции во входном комплекте перехода,  $\#(p_i, I(t_j))$ . Аналогично кратность выходной позиции  $p_i$  для перехода  $t_j$  есть число появлений позиции в выходном комплекте перехода,  $\#(p_i, O(t_j))$ . Если входная и выходная функции являются множествами (а не комплектами), то кратность каждой позиции есть либо 0, либо 1.

Существует и такой вариант определения входной и выходной функций переходов.  $I$  – входная функция переходов, которая определяется как отображение  $I: P \times T \rightarrow N_0$ ;  $O$  – выходная функция переходов, которое определяется как отображение  $O: T \times P \rightarrow N_0$ , где  $N_0$  – множество натуральных чисел и ноль.

Пример.  $S = (P, T, I, O)$ ,  $P = \{p_1, p_2, p_3, p_4, p_5\}$ ,  $T = \{t_1, t_2, t_3, t_4\}$ ,

$I(t_1) = \{p_1\}$ ,  $O(t_1) = \{p_2, p_3, p_5\}$ ,

$I(t_2) = \{p_2, p_3, p_5\}$ ,  $O(t_2) = \{p_5\}$ ,

$I(t_3) = \{p_3\}$ ,  $O(t_3) = \{p_4\}$ ,

$I(t_4) = \{p_4\}$ ,  $O(t_4) = \{p_2, p_3\}$ ,

*Графическое представление сети Петри.* Структура сети Петри представляет собой совокупность позиций и переходов. В соответствии с этим граф сети Петри обладает двумя типами узлов. Кружок  $\circ$  является позицией, а планка  $|$  - переходом.

Ориентированные дуги (стрелки) соединяют позиции и переходы, при этом некоторые дуги направлены от позиций к переходам, а другие — от переходов к позициям. Дуга, направленная от позиции  $p_i$  к переходу  $t_j$ , определяет позицию, которая является входом перехода. Кратные входы в

переход указываются кратными дугами из входных позиций в переход. Выходная позиция указывается дугой от перехода к позиции. Кратные выходы также представлены кратными дугами.

Граф  $G$  сети Петри есть двудольный ориентированный мультиграф,  $G = (V, A)$ , где  $V = \{v_1, v_2, \dots, v_s\}$  — множество вершин, а  $A = \{a_1, a_2, \dots, a_r\}$  — комплект направленных дуг,  $a_i = (v_j, v_k)$ , где  $v_j, v_k \in V$ . Множество  $V$  может быть разбито на два непересекающихся подмножества  $P$  и  $T$ , таких, что  $V = P \cup T$ ,  $P \cap T = \emptyset$ , и для любой направленной дуги  $a_i \in A$ , если  $a_i = (v_j, v_k)$ , тогда либо  $v_j \in P$  и  $v_k \in T$ , либо  $v_j \in T$ , а  $v_k \in P$ .

*Матричное представление сети Петри.* Альтернативным по отношению к определению сети Петри в виде  $(P, T, I, O)$  является определение двух матриц  $D^-$  и  $D^+$ , представляющих входную и выходную функции. Каждая матрица имеет  $n$  строк (по одной на позицию) и  $m$  столбцов (по одному на переход). Определим  $D^- [i, j] = \#(p_i, I(t_j))$ , а  $D^+ [i, j] = \#(p_i, O(t_j))$ .  $D^-$  определяет входы в переходы,  $D^+$  — выходы. Матричная форма определения сети Петри  $(P, T, D^-, D^+)$  эквивалентна стандартной форме, но позволяет дать определения в терминах векторов и матриц.

*Маркировка сетей Петри.* Маркировка  $\mu$  есть присвоение фишек позициям сети Петри. Фишка — это примитивное понятие сетей Петри (подобно позициям и переходам). Фишки присваиваются (можно считать, что они принадлежат) позициям. Количество и положение фишек при выполнении сети Петри могут изменяться. Фишки используются для определения выполнения сети Петри.

Маркировка  $\mu$  сети Петри  $S = (P, T, I, O)$  есть функция, отображающая множество позиций  $P$  в множество неотрицательных целых чисел  $N$ :  $\mu: P \rightarrow N$ . Маркировка  $\mu$  может быть также определена как вектор  $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ , где  $n = |P|$  и каждое  $\mu_i \in N$ ,  $i = 1, \dots, n$ . Вектор  $\mu$  определяет для каждой позиции  $p_i$  сети Петри количество фишек в этой позиции. Количество фишек в позиции  $p_i$  есть  $\mu_i$ ,  $i = 1, \dots, n$ . Связь между определениями марки-

ровки как функции и как вектора очевидным образом устанавливается соотношением  $\mu(p_i) = \mu_i$ .

Маркированная сеть Петри  $M = (C, \mu)$  есть совокупность структуры сети Петри  $C = (P, T, I, O)$  и маркировки  $\mu$  и может быть записана в виде  $M = (P, T, I, O, \mu)$ . На графе сети Петри фишки изображаются маленькой точкой в кружке, который представляет позицию сети Петри. На рис. 7.1 приведен пример графического представления маркированной сети Петри.

*Правила выполнения сетей Петри.* Сеть Петри выполняется посредством запусков переходов. Переход может запускаться только в том случае, когда он разрешен.

Переход  $t_j \in T$  в маркированной сети Петри  $C = (P, T, I, O)$  с маркировкой  $\mu$  разрешен, если для всех  $p_i \in P$   $\mu(p_i) \geq \#(p_i, I(t_j))$  (каждая из входных позиций перехода имеет число фишек, по крайней мере, равное числу дуг из позиции в переход).

Переход  $t_j$  в маркированной сети Петри с маркировкой  $\mu$ , может быть запущен всякий раз, когда он разрешен. В результате запуска разрешенного перехода  $t_j$  образуется новая маркировка  $\mu'$ , определяемая следующим соотношением:  $\mu'(p_i) = \mu(p_i) - \#(p_i, I(t_j)) + \#(p_i, O(t_j))$ . Это означает, что переход запускается удалением всех разрешающих фишек из его входных позиций и последующим помещением в каждую из его выходных позиций по одной фишке для каждой дуги. Кратные фишки создаются для кратных дуг.

В качестве примера рассмотрим маркированную сеть Петри, изображенную на рис. 7.1. При такой маркировке разрешены только три перехода:  $t_1$ ,  $t_3$  и  $t_4$ . Переход  $t_2$  не разрешен, так как ни позиция  $p_2$ , ни позиция  $p_3$ , являющиеся входами перехода  $t_2$ , не содержат ни одной фишки. Так как переходы  $t_1$ ,  $t_3$  и  $t_4$  разрешены, любой из них может быть запущен. Если запущен переход  $t_4$ , то происходит удаление фишки из каждого входа и по-

мещение фишки в каждый выход. При этом одна фишка удаляется из  $p_5$ , одна фишка помещается в  $p_3$ , а количество фишек в  $p_4$  увеличивается с двух до трех. Новая маркировка, полученная в результате запуска перехода  $t_4$ , показана на рис. 7.2.

В маркированной сети Петри, изображенной на рис. 7.2, разрешены только переходы  $t_1$  и  $t_3$ . При запуске перехода  $t_1$  осуществляется удаление фишки из  $p_1$  и помещение фишек в  $p_2$ ,  $p_3$  и  $p_4$  (в  $p_4$  — две фишки, так как эта позиция является кратным выходом перехода  $t_1$ ). Эта операция образует маркировку, приведенную на рис. 7.3.

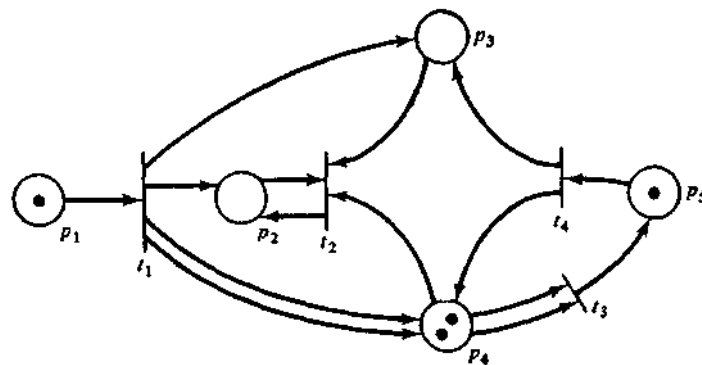


Рис. 7.1. Маркированная сеть Петри. Переходы  $t_1$ ,  $t_3$  и  $t_4$  разрешены.

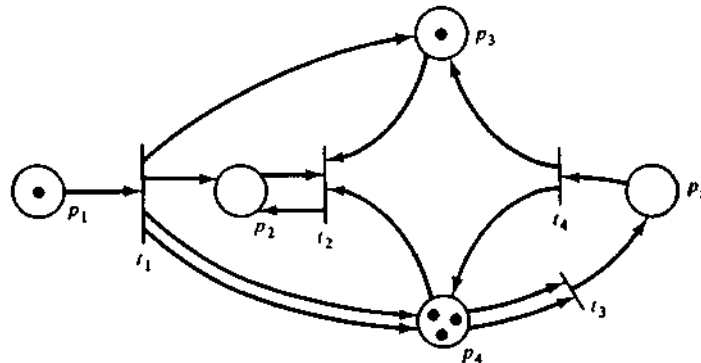


Рис. 7.2. Маркировка, полученная в результате запуска перехода  $t_4$  в сети на рис. 7.1.

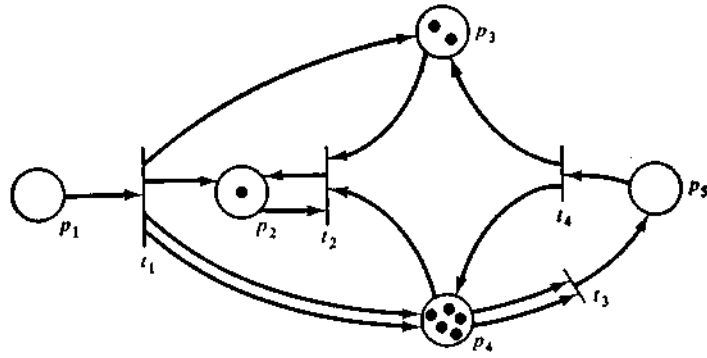


Рис. 7.3. Маркировка, полученная при запуске перехода  $t_1$  в сети на рис. 7.2.

Запуски могут осуществляться до тех пор, пока существует хотя бы один разрешенный переход. Когда не останется ни одного разрешенного перехода, выполнение прекращается.

*Пространство состояний сети Петри.* Состояние сети Петри определяется ее маркировкой. Запуск перехода изменяет состояние сети Петри посредством изменения маркировки сети. Пространство состояний сети Петри, обладающей  $n$  позициями, есть множество всех маркировок, т. е.  $\mathbb{N}^n$ . Изменение в состоянии, вызванное запуском перехода, определяется функцией изменения  $\delta$ , которую мы назовем функцией следующего состояния. Когда эта функция применяется к маркировке  $\mu$  (состоянию) и переходу  $t_j$  (он должен быть разрешен), она образует новую маркировку (состояние), которая получается при запуске перехода  $t_j$  в маркировке  $\mu$ .

Функция следующего состояния  $\delta: \mathbb{N}^n \times T \rightarrow \mathbb{N}^n$  для сети Петри  $S = (P, T, I, O)$  с маркировкой  $\mu$  и переходом  $t_j \in T$  определена тогда и только тогда, когда  $\mu(p_i) \geq \#(p_i, I(t_j))$  для всех  $p_i \in P$ . Если  $\delta(\mu, t_j)$  определена, то  $\delta(\mu, t_j) = \mu'$ , где  $\mu'(p_i) = \mu(p_i) - \#(p_i, I(t_j)) + \#(p_i, O(t_j))$  для всех  $p_i \in P$ .

Пусть дана сеть Петри  $S = (P, T, I, O)$  с начальной маркировкой  $\mu^\circ$ . Эта сеть может быть выполнена последовательными запусками переходов. Запуск разрешенного перехода  $t_j$  в начальной маркировке образует новую маркировку  $\mu^1 = \delta(\mu^\circ, t_j)$ . В этой новой маркировке можно запустить любой

другой разрешенный переход, например  $t_k$ , образующий новую маркировку  $\mu^2 = \delta(\mu^1, t_k)$ . Этот процесс будет продолжаться до тех пор, пока в маркировке будет существовать хотя бы один разрешенный переход. Если же получена маркировка, в которой ни один переход не разрешен, то никакой переход не может быть запущен, функция следующего состояния не определена для всех переходов, и выполнение сети должно быть закончено.

При выполнении сети Петри получают две последовательности: последовательность маркировок  $(\mu^0, \mu^1, \mu^2, \dots)$  и последовательность переходов, которые были запущены  $(t_{j_0}, t_{j_1}, t_{j_2}, \dots)$ . Эти две последовательности связаны следующим соотношением:  $\delta(\mu^k, t_{j_k}) = \mu^{k+1}$  для  $k = 0, 1, 2, \dots$ . Имея последовательность переходов, легко получить последовательность маркировок сети Петри, а имея последовательность маркировок, легко получить последовательность переходов, за исключением нескольких вырожденных случаев. Таким образом, обе эти последовательности представляют описание выполнения сети Петри.

Для сети Петри  $S = (P, T, I, O)$  с маркировкой  $\mu$  маркировка  $\mu'$  называется непосредственно достижимой из  $\mu$ , если существует переход  $t_j \in T$ , такой, что  $\delta(\mu, t_j) = \mu'$ .

Можно распространить это понятие на определение множества достижимых маркировок данной маркированной сети Петри. Если  $\mu'$  непосредственно достижима из  $\mu$ , а  $\mu''$  — из  $\mu'$ , говорят, что  $\mu''$  достижима из  $\mu$ . Определим множество достижимости  $R(S, \mu)$  сети Петри  $S$  с маркировкой  $\mu$  как множество всех маркировок, достижимых из  $\mu$ . Маркировка  $\mu'$  принадлежит  $R(S, \mu)$ , если существует какая-либо последовательность запусков переходов, изменяющих  $\mu$  на  $\mu'$ .

## 7.2. Сети Петри для моделирования

Сети Петри были разработаны и используются в основном для моделирования. С их помощью могут быть промоделированы многие системы, в особенности системы с независимыми компонентами, например аппаратное и программное обеспечение ЭВМ, физические системы, социальные и др. Сети Петри применяются для моделирования возникновения различных событий в системе. В частности, сети Петри могут моделировать поток информации или другие ресурсы системы.

Простое представление системы сетью Петри основано на двух основополагающих понятиях: событиях и условиях. События — это действия, имеющие место в системе. Возникновением событий управляет состояние системы. Состояние системы может быть описано множеством условий. Условие — есть предикат или логическое описание состояния системы. Условие может принимать либо значение «истина», либо значение «ложь».

Так как события являются действиями, то они могут происходить. Для того чтобы событие произошло, необходимо выполнение соответствующих условий. Эти условия называются предусловиями события. Возникновение события может вызвать нарушение предусловий и может привести к выполнению других условий, постусловий.

В качестве примера рассмотрим задачу моделирования простого автомата-продавца. Автомат-продавец находится в состоянии ожидания до тех пор, пока не появится заказ, который он выполняет и посылает на доставку. Условиями для такой системы являются: а) автомат-продавец ждет; б) заказ прибыл и ждет; в) автомат-продавец выполняет заказ; г) заказ выполнен. Событиями будут: 1. Заказ поступил. 2. Автомат-продавец начинает выполнение заказа. 3. Автомат-продавец заканчивает выполнение заказа. 4. Заказ посылается на доставку. Предусловия события 2 (автомат-продавец

начинает выполнение заказа) очевидны: (а) автомат-продавец ждет; (б) заказ прибыл и ждет. Постусловие для события 2: (в) автомат-продавец выполняет заказ. Аналогично мы можем определить предусловия и постусловия для всех остальных событий:

События	Предусловия	Постусловия
1	нет	б
2	а, б	в
3	в	г, а
4	г	нет

Такое представление системы легко моделировать сетью Петри. В сети Петри условия моделируются позициями, события — переходами. При этом входы перехода являются предусловиями соответствующего события; выходы — постусловиями. Возникновение события равносильно запуску соответствующего перехода. Выполнение условия представляется фишкой в позиции, соответствующей этому условию. Запуск перехода удаляет разрешающие фишки, представляющие выполнение предусловий и образует новые фишки, которые представляют выполнение постусловий. Сеть Петри на рис. 7.4 иллюстрирует модель приведенного выше автомата-продавца.

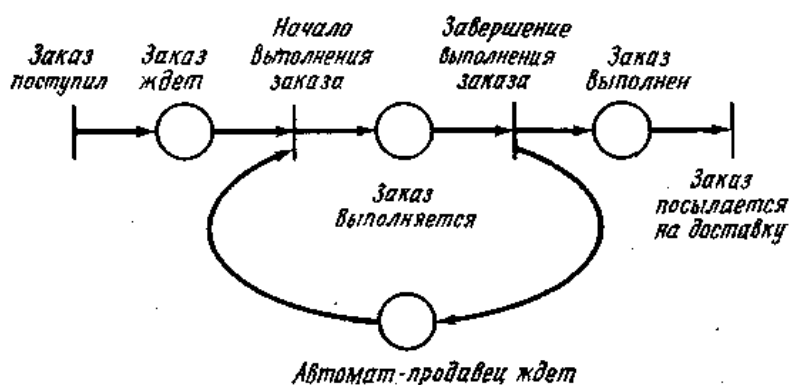


Рис. 7.4. Сеть Петри для простого автомата-продавца

Аналогичный пример можно привести для вычислительной системы, которая обрабатывает задания, поступающие с устройства ввода, и выводит результаты на устройство вывода. Задания поступают на устройство ввода. Когда процессор свободен и в устройстве ввода есть задание, процессор начинает обработку задания. Когда задание выполнено, оно посылается в устройство вывода; процессор же либо продолжает обрабатывать другое задание, если оно имеется, либо ждет прихода задания, если устройство ввода еще не получило такового. Эта система может быть промоделирована сетью Петри, показанной на рис. 7.5.

*Одновременность и конфликт.* Приведенные примеры иллюстрируют некоторые особенности сетей Петри и систем, моделируемых с их помощью. Одной из особенностей является свойственный сетям и их моделям параллелизм, или одновременность. В модели сети Петри два разрешенных не взаимодействующих события могут происходить независимо друг от друга. Синхронизировать события, пока это не потребуется моделируемой системе, нет нужды. Но, когда синхронизация необходима, моделировать ее легко. Таким образом, сети Петри представляются идеальными для моделирования систем с распределенным управлением, в которых несколько процессов выполняются одновременно.

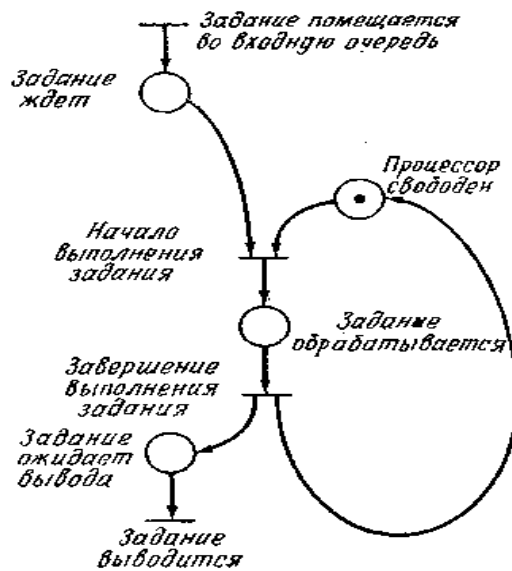


Рис. 7.5. Моделирование простой вычислительной системы

Другая важная особенность сетей Петри — это их асинхронная природа. В сети Петри отсутствует измерение времени или течение времени. Это отражает философский подход к понятию времени, утверждающий, что одно из важнейших свойств времени, с логической точки зрения, — это определение частичного упорядочения событий. В реальной жизни различные события укладываются в различные интервалы времени, и это отражено в модели сети Петри независимостью от времени управления последовательностью событий. Структура сети Петри такова, что содержит в себе всю необходимую информацию для определения возможных последовательностей событий. Таким образом, на рис. 7.5 событие «завершение выполнения задания» должно следовать за соответствующим событием «начало выполнения задания». Однако нет и не требуется никакой информации, связанной с количеством времени, необходимым на выполнение задания.

Выполнение сети Петри (или поведение моделируемой системы) рассматривается здесь как последовательность дискретных событий. Порядок появления событий является одним из возможных, допускаемых основной

структурой. Это приводит к явной недетерминированности в выполнении сети Петри. Если в какой-то момент времени разрешено более одного перехода, то любой из нескольких возможных переходов может стать «следующим» запускаемым. Выбор запускаемого перехода осуществляется недетерминированным образом, т. е. случайно. Эта особенность сети Петри отражает тот факт, что в реальной жизненной ситуации, где несколько действий происходит одновременно, возникающий порядок появления событий — не однозначен; скорее может возникнуть любая из множества последовательностей событий. Однако частичный порядок появления события — единственен.

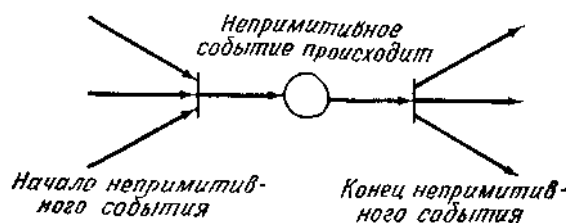


Рис. 7.6. Моделирование "непримитивного события".

Для простоты обычно вводят следующее ограничение. Запуск перехода (и соответствующего события) рассматривается как мгновенное событие, занимающее нулевое время, и возникновение двух событий одновременно невозможно. Моделируемое таким образом событие называется примитивным; примитивные события мгновенны и не одновременны. Непримитивными называются такие события, длительность которых отлична от нуля. Они не являются не одновременными и, следовательно, могут пересекаться во времени. Так как осуществление большинства событий в реальном мире занимает некоторое время, то они являются непримитивными событиями и поэтому не могут должным образом моделироваться переходами в сети Петри. Однако это не приводит к возникновению проблем при моделировании систем. Непримитивное событие может быть представлено в виде

двух примитивных событий: «начало непримитивного события», «конец непримитивного события» и условия «непримитивное событие происходит». Эта ситуация может быть промоделирована с помощью сети, показанной на рис. 7.6.

Недетерминированность и неодновременность запусков переходов в моделировании параллельной системы показываются двумя способами. Один из них представлен на рис. 7.7. В этой ситуации два разрешенных перехода в любом случае не влияют друг на друга, и в число возможных последовательностей событий входит последовательность, в которой первым срабатывает один переход, и последовательность, в которой первым будет запущен другой переход. Это называется одновременностью.

Другая ситуация, в которой одновременное выполнение затруднено и которая характеризуется невозможностью одновременного возникновения событий, показана на рис. 7.8. Здесь два разрешенных перехода находятся в конфликте. Может быть запущен только один переход, так как при запуске он удаляет фишку из общего входа и запрещает другой переход. Рассмотренные ситуации требуют внимательного изучения моделируемых сетями Петри систем для правильного отображения их поведения.

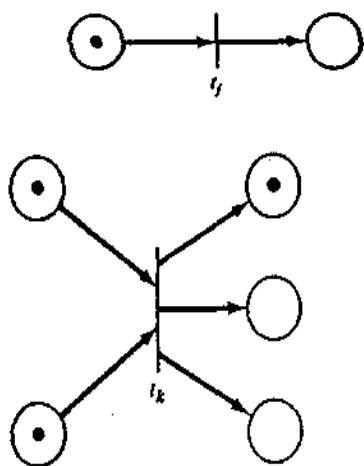


Рис. 7.7. Одновременность. Эти два перехода могут быть запущены в любом порядке

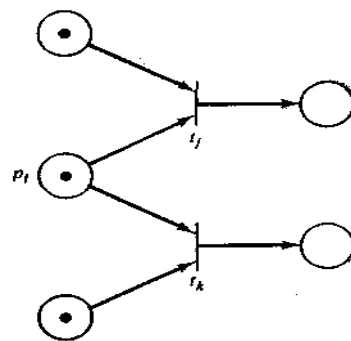


Рис. 7.8. Конфликт. Переходы  $t_i$  и  $t_k$  находятся в конфликте, т.е. запуск одного из них удаляет фишку из  $p_i$  и тем самым запрещает другой

### 7.3. Анализ сетей Петри

#### 7.3.1. Задачи анализа сетей Петри

*Безопасность.* Позиция  $p_i \in P$  сети Петри  $C = (P, T, I, O)$  с начальной маркировкой  $\mu$  является безопасной, если  $\mu'(p_i) \leq 1$  для любой  $\mu' \in R(C, \mu)$ . Сеть Петри безопасна, если безопасна каждая ее позиция. Безопасность — очень важное свойство для устройств аппаратного обеспечения. Если позиция безопасна, то число фишек в ней равно 0 или 1. Следовательно, позицию можно реализовать одним триггером.

*Ограниченность.* Безопасность — это частный случай более общего свойства ограниченности. Позиция  $p_i \in P$  сети Петри  $C = (P, T, I, O)$  с начальной маркировкой  $\mu$  является  $k$ -безопасной, если  $\mu'(p_i) \leq k$  для всех  $\mu' \in R(C, \mu)$ . 1-безопасная позиция называется просто безопасной. Заметим, что граница  $k$  по числу фишек, которые могут находиться в позиции, может быть функцией от позиции (например, позиция  $p_1$  может быть 3-безопасной, тогда как позиция  $p_2$  — 8-безопасной). Однако, если позиция  $p_i$   $k$ -безопасна, то она также и  $k'$ -безопасна для всех  $k' \geq k$ . Поскольку число позиций конечно, можно выбрать  $k^*$ , равное максимуму из границ каждой позиции, и определить сеть Петри  $k^*$ -безопасной, если каждая позиция сети  $k^*$ -безопасна.

Иногда интересуются только тем, является число фишек в позиции ограниченным или нет, а не конкретным значением границы. Позиция называется ограниченной, если она  $k$ -безопасна для некоторого  $k$ ; сеть Петри ограничена, если все ее позиции ограничены. Ограниченную сеть Петри можно реализовать аппаратно, тогда как сеть Петри с неограниченными позициями в общем случае реализовать аппаратно нельзя.

*Сохранение.* Сети Петри можно использовать для моделирования систем распределения ресурсов. В этом случае некоторые фишки могут представлять собой ресурсы. Для сетей Петри такого типа важным свойством является сохранение.

Сеть Петри  $C = (P, T, I, O)$  с начальной маркировкой  $\mu$  называется строго сохраняющей, если для всех  $\mu' \in R(C, \mu)$

$$\sum_{p_i \in P} \mu'(p_i) = \sum_{p_i \in P} \mu(p_i).$$

Строгое сохранение — это очень сильное ограничение. Например, из него немедленно следует, что число входов в каждый переход должно равняться числу выходов,  $|I(t_j)| = |O(t_j)|$ . Если бы это было не так, запуск перехода изменил бы число фишек в сети.

Сеть Петри должна сохранять ресурсы, которые она моделирует. Однако взаимно однозначного соответствия между фишками и ресурсами нет. Фишка может представлять один элемент или несколько ресурсов сразу. Во втором случае фишка позже используется для создания кратных фишек (по одной на ресурс) путем запуска перехода с большим числом выходов, чем входов. В общем, следует определить взвешивание фишек. Взвешенная сумма для всех достижимых маркировок должна быть постоянной. Фишкам, не являющимся важными, можно присвоить вес 0; другим фишкам можно присвоить веса 1, 2, 3 или любое другое целое. Фишка определяется ее позицией в сети, все фишки в позиции неразличимы. Следовательно, веса связываются с каждой позицией сети. Вектор взвешивания  $v = (v_1, v_2, \dots, v_n)$  определяет вес  $v_i$  для каждой позиции  $p_i \in P$ .

Сеть Петри  $C = (P, T, I, O)$  с начальной маркировкой  $\mu$  называется сохраняющей по отношению к вектору взвешивания  $v = (v_1, v_2, \dots, v_n)$ ,  $n = |P|$ ,  $v_i \geq 0$ , если для всех  $\mu' \in R(C, \mu)$

$$\sum_i v_i \cdot \mu'(p_i) = \sum_i v_i \cdot \mu(p_i).$$

Строго сохраняющая сеть Петри является сохраняющей по отношению к вектору взвешивания  $(1, 1, \dots, 1)$ . Все сети Петри являются сохраняющими по отношению к вектору взвешивания  $(0, 0, \dots, 0)$ . Этот факт вносит в теорию некоторую нестройность. Поэтому можно потребовать, чтобы сеть Петри называлась сохраняющей, если она сохраняющая по отношению к некоторому положительному ненулевому вектору взвешивания  $v > 0$  (с положительными ненулевыми весами  $v_i > 0$ ).

*Активность.* Причиной рассмотрения сохранения в сети Петри было распределение ресурсов, другая задача, которая может возникнуть при распределении ресурсов — тупики. Тупики служат предметом многих исследований, например в области вычислительной техники.

Тупик в сети Петри — это переход (или множество переходов), которые не могут быть запущены. Переход называется активным, если он не заблокирован (нетупиковый). Это не означает, что переход разрешен, скорее он может быть разрешенным. Переход  $t_j$  сети Петри  $S$  называется потенциально запускимым в маркировке  $\mu$ , если существует маркировка  $\mu' \in R(S, \mu)$ , в которой  $t_j$  разрешен. Переход активен в маркировке  $\mu$ , если потенциально запустим во всякой маркировке из  $R(S, \mu)$ . Следовательно, если переход активен, то всегда возможно перевести сеть Петри из ее текущей маркировки в маркировку, в которой запуск перехода станет разрешенным.

Существуют другие, связанные с активностью понятия, которые рассматривались при изучении тупиков. Их можно разбить на категории по уровню активности и определить для сети Петри  $S$  с маркировкой  $\mu$  следующим образом:

Уровень 0: Переход  $t_j$  обладает активностью уровня 0 и называется пассивным или мертвым, если он никогда не может быть запущен.

Уровень 1: Переход  $t_j$  обладает активностью уровня 1 и называется потенциально активным или живым, если он потенциально запустим, т. е. если существует такая  $\mu' \in R(S, \mu)$ , что  $t_j$  разрешен в  $\mu'$ .

Уровень 2: Переход  $t_j$  обладает активностью уровня 2, если для всякого целого  $n$  существует последовательность запусков, в которой  $t_j$  присутствует по крайней мере  $n$  раз.

Уровень 3: Переход  $t_j$  обладает активностью уровня 3, если существует бесконечная последовательность запусков, в которой  $t_j$  присутствует неограниченно часто.

Уровень 4: Переход  $t_j$  обладает активностью уровня 4 и называется активным или живым, если для всякой  $\mu' \in R(C, \mu)$  существует такая последовательность запусков  $\sigma$ , что  $t_j$  разрешен в  $\delta(\mu', \sigma)$ .

Сеть Петри обладает активностью уровня  $i$ , если каждый ее переход обладает активностью уровня  $i$ .

*Устойчивость переходов и СП.* Переход  $t_i$  СП  $C = (P, T, I, O)$  с маркировкой  $\mu$  называется устойчивым, если он активен при данной маркировке и никакой другой переход  $t_k$ , тоже активный при маркировке  $\mu$ , не может, сработав, сделать переход  $t_i$  неактивным, т.е. лишив его возможности срабатывания. СП  $C = (P, T, I, O)$  с маркировкой  $\mu$  называется устойчивой, если все ее переходы являются устойчивыми.

*Достижимость и покрываемость.* Большинство задач, к которым мы до сих пор обращались, касаются достижимых маркировок.

*Задача достижимости.* Для данной сети Петри  $C$  с маркировкой  $\mu$  и маркировки  $\mu'$  определить:  $\mu' \in R(C, \mu)$ ? Задача достижимости, быть может, основная задача анализа сетей Петри; многие другие задачи анализа можно сформулировать в терминах задачи достижимости.

*Задача покрываемости.* Эта задача аналогична достижимости, но несколько отличается. Для данной сети Петри  $C$  с начальной маркировкой  $\mu$  и маркировки  $\mu'$  определить, существует ли такая достижимая маркировка  $\mu'' \in R(C, \mu)$ , что  $\mu'' \geq \mu'$ .

В других возможных задачах типа достижимости могло бы игнорироваться содержимое некоторых позиций и приниматься во внимание сравнение или покрытие содержимого нескольких важных позиций. Эти задачи называются задачами *достижимости подмаркировки* и *покрываемости подмаркировки*.

*Последовательности запусков.* Другой предложенный к анализу подход основан не на позициях, а на последовательностях запусков переходов, т. е. связан с активностью, поскольку уместен вопрос: может ли переход быть запущен (иначе, не является ли он пассивным)? В более общем случае можно потребовать определить, возможна ли заданная последовательность запусков переходов или возможна ли какая-либо последовательность из множества последовательностей запусков.

*Задачи эквивалентности и подмножества.* Последний класс задач порожден соображениями оптимизации. Пусть сети Петри присуще некоторое поведение, определяемое ее множеством последовательностей запусков переходов или ее множеством достижимости. Можно ли изменить (оптимизировать) сеть Петри, не изменяя ее поведения? Изменение означает удаление пассивных переходов (которые никогда нельзя запустить) или пассивных позиций {которые никогда не могут быть маркированы}, или переопределение некоторых переходов. Можно ли показать, что две различные маркированные сети Петри с одинаковым числом переходов (но с различным числом позиций) будут порождать одинаковые последовательности запусков переходов или что две различные маркированные сети Петри с одинаковым числом позиций (но с различным числом переходов) будут порождать одно множество достижимости? Это позволило бы модифицировать сети Петри для увеличения параллелизма, уменьшения стоимости реализации или улучшения других оптимизируемых функционалов. В этих случаях затрагивается проблема определения того, являются ли две сети Петри эквивалентными или является ли одна из них подмножеством другой. Для точного определения понятия эквивалентности или включения необходимо быть особенно внимательным. Если определить эквивалентность как равенство множеств достижимости, тогда нельзя будет изменять число позиций, если потребовать равенства множеств после-

довательностей запусков переходов — нельзя будет изменять переходы. Поэтому определение задачи, которое дается, исключительно важно.

### 7.3.2. Методы анализа

Существуют два основных метода анализа сетей Петри, которые описывают механизмы решения некоторых из уже перечисленных задач. Первый метод анализа, используемый для сетей Петри, — это дерево достижимости, второй метод связан с матричными уравнениями. Обсудим подробнее первый из них.

*Дерево достижимости.* Дерево достижимости представляет множество достижимости сети Петри. Рассмотрим, например, маркированную сеть Петри на рис. 7.9. Начальная маркировка ее —  $(1, 0, 0)$ . В этой начальной маркировке разрешены два перехода:  $t_1$  и  $t_2$ . Пусть корнем является вершина, соответствующая начальной маркировке, определим новые вершины в дереве достижимости для (достижимых) маркировок, получающихся в результате запуска каждого из этих двух переходов. Дуга, помеченная запускаемым переходом, приводит из начальной маркировки к каждой из новых маркировок (рис. 7.10). Это (частичное) дерево показывает все маркировки, непосредственно достижимые из начальной маркировки.

Теперь необходимо рассмотреть все маркировки, достижимые из новых маркировок. Из маркировки  $(1, 1, 0)$  можно снова запустить  $t_1$  (получая  $1, 2, 0$ ) и  $t_2$  (получая  $(0, 2, 1)$ ); из  $(0, 1, 1)$  можно запустить  $t_3$  (получая  $(0, 0, 1)$ ). Это порождает дерево, изображенное на рис. 7.11. Начиная с трех новых маркировок, необходимо повторить этот процесс, порождая новые маркировки, которые нужно ввести в дерево, как показано на рис. 7.12.

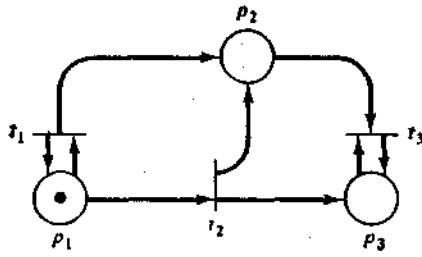


Рис. 7.9. Маркированная сеть Петри, для которой строится дерево достижимости.

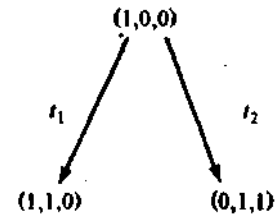


Рис. 7.10. Первый шаг построения дерева достижимости

Заметим, что маркировка  $(0, 0, 1)$  пассивная; никакой переход в ней не является разрешенным, поэтому никакие новые маркировки из этой пассивной маркировки в дереве породиться не будут. Кроме того, необходимо отметить, что маркировка  $(0, 1, 1)$ , порождаемая запуском  $t_3$  в  $(0, 2, 1)$ , была уже порождена непосредственно из начальной маркировки запуском  $t_2$ .

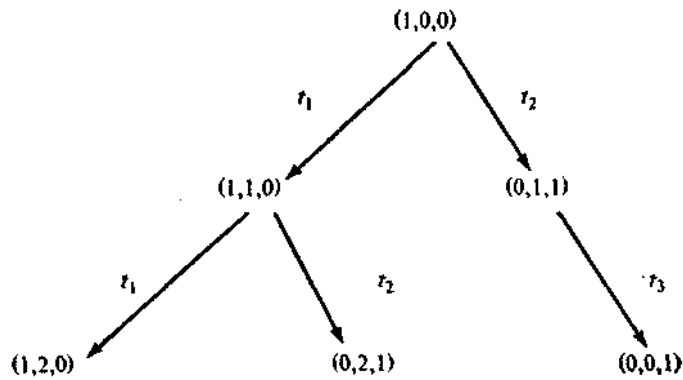


Рис. 7.11. Второй шаг построения дерева достижимости.

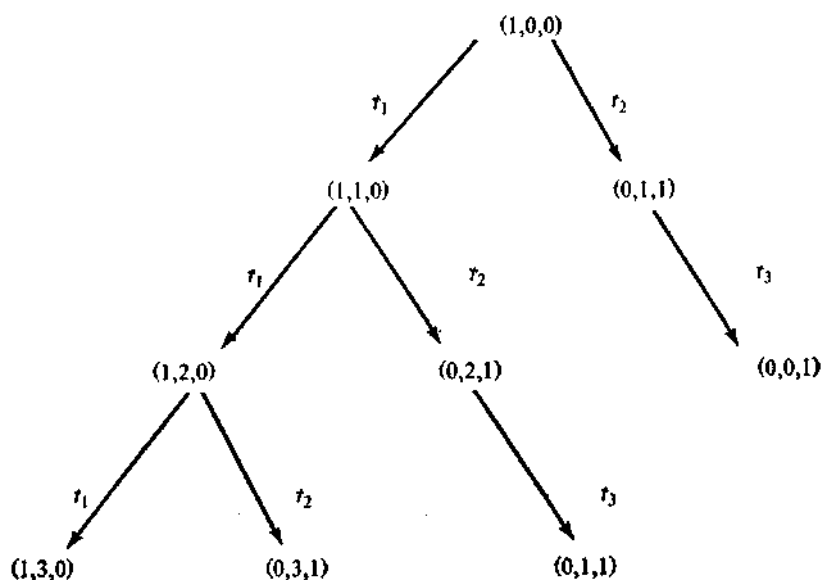


Рис. 7.12. Третий шаг построения дерева достижимости.

Если эту процедуру повторять снова и снова, каждая достижимая маркировка окажется порожденной. Однако получившееся дерево достижимости может оказаться бесконечным. Будет порождена каждая маркировка из множества достижимости, поэтому для любой сети Петри с бесконечным множеством достижимости соответствующее дерево также должно быть бесконечным. Даже сеть Петри с конечным множеством достижимости может иметь бесконечное дерево. Дерево представляет все возможные последовательности запусков переходов. Всякий путь в дереве, начинающийся в корне, соответствует допустимой последовательности переходов. Для превращения дерева в полезный инструмент анализа необходимо найти средства ограничения его до конечного размера.

При графическом представлении множества достижимости в дереве достижимости предполагается, что равные маркировки, которым соответствуют различные последовательности переходов, изображаются различными вершинами на дереве. В этом случае в каждую вершину дерева будет вести единственный ориентированный путь от корневой вершины, которому соответствует единственная последовательность переходов. Если же с целью удобства и сокращения геометрических размеров дерева достижимости равным маркировкам ставить в соответствие одну вершину, то полученный в результа-

те граф получил название диаграммы достижимых маркировок СП или диаграммы состояний СП. В последнем случае от корневой вершины диаграммы в каждую из ее вершин может вести несколько ориентированных путей, каждому из которых будет соответствовать своя последовательность переходов.

Приведение к конечному представлению осуществляется несколькими способами. Нам необходимо найти те средства, которые ограничивают введение новых маркировок (называемых граничными вершинами) на каждом шаге. Здесь могут помочь пассивные маркировки — маркировки, в которых нет разрешенных переходов. Эти пассивные маркировки называются терминальными вершинами. Другой класс маркировок — это маркировки, ранее встречавшиеся в дереве. Такие дублирующие маркировки называются дублирующими вершинами; никакие последующие маркировки рассматривать нет нужды — все они будут порождены из места первого появления дублирующей маркировки в дереве. Таким образом, в дереве на рис. 7.12 маркировка  $(0, 1, 1)$ , получившаяся в результате выполнения последовательности  $t_1 t_2 t_3$ , не будет порождать какие-либо новые вершины в дереве, поскольку она ранее встречалась в дереве в результате выполнения последовательности  $t_3$  из начальной маркировки.

Для сведения дерева достижимости к конечному представлению используется еще одно средство. Рассмотрим последовательность запусков переходов  $\sigma$ , начинающуюся в начальной маркировке  $\mu$  и кончающуюся в маркировке  $\mu'$ ,  $\mu' > \mu$ . Маркировка  $\mu'$  совпадает с маркировкой  $\mu$ , за исключением того, что имеет некоторые «дополнительные» фишки в некоторых позициях, т. е.  $\mu' = \mu + (\mu' - \mu)$  и  $(\mu' - \mu) > 0$ . Теперь, поскольку на запуски переходов лишние фишки не влияют, последовательность  $\sigma$  можно запустить снова, начиная с  $\mu'$ , приходя к маркировке  $\mu''$ . Так как действие последовательности переходов  $\sigma$  добавило  $\mu' - \mu$  фишек к маркировке  $\mu$ , она добавит также  $\mu' - \mu$  фишек и к маркировке  $\mu'$ , поэтому  $\mu'' = \mu' + (\mu' - \mu)$  или  $\mu'' = \mu + 2(\mu' - \mu)$ . В общем можно запустить последовательность  $\sigma$   $n$

раз, получив в результате маркировку  $\mu + n(\mu' - \mu)$ . Следовательно, для тех позиций, которые увеличивают число фишек последовательностью  $\sigma$ , можно создать произвольно большое число фишек, просто повторяя последовательность  $\sigma$  столько, сколько это нужно. В сети Петри на рис. 7.9, например, можно запустить переход  $t_1$  столько раз, сколько необходимо для того, чтобы получить произвольное число фишек в  $p_2$ .

Представим бесконечное число маркировок, получающихся из циклов такого типа, с помощью специального символа  $\omega$ , который обозначает «бесконечность». Для любого постоянного  $a$  определим

$$\omega + a = \omega \quad a < \omega,$$

$$\omega - a = \omega \quad \omega \leq \omega.$$

Для построения дерева достижимости необходимы только эти операции над  $\omega$ .

Теперь можно точно сформулировать действительный алгоритм построения дерева достижимости. Каждая вершина  $i$  дерева связывается с расширенной маркировкой  $\mu[i]$ ; в расширенной маркировке число фишек в позиции может быть либо неотрицательным целым, либо  $\omega$ . Каждая вершина классифицируется как граничная вершина, терминальная вершина, дублирующая вершина, или как внутренняя вершина. Граничными являются вершины, которые еще не обработаны алгоритмом; алгоритм превратит их в терминальные, дублирующие или внутренние вершины.

Алгоритм начинает с определения начальной маркировки корнем дерева, т. е. граничной вершиной. До тех пор пока имеются граничные вершины, они обрабатываются алгоритмом.

Пусть  $x$  — граничная вершина, которую необходимо обработать.

Если в дереве имеется другая вершина  $y$ , не являющаяся граничной, и с ней связана та же маркировка,  $\mu[x] = \mu[y]$ , то вершина  $x$  — дублирующая.

Если для маркировки  $\mu[x]$  ни один из переходов не разрешен (т. е.  $\delta(\mu[x], t_j)$  не определено для всех  $t_j \in T$ ), то  $x$  — терминальная вершина.

Для всякого перехода  $t_j \in T$ , разрешенного в  $\mu[x]$  (т. е.  $\delta(\mu[x], t_j)$  определено), создать новую вершину  $z$  дерева достижимости. Маркировка  $\mu[z]$ , связанная с этой вершиной, определяется для каждой позиции  $p_i$  следующим образом:

а) Если  $\mu [x]_i = \omega$ , то  $\mu [z]_i = \omega$ .

б) Если на пути от корневой вершины к  $x$  существует вершина  $y$  с  $\mu [y] \leq \delta (\mu [x], t_j)$  и  $\mu [y]_i < \delta (\mu [x], t_j)_i$ , то  $\mu [z]_i = \omega$ .

в) В противном случае  $\mu [z]_i = \delta (\mu [x], t_j)_i$ .

Дуга, помеченная  $t_j$ , направлена от вершины  $x$  к вершине  $z$ . Вершина  $x$  переопределяется как внутренняя, вершина  $z$  становится граничной.

Когда все вершины дерева — терминальные, дублирующие или внутренние, алгоритм останавливается.

На рис. 7.13 представлено дерево достижимости для сети Петри на рис. 7.9.

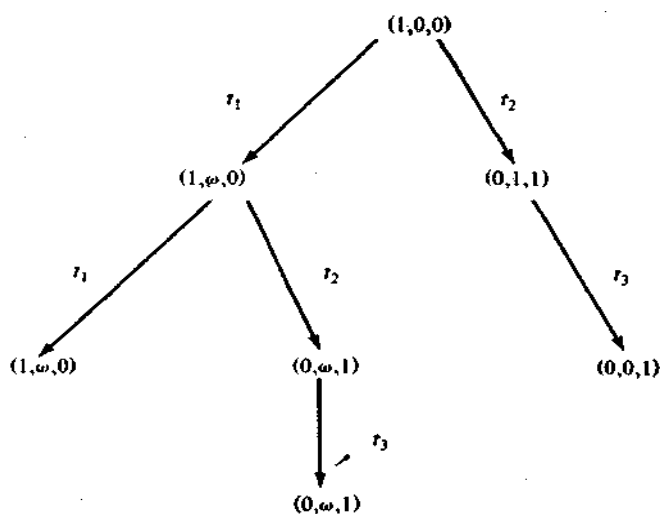


Рис. 7.13. Дерево достижимости сети Петри, приведенной на рис. 7.9.

Дерево достижимости — очень полезный инструмент анализа сетей Петри. Покажем, как его можно использовать для решения некоторых задач, представленных выше.

Безопасность и ограниченность. Сеть Петри безопасна, если число фишек в каждой позиции не может превысить 1; сеть Петри ограничена, если существует такое целое  $k$ , что число фишек в любой позиции не может превысить  $k$ . Оба этих свойства можно проверить с помощью дерева достижимости. Сеть Петри ограничена тогда и только тогда, когда символ  $\omega$  отсутствует в ее дереве достижимости. Присутствие символа  $\omega$  в дереве достижимости означает, что число фишек потенциально не ограниче-

но; существует последовательность запусков переходов, которую можно повторить произвольное число раз, увеличивая количество фишек до произвольно большого числа. Таким образом, если имеется символ  $\omega$ , сеть неограниченна. Кроме того, положение символа  $\omega$  показывает, какие позиции неограниченны.

Обратное утверждение также является верным, если сеть Петри неограниченна, то число достижимых маркировок бесконечно. Поскольку дерево достижимости конечно, бесконечное число достижимых маркировок отражает присутствие символа  $\omega$ .

Если сеть Петри ограничена и символ  $\omega$  отсутствует в дереве достижимости, то сеть Петри представляет систему конечных состояний. Дерево достижимости, по существу, является графом состояний и будет содержать вершину, соответствующую всякой достижимой маркировке. Это позволяет решить вопросы анализа простым перебором и проверкой конечного множества всех достижимых маркировок. Например, чтобы определить границу для заданной позиции, нужно построить дерево достижимости и найти наибольшее значение компоненты маркировки, соответствующей этой позиции.

Найденное значение является границей числа фишек для заданной позиции. Если граница для всех позиций не превышает 1, сеть безопасна.

Отметим, что по дереву достижимости даже для сетей Петри, не являющихся ограниченными (вследствие неограниченности некоторой позиции), можно определить границы для тех позиций, которые являются ограниченными. Таким образом, дерево достижимости позволяет эффективно решить задачи анализа сетей Петри по определению ограниченности и безопасности для отдельных позиций и целых сетей.

Сохранение. Сеть Петри является сохраняющей по отношению к вектору взвешивания, если взвешенная сумма фишек постоянна для всех дос-

тижимых маркировок. Свойство сохранения эффективно проверяется с помощью дерева достижимости. Так как дерево достижимости конечно, для каждой маркировки можно вычислить взвешенную сумму. Если сумма одинакова для каждой достижимой маркировки, сеть — сохраняющая по отношению к данному весу. Если суммы не равны, сеть — несохраняющая.

При оценке сохранения необходимо быть внимательным с символом  $\omega$ . Если маркировка имеет  $\omega$  в качестве маркировки позиции  $p_i$ , тогда для того, чтобы сеть была сохраняющей, вес этой позиции должен быть равным 0. Напомним, что символ  $\omega$  представляет бесконечное множество значений. Так как все веса неотрицательны, вес должен равняться либо нулю (тем самым означая, что число фишек в данной позиции не важно), либо быть положительным. Если вес положителен, то сумма будет разной для двух маркировок, различающихся в своей  $\omega$ -компоненте. Следовательно, если какая-либо маркировка с ненулевым весом равна  $\omega$ , сеть — несохраняющая.

Эти рассуждения относятся к сохранению по отношению к определенному взвешиванию. Сеть Петри является сохраняющей, если она сохраняющая по отношению к некоторому вектору  $v$ ,  $v_i > 0$ . Дерево достижимости можно использовать для определения того, является сеть сохраняющей или нет, путем нахождения вектора весов  $v$  (если он существует). Заметим прежде всего, что для того, чтобы сеть Петри была сохраняющей по отношению к положительному вектору весов, она должна быть ограниченной. Как было указано выше, неограниченная позиция должна иметь нулевой вес, что недопустимо в сети с положительным вектором весов. Теперь, если сеть сохраняющая, существуют взвешенная сумма, обозначим ее  $s$ , и вектор весов  $v = (v_1, v_2, \dots, v_n)$ . Для каждой маркировки  $\mu [x]$  дерева достижимости имеем

$$v_1 * \mu[x]_1 + v_2 * \mu[x]_2 + \dots + v_n * \mu[x]_n = s.$$

Это равенство определяет для  $k$  вершин дерева достижимости совокупность из  $k$  линейных уравнений с  $n + 1$  неизвестными. Добавим к ним ограничения:  $v_i > 0$ ,  $i = 1, \dots, n$ , в результате чего определим ограничения для вектора весов.

Решение этой системы линейных уравнений — хорошо известная задача, имеющая множество алгоритмов решения. Можно рассматривать ее как задачу линейного программирования или просто как систему линейных уравнений. В любом случае, если решение существует, оно будет вычислено. Если ограничения, накладываемые на веса, являются чрезмерно жесткими и, следовательно, вектора взвешиваний не существует, это будет определено. В любом случае можно определить, является или нет сеть Петри сохраняющей, и если это так, получить вектор весов.

Покрываемость. Последняя задача, которую можно решить с помощью дерева достижимости, — задача покрываемости. В задаче покрываемости хотят для данной маркировки  $\mu'$  определить, достижима ли маркировка  $\mu'' \geq \mu'$ . Данная задача решается проверкой дерева достижимости. Строим для начальной маркировки  $\mu$  дерево достижимости. Затем ищем любую вершину  $x$  с  $\mu[x] \geq \mu'$ . Если такой вершины не существует, маркировка  $\mu'$  не покрывается никакой достижимой маркировкой; если она найдена,  $\mu[x]$  дает достижимую маркировку, покрывающую  $\mu'$ .

Путь от корня к покрывающей маркировке определяет последовательность переходов, которые приводят из начальной маркировки к покрывающей маркировке, а маркировка, связанная с этой вершиной, определяет покрывающую маркировку. Символ  $\omega$  вновь должен рассматриваться как обозначение бесконечного множества значений. Если компонента покрывающей маркировки —  $\omega$ , то в пути от корня к покрывающей маркировке имеется «цикл». Для увеличения соответствующей компоненты с тем, что-

бы она была не меньше, чем в данной маркировке, необходимо достаточное число раз повторить этот цикл.

Ограниченность дерева достижимости. Как видим, дерево достижимости можно использовать для решения задач безопасности, ограниченности, сохранения и покрываемости. Но, в общем случае его нельзя использовать для решения задач достижимости и активности, а также для определения возможной последовательности запусков. Решение этих задач ограничено существованием символа  $\omega$ . Символ  $\omega$  означает потерю информации; конкретные количества фишек отбрасываются, учитывается только существование их большого числа.

#### 7.4. Обобщения сетей Петри

*Цветные сети Петри.* Появление сетей этого класса связано с концепцией использования различных меток. Ранее все метки предполагались одинаковыми. Механизм функционирования сетей был связан только лишь с количествами меток во входных позициях переходов и определялся общими для всех меток условиями возбуждения переходов и правилами изменения различных позиций при выполнении сети. В цветных сетях каждая метка получает свой цвет. Условия возбуждения и правила срабатывания переходов для меток каждого цвета задаются независимо. Множество используемых при реализации цветных сетей красок выбирается конечным или бесконечным (например, счётным). При моделировании систем цветные сети чаще всего используются для построения компактных формальных и графических представлений, в составе которых имеются однотипные по структуре и характеру функционирования группы объектов.

*Сети Петри со сдерживающими дугами.* Сдерживающая дуга из позиции  $p_i$  в переход  $t_j$  имеет маленький кружок (а не стрелку). Кружок означа-

ет отрицание (“не”). Правила запуска изменяются следующим образом: переход является разрешённым, когда фишки присутствуют во всех его (обычных) входах и отсутствуют в сдерживающих входах. Переход запускается удалением фишек из всех его (обычных) входов. Так можно описывать переходы «исключающее ИЛИ». В обычных СП переход запускается, когда все его входы имеют фишки (логика И). Переход “исключающее ИЛИ” запускается тогда и только тогда, когда только один из его входов имеет фишки, а все другие фишек не имеют. Когда переход запускается, он удаляет фишку только из входа с фишками.

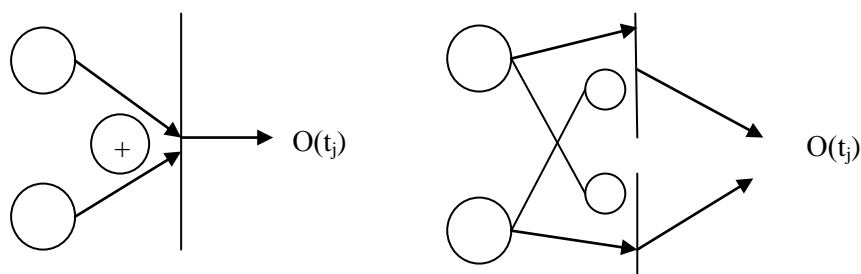


Рис. 7.14. Интерпретация перехода “исключающее ИЛИ” с помощью сдерживающих дуг

Имеются ещё два других важных расширения СП. Переходам могут быть поставлены в соответствие приоритеты так, что если  $t_i$  и  $t_k$  оба допустимы, то переход с высшим приоритетом будет запущен первым. Механизм назначения приоритетов может устанавливать порядок срабатывания переходов при возникновении конфликтов. Во-вторых, используют временные сети Петри. Во временных сетях Петри каждому переходу  $t_j$  сопоставляются два момента времени  $\tau_{1j}$  и  $\tau_{2j}$ . Переход  $t_j$  может быть запущен, только если он был разрешён к моменту времени  $\tau_{1j}$ . Если он является разрешённым, то должен быть запущен до наступления момента времени  $\tau_{2j}$ . Рассмотрим временные сети более подробно.

Формально временные сети задаются набором  $(P, T, I, O, \mu, Z)$ , где  $P, T, I, O, \mu$  имеют обычный смысл, а  $Z: P \rightarrow \mathbb{R}^+$  функция времени задержки меток в позициях сети. Работа временных сетей подчиняется следующим правилам:

- метки в позициях могут быть доступными или же недоступными;
- переходы считаются возбуждёнными, если все их входные позиции имеют метки и эти метки доступны;
- переходы срабатывают мгновенно в тот самый момент, как только будут выполнены условия их возбуждения. Правила перехода меток во временных сетях совпадают с аналогичными правилами для сетей Петри;
- каждая метка, совершившая переход из  $p_i \in P$  в  $p_m \in P$ , будет недоступной в  $p_m$  в течении времени  $z_m$ , начиная с момента её появления в  $p_m$ . По истечению времени  $z_m$  метка становится доступной.

Одним из интересных применений временных сетей являются задачи анализа периодических режимов функционирования систем. В сети можно обеспечить периодический режим работы с минимальным периодом. Такой режим достигается при использовании специального расписания включения переходов сети.

Характерным для имитационных моделей формализующим фактором является применённый в них новый механизм описания состояния сети. Метки в расширенных сетях Петри являются носителями определённого количества атрибутов, в качестве которых могут выступать числа, логические переменные, текстовые конструкции, массивы, таблицы. Атрибуты меток могут быть функциями времени. Переходы меток при выполнении сети сопровождаются изменениями значений атрибутов. Эти изменения подчиняются специально определяемым в модели правилам (процедурам перехода). В расширенных сетях Петри средства описания процессов син-

хронизации событий значительно более развиты, чем рассмотренный ранее механизм блокировки меток во временных сетях.

А теперь рассмотрим конкретные варианты реализации расширенных сетей Петри.

#### 7.4.1. Временные сети событий (ВСС)

ВСС= $(P, T, E)$  – связный биограф, где  $P$  – конечное непустое множество позиций,  $T$  – конечное непустое множество переходов,  $E$  – отображение вида  $T \rightarrow \Phi(P)$ ;  $\Phi(P)$  – множество всех подмножеств  $P$ . Позиции сети  $p_i \in P$ ,  $i = 1, 2, \dots, |P|$  отображают в модели все основные состояния процессов. Переходы  $t_j \in T$ ,  $j = 1, 2, \dots, |T|$  соответствуют событиям. Отображение  $E(t_j)$  находит все смежные с  $t_j$  позиции из множества  $P$ . Состояние ВСС определяется маркировкой позиций  $M: P \rightarrow N$  – функция маркировки  $(M(p_1), \dots, M(p_{|P|}))$  – разметка сети. Применяемые в ВСС метки имеют атрибуты:  $(a_1, a_2, \dots, a_{k_i})$  – вектор атрибутов метки в позиции  $p_i \in P$ . Значения некоторых атрибутов могут быть функциями времени. Например, если позиция  $p_i$  моделирует в сети состояние “Идёт обработка детали”, то атрибут  $a_m$  в списке атрибутов метки  $(a_1, \dots, a_m, \dots, a_{k_i})$  может означать время, оставшееся до конца обработки.

Как правило,  $M(p_i) \leq 1$ . Равенству соответствует тот факт, что состояние  $p_i$  некоторого процесса в системе, представленного в модели некоторым элементарным высказыванием, реализовалось. Если  $M(p_i) > 1$ , то наиболее простыми интерпретациями в этом случае являются состояния, возникающие при накоплении объектов, участвующих в процессе. Важно указать, что при  $M(p_i) > 1$  необходимо упорядочивать множество меток в

позиции. Например, если  $p_i$  моделирует стек или очередь, то метки в позиции упорядочиваются линейно.

Механизм изменения состояний ВСС связан с выполнением переходов  $t_j \in T$ . Каждый переход можно записать в виде тройки  $(U_j, S_j, V_j)$ . Здесь  $U_j$  - условия возбуждения,  $S_j$  - схема выполнения,  $V_j$  - процедура перехода. Условие возбуждения  $U_j$  - это предикат, определённый на множестве позиций из  $E(t_j)$ , истинный только в том случае, когда реализуется некоторая заданная разметка позиций множества  $E(t_j)$ . Кроме того, условие возбуждения, может включать ещё проверку значений атрибутов меток. Если значение какого-то атрибута является функцией времени, то обеспечивается соответствующая значению данной функции временная задержка. Схема выполнения определяет изменение разметки позиций сети при срабатывании перехода. Пусть  $\bar{E}(t_j)$  - упорядоченное множество позиций, смежных с  $t_j$ . Тогда схема  $S_j$  - это вектор, элементы которого образуют биекцию с элементами множества  $\bar{E}(t_j)$ . Любое положительное целое число в  $S_j$  означает число меток, помещённых в соответствующую позицию после выполнения перехода. Целые отрицательные числа в  $S_j$  указывают число удалённых меток из соответствующих позиций. Нулевые значения компонент отмечают изменяемые позиции (каждая изменяемая позиция не обязательно имеет единственную метку и образует с переходом  $t_j$  цикл длины 2). Не изменяемые в результате срабатывания перехода позиции отмечаются в  $S_j$  символом (\*). Процедура перехода  $V_j$  представляет собой правила вычисления атрибутов (для позиций, отмеченных в  $S_j$  нулём или целым положительным числом).

Графическое изображение ВСС имеет некоторые отличительные особенности. Каждая позиция ВСС обозначается кружком, каждый переход -

барьером, позиции и переходы соединяются ориентированными или неориентированными рёбрами в соответствии с отображением  $E$  и схемой выполнения перехода  $S_j$ . Ребро не ориентируется, если в схеме  $S_j$  для соответствующей позиции указывается символ (\*). Ребро ориентируется от позиции к переходу, если в  $S_j$  для этой позиции задано отрицательное число, от перехода к позиции – если положительное число. Ребро будет двунаправленным, если задан 0.

#### 7.4.2. E-сети

E-сети применяются в качестве имитационных моделей систем, структурируемых в виде сетей событий. Формально они определяются набором из 8 элементов:

$$E = ((P, B, R), A, (I(A), O(A)), Z, V, Q, \Psi, M_0)$$

$P$  – конечное непустое множество позиций сети  $P = \{p_i\}$ ;  $B = \{b_k\} \subset P$  – множество периферийных (не внутренних) позиций;  $R = \{r_m\} \subset P$  – множество решающих позиций. Позиции  $p_i \in P \setminus R$  играют в E-сетях ту же самую роль, что и в сетях Петри. Обозначаются графически также кружком. Периферийные позиции  $b_k \in B$  используются для моделирования взаимосвязей системы с внешней средой. Решающие позиции не имеют прямых аналогов в СП. С их помощью в E-сетях обеспечивается разрешение конфликтов и синхронизация событий. С каждой решающей позицией связан некоторый список предикатов, применяемый для формального описания условий выполнения переходов. Задаются эти позиции шестиугольниками.

В каждом состоянии сети позиции могут иметь или не иметь метку. Число меток в каждой позиции  $\leq 1$  ( $M: P \rightarrow \{0,1\}$ ). Отмечаются метки жирной точкой.

Второй элемент набора - конечное непустое множество переходов  $A = \{a_n\}$ , которые задаются барьерами.

Третий элемент набора  $I(A)$ ,  $O(A)$  – множество позиций, смежных с переходами по входу (I) и выходу (O). Пары  $(p_i, a_n) \in I(A) \times A$  и  $(a_n, p_i) \in A \times O(A)$ , составленные из смежных позиций и переходов, соответствуют дугам сети.

Четвёртый элемент набора функция  $Z: A \rightarrow \mathbb{R}^+$ , с помощью которой в Е-сетях задаются значения времени выполнения переходов, т.о. имитируются временные задержки, связанные с реализацией моделируемых событий.

Пятый элемент набора  $V = \{v_s\}$  задаёт непустое конечное множество переменных – количественных оценок состояния модели.

Шестой элемент набора  $Q = \{q_m\}$  описывает множество решающих процедур, применяемых для разрешения конфликтов на переходах и синхронизации событий.

Седьмой элемент набора - множество  $\Psi = \{\psi_n\}$  процедур перехода.

Восьмой элемент набора  $M_0 = (M_0(p_1), \dots, M_0(p_{|p|}))$  обозначает начальную маркировку позиций.

При построении Е-сетей используется ограниченный набор типов переходов: T, I, F, X, Y, MX, MY (см. табл.7.1). Два последних являются макрорасширениями X и Y переходов.

Макропереходы определяются несколько сложнее, чем соответствующие схемы X и Y переходов. Однако, главным в них по-прежнему остаётся сравнение значений решающей позиции и меток инцидентных переходу дуг.

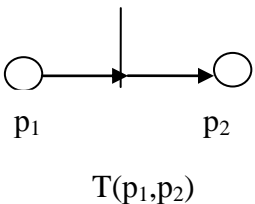
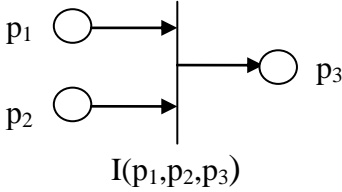
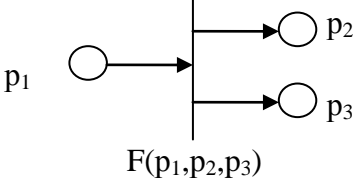
Каждый переход Е-сети имеет три характеристики и записывается как  $a_n = (\pi_n, z_n, \Psi_n)$ , где  $\pi_n \in \Pi$  - тип перехода;  $z_n \in Z$  - время перехода;  $\Psi_n$  - процедура перехода. Переход  $a_n \in A$  выполняется в 3 этапа:

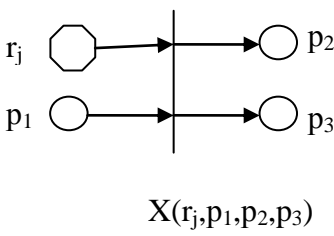
1. Проверяются условия активности перехода, а для  $X$  и  $Y$  переходов ещё и находятся значения решающей позиции и определяется конкретная схема срабатывания.

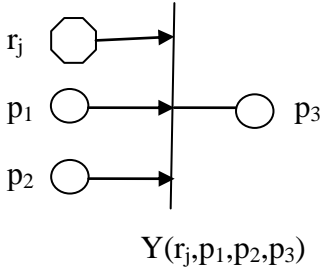
2. Реализуется задержка выполнения перехода на время  $Z$  и пересчитываются значения атрибутов метки по правилам, указанным в процедуре  $\Psi_n$ .

3. В соответствии со схемой перехода изменяются маркировки его входных и выходных позиций.

Таблица 7.1. Схемы выполнения переходов каждого типа

Графические и символ-ные изображения	Комментарий
 <p style="text-align: center;"><math>T(p_1, p_2)</math></p>	<p><math>(1,0) \rightarrow (0,1)</math> Переход активный, если <math>p_1</math> имеет метку а <math>p_2</math> – свободна. После срабатывания перехода появляется метка в <math>p_2</math>, метка из <math>p_1</math> удаляется</p>
 <p style="text-align: center;"><math>I(p_1, p_2, p_3)</math></p>	<p><math>(1,1,0) \rightarrow (0,0,1)</math>. Переход активный, если <math>p_1</math> и <math>p_2</math> имеют метки, а <math>p_3</math> – свободна. После срабатывания перехода появляется метка в <math>p_3</math>, метки из <math>p_1</math> и <math>p_2</math> удаляются</p>
 <p style="text-align: center;"><math>F(p_1, p_2, p_3)</math></p>	<p><math>(1,0,0) \rightarrow (0,1,1)</math> Переход активный, если <math>p_1</math> имеет метку, а <math>p_2</math> и <math>p_3</math> свободны. После срабатывания перехода появляются метки в <math>p_2</math> и <math>p_3</math>, из <math>p_1</math> метка удаляется</p>

 <p style="text-align: center;"><math>X(r_j, p_1, p_2, p_3)</math></p>	$(0,1,0,0) \rightarrow (*,0,1,0)$ $(1,1,0,0) \rightarrow (*,0,0,1)$ $(0,1,1,0) \rightarrow (*,0,1,1)$ $(0,1,0,1) \rightarrow (*,0,1,1)$ $(1,1,1,0) \rightarrow (*,0,1,1)$ $(1,1,0,1) \rightarrow (*,0,1,1)$	<p>Если <math>p_2</math> и <math>p_3</math> свободны, а в <math>p_1</math> имеется метка, то на переходе возникает состояние конфликта. Конфликт исключается по значению решающей позиции: если <math>M(r_j) = 0</math>, то метка “переходит” из <math>p_1</math> в <math>p_2</math>; если <math>M(r_j) = 1</math>, то из <math>p_1</math> в <math>p_3</math>. Значение <math>M(r_j)</math> находится в результате применения решающей</p>
<p>процедуры <math>q(r_j) \in Q</math>. Её выполнение заключается в проверке значений предикатов <math>pr1(r_j)</math> и <math>pr2(r_j)</math>. При истинности какого-то одного из них <math>M(r_j) = 1</math>, при истинности другого <math>M(r_j) = 0</math>. Если оба ложные, то <math>M(r_j) = \infty</math>. В этом случае решающая позиция блокирует переход до момента результативного выполнения процедуры <math>q(r_j)</math>. (* означает, что необходимо снова определить значение <math>M(r_j)</math>). Если на переходе конфликта нет (свободна только какая-то одна выходная позиция) и имеется метка в позиции <math>p_1</math>, то переход будет выполняться по T-схеме и по единственному возможному пути, независимо от того, какое значение имеет при этом решающая позиция</p>		

 <p style="text-align: center;"><math>Y(r_j, p_1, p_2, p_3)</math></p>	$(0,1,1,0) \rightarrow (*,0,1,1)$ $(1,1,1,0) \rightarrow (*,1,0,1)$ $(0,1,0,0) \rightarrow (*,0,0,1)$ $(0,0,1,0) \rightarrow (*,0,0,1)$ $(1,1,0,0) \rightarrow (*,0,0,1)$ $(1,0,1,0) \rightarrow (*,0,0,1)$	Аналогично вышеописанному
---	--	---------------------------

### 7.4.3. Комби-сети

Наиболее полный набор средств описания дискретных и дискретно-непрерывных систем в виде сетей событий применяется при построении Комби-сетей. Множество  $P$  позиций  $K$ -сетей объединяет 6 подмножеств:  $P_э$  – элементарные,  $P_т$  – с временной задержкой,  $P_д$  – долгоживущие,  $P_г$  – гибридные,  $P_к$  – комплексные,  $P_м$  – макропозиции. Каждой позиции отвечает некоторый характерный для данной позиции набор атрибутивных переменных  $ATTR(p_i)$ ,  $p_i \in P$ . Позиции сети могут не иметь маркеров, иметь единственный маркер, накапливать маркеры ( $M: P \rightarrow N$ ). При выполнении сети общее число действующих в ней маркеров изменяется и в каждый текущий момент времени равно  $MA$ . Всё множество  $MA$  маркеров разбивается на 4 класса: булевские, с атрибутивными переменными, долгоживущие, изменяющие вид с элементарного на долгоживущие и обратно. С двумя первыми классами связаны элементарные маркировки (запрещены для позиций  $P_д$ ), третьему и четвёртому классам соотносятся долгоживущие маркировки (запрещены для позиций  $P_э, P_т, P_м$ ). Каждый класс МК использует определённую структуру данных, все маркеры класса МК, за ис-

ключением булевского класса, имеют множество атрибутивных переменных  $ATTR(MK)$ .

Множество  $T$  переходов  $K$ -сети состоит из 5-ти подмножеств:  $T_э$  - элементарных,  $T_{пр}$  - с прерыванием,  $T_д$  - долгоживущих,  $T_к$  - комплексных,  $T_м$  - макропереходов. Каждый  $t_j \in T$  характеризуется набором атрибутивных переменных  $ATTR(t_j)$ .

Объединение всех наборов атрибутивных переменных сети образует множество  $ATTR$  её локальных переменных.

Связи  $E$  между позициями и переходами в  $K$ -сетях задаются обычным для сетей Петри способом:  $E \subseteq (P \times T) \cup (T \times P)$ .

Необходимые условия возбуждения переходов описываются логическими выражениями. Переход  $t_j \in T$  может перейти в возбуждённое состояние, если соответствующее этому переходу логическое выражение будет истинным. Другое необходимое условие возбуждения связано с текущей разметкой входных и выходных позиций перехода. В тех случаях, когда количество маркеров в  $p_i \in I(t_j)$  больше, чем это необходимо для возбуждения перехода, возникает вопрос о механизме выбора удаляемых маркеров. Такими механизмами могут быть: FIFO, LIFO, HVF, LVF, согласно заданным приоритетам или по указанию пользователя.

После перехода в позицию  $p_i \in P_T \subset P$  маркер становится недоступным в ней на время задержки  $\tau(p_i)$ . Значения  $\tau(p_i)$  могут быть определены различными способами: выбор случайной величины с заданным законом распределения, назначение констант из множества неотрицательных целых чисел, выбором по указанию пользователя и т.д.

Если  $|O(t_j)| > 1$ , необходимо указать конкретную позицию  $p_i \in O(t_j)$ , в которую переходят маркеры. Могут быть различные механизмы: равноверо-

ятный по тестовым условиям, по значениям атрибутивных переменных маркеров или по указанию пользователя.

Ещё одна особенность выполнения К-сетей обусловлена структурой её связей: в позициях  $p_i \in P \setminus P_d$ ,  $|O(p_i)| > 1$  возможны конфликтные ситуации, связанные с необходимостью выбора какого-то одного из нескольких возбуждённых переходов, смежных с  $p_i$ . Для исключения конфликтов этого вида в К-сетях применяются специальные механизмы случайного выбора, выбора по номеру перехода или по вероятности и другие. Большое разнообразие возможных вариантов построения схемы переходов в К-сетях обусловило разработку основного набора таких схем, а также способов построения составных схем, создаваемых на базе элементов из основного набора.

### 7.5. Вопросы для самопроверки

1. Назовите основные способы задания сетей Петри.
2. Сформулируйте правило выполнения сетей Петри.
3. Что такое непримитивное событие и каким образом оно сводится к примитивным событиям?
4. Перечислите основные задачи анализа сетей Петри.
5. Какие существуют методы анализа сетей Петри?
6. Какие задачи анализа сетей Петри нельзя решить с помощью построения дерева достижимости?
7. Назовите направления развития базового математического аппарата сетей Петри. Приведите примеры реализаций этих направлений.

## ГЛАВА 8. МОДЕЛИРОВАНИЕ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ МЕТОДОМ СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ

### 8.1. Основные положения

Ранее мы познакомились с методами построения математических моделей, дающих возможность установить аналитическую (формульную) связь между заданными условиями операции и результатом (исходом) операции, характеризующимся одним или несколькими параметрами – показателями эффективности. Если в ход операции вмешиваются случайные факторы, то она представляет собой случайный процесс, а показатель эффективности – вероятность какого-то события или же математическое ожидание какой-то случайной величины. Иногда удастся построить аналитическую модель случайного процесса (например, систему дифференциальных уравнений для вероятностей состояния или алгебраических уравнений для предельных вероятностей состояния) и связать заданные условия операции с ее исходом аналитическими зависимостями. Однако, это удастся далеко не всегда - главным образом, в тех случаях, когда случайный процесс, протекающий в рассматриваемой системе, марковский или близок к марковскому.

Для произвольных потоков событий, переводящих систему из состояния в состояние, аналитические решения получены только для отдельных частных случаев, а в общем случае удовлетворительных методов математического описания соответствующих процессов не существует.

В тех случаях, когда построение аналитической модели явления по той или иной причине трудно осуществимо, применяется другой метод моделирования, известный под названием метода статистических испытаний или, иначе, метода Монте-Карло.

Существо метода. Вместо того, чтобы описывать случайное явление с помощью аналитических зависимостей, производится «розыгрыш» - моделирование случайного явления с помощью некоторой процедуры, дающей случайный результат. Так же как в жизни, когда конкретное осуществление процесса складывается каждый раз по иному, так и в результате «розыгрыша» мы получаем один экземпляр – одну реализацию случайного явления. Произведя такой «розыгрыш» очень большое число раз, мы получим статистический материал – множество реализаций случайного явления – который можно обработать обычными методами математической статистики.

Нередко такой прием оказывается проще, чем попытки построить аналитическую модель явления и исследовать зависимость между его параметрами на этой модели. Для сложных операций, в которых участвует большое число элементов (машин, систем, людей, коллективов) и в которых случайные факторы сложным образом взаимодействуют между собой, метод статистических испытаний, как правило, оказывается проще аналитического.

В сущности, методом «розыгрыша» может быть решена любая вероятностная задача: однако оправданным он становится только в случае, когда процедура «розыгрыша» проще, а не сложнее применения аналитических методов.

Пример 1. Решается задача: по некоторой цели производится 4 независимых выстрела, каждый из которых попадает в нее с  $p=0,5$ . Для поражения цели требуется не менее двух попаданий. Определить вероятность поражения цели.

Аналитический способ. Вероятность поражения цели вычисляется через вероятность противоположного события. Вероятность непоражения цели = сумме вероятностей ни одного попадания и ровно одного попада-

ния; вероятность ни одного попадания  $0,5^4$ ; вероятность одного попадания равна:

$$C_4^1 \cdot 0,5^1 \cdot 0,5^3 = 4 \cdot 0,5^4, \text{ следовательно}$$

$$W = 1 - (0,5^4 + 4 \cdot 0,5^4) \approx 0,688.$$

Розыгрыш. Будем моделировать процедуру стрельбы с помощью другой, тоже случайной процедуры. Будем бросать 4 монеты: условимся - герб попадание, решка – промах. Цель поражена - не менее двух гербов. Розыгрыш в нашем случае – бросание четырех монет, результат этого опыта - «поражение или не поражение цели». Повторим такой опыт (бросание 4 монет) очень много раз подряд. Тогда согласно теореме Бернулли, частота поражения цели почти наверняка будет мало отличаться от вероятности этого события  $W$ ; значит, если мы бросим четыре монеты большое число раз  $N$ , мы почти наверняка получим число близкое к  $W$ , т.е. к  $0,688$ . В данном случае определение вероятности  $W$  розыгрышем было несравненно трудней, чем аналитическим расчетом.

Пример 2. Анализ поведения дискретного объекта (дискретные входные и выходные переменные) - «задача о пьяном прохожем или задача о случайном блуждании». Прохожий решил прогуляться, стоя на углу улиц. Пусть вероятность того, что, достигнув очередного перекрестка, он пойдет на север, юг, восток и запад, одинакова. Какова вероятность того, что пройдя 10 кварталов, прохожий окажется не далее 2 кварталов от места, где он начал прогулку.

Обозначим его местонахождение на каждом перекрестке двумерным вектором  $(x_1, x_2)$  («выход»), где  $x_1$  – направление с востока на запад и  $x_2$  – направление с севера на юг. Каждое перемещение на один квартал к востоку  $(x_1 + 1)$ , а каждое перемещение на один квартал к западу  $(x_1 - 1)$  ( $x_1$  - дискретная переменная). К северу  $x_2 + 1$ , к югу  $x_2 - 1$ . Начальное положение  $(0,0)$ .

Если в конце прогулки абсолютные значения  $x_1$  и  $x_2$  будут больше 2, то будем считать, что он ушел дальше двух кварталов в конце прогулки протяженностью в 10 кварталов. Т.к. вероятность движения нашего прохожего в любом из 4 направлений по условию одинакова и равна 0,25, то можно оценить его передвижение с помощью таблицы случайных чисел. Условимся, что если случайное число (СЧ) лежит в пределах от 0 до 24, пьяный пойдет на восток и мы увеличим  $x_1$  на 1; если от 25 до 49, то он пойдет на запад и  $x_1 - 1$ ; если от 50 до 74, он пойдет на север и  $x_2 + 1$ ; если от 75 до 99, то на юг и  $x_2 - 1$ . Блок - схема соответствующей программы приведена на рис. 8.1.

Нужно провести достаточно большое число «машинных опытов», чтобы получить достоверный результат. Другими методами такую задачу решить практически невозможно.

В литературе этот метод получил название метода имитационного моделирования, а также машинного, статистического, вероятностного, Монте-Карло или метода машинной имитации. В принципе имитационное моделирование можно осуществлять на широком спектре устройств, начиная с аналоговых ЭВМ и кончая листом бумаги с карандашом, однако, как правило, ориентируются на использование ЭВМ.

Метод имитационного моделирования может рассматриваться как своеобразный экспериментальный метод. Отличие от обычного эксперимента заключается в том, что в качестве объекта экспериментирования выступает имитационная модель, реализованная в виде программы на ЭВМ. При таком экспериментировании с моделью (в отличие от «решения» модели при аналитическом, например, моделировании) могут быть применены статистические методы.

Заметим, что методом статистических испытаний можно находить не только вероятности событий, но и средние значения (математические ожи-

дания) случайных величин. При этом применяется закон больших чисел (теорема Чебышева). Согласно этой теореме, при большом числе опытов среднее арифметическое наблюдаемых значений случайной величины почти наверняка мало отличается от её математического ожидания. Аналогичным образом могут быть найдены не только математические ожидания, но и дисперсии интересующих нас случайных величин.

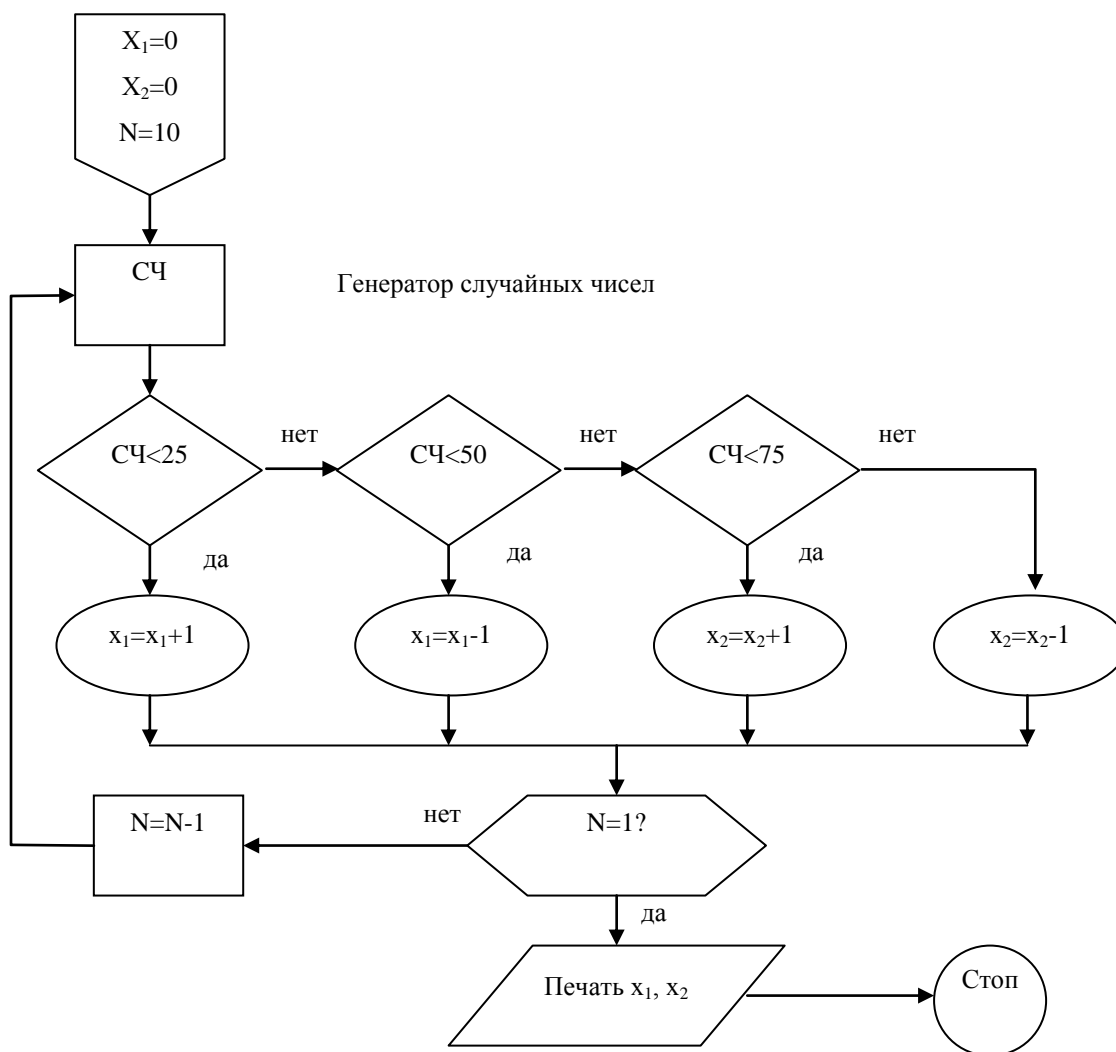


Рис. 8.1. Блок-схема программы имитации поведения прохожего

Метод Монте-Карло есть метод математического моделирования случайных явлений, в которых сама случайность непосредственно включается в процесс моделирования и представляет собой его существенный элемент.

Каждый раз, когда в ход операции вмешивается тот или другой случайный фактор, его влияние имитируется с помощью специально организованного «розыгрыша» или жребия. Таким образом, строится одна реализация случайного явления, представляющая собой как бы результат одного «опыта». При большом числе реализаций интересующие нас характеристики случайного явления (вероятности, математические ожидания) находятся так же, как они находятся из опыта.

Моделирование случайных явлений методом Монте-Карло имеет общие черты с процессом набора опыта отдельными людьми и человеческими коллективами. И тут, и там каждая отдельная реализация случайна; устойчивые закономерности обнаруживаются лишь при многократном наблюдении явления, при обширном опыте. Большое число реализаций, требующееся при применении метода Монте-Карло, делает его вообще громоздким и трудоемким.

Метод имитационного моделирования свободен от каких-либо ограничений на класс решаемых задач, нагляден, прост в освоении, обладает повышенной устойчивостью к сбоям ЭЦВМ; все промежуточные результаты легко интерпретируются и контролируются.

Недостатками имитационного моделирования являются:

- большой расход машинного времени;
- малая точность вероятностных характеристик редких событий;
- трудность получения обобщающих выводов и рекомендаций;
- сложность оптимизации системы (поиск оптимума требует многовариантных расчетов и ведется при наличии вероятностных помех – случайных ошибок в результатах);
- вероятностная оценка погрешностей (впрочем, для вероятностных задач такая оценка вполне естественна).

Перечисленные недостатки имитационного моделирования несколько смягчаются постоянным ростом технических характеристик современных ЭВМ и использование специальных методов.

По указанным причинам применение имитационного моделирования (особенно в квалификационных работах) нужно сводить к разумному минимуму. Такое применение представляется целесообразным:

- для накопления первичных данных об изучаемом явлении, если эти данные нельзя получить в натурном эксперименте;
- для проверки правомерности допущений, сделанных разработчиком в целях перехода к аналитическим методам;
- для демонстрации конечных результатов исследования на достаточно полной модели реальной ситуации;
- при «безысходности», когда сложность ситуации намного превосходит возможности аналитических методов, известных разработчику.

## **8.2. Способы организации единичного жребия**

Основным элементом, из совокупности которых складывается монте-карловская модель, является одна случайная реализация моделируемого явления, например: один «обстрел» цели», один «день работы» транспорта, одна «эпидемия» и т.п.

Реализация представляет собой как бы один случай осуществления моделируемого случайного явления (процесса) со всеми присущими ему случайностями. Она разыгрывается с помощью специально разработанной процедуры или алгоритма, в котором важную роль играет собственно «розыгрыш» или бросание жребия». Каждый раз, когда в ход моделируемого процесса вмешивается случайность, её влияние учитывается не расчетом, а бросанием жребия.

Предположим, что в ходе моделируемого процесса наступил момент, когда его дальнейшее развитие (а значит и результат) зависит от того, появилось ли на данном этапе событие А или не появилось (например: произошло ли попадание в цель, обнаружен ли некоторый объект, исправна ли некоторая аппаратура и т.д). Тогда нужно «бросанием жребия» решить во-

прос: появилось событие  $A$  или не появилось? Для этого нужно привести в действие некоторый случайный механизм розыгрыша (бросить игральную кость, несколько монет или выбрать число из таблицы случайных чисел) и условиться о том, какой результат жребия означает появление, а какой – непоявление события  $A$ ). Ниже мы увидим, что розыгрыш всегда можно организовать так, чтобы событие  $A$  имело любую наперед заданную вероятность. Кроме событий, появляющихся случайным образом, на ход и исход операции могут так же влиять разные случайные величины (время, координаты и т.д.). С помощью жребия можно разыграть значения любой случайной величины или совокупность значений нескольких случайных величин. Условимся называть единичным жребием любой элементарный опыт, в котором решается один из вопросов:

1. Произошло или не произошло событие  $A$ ?
2. Какое из возможных событий  $A_1, A_2, \dots, A_k$  произошло?
3. Какое значение приняла случайная величина  $X$ ?
4. Какую совокупность значений приняла система случайных величин  $X_1, X_2, \dots, X_k$ ?

Рассмотрим способы организации всех разновидностей единичного жребия. При любой организации жребия должен быть пущен в ход какой-то механизм случайного выбора. Механизмы могут быть самыми разнообразными, однако любой из них может быть заменен стандартным механизмом, позволяющим решить одну задачу: получить случайную величину, распределенную с постоянной плотностью от 0 до 1. Условимся для краткости называть такую случайную величину «случайное число от 0 до 1» и обозначать  $R$ .

1. *Появилось или нет событие  $A$ ?*

Пусть вероятность события  $A$  равна  $p$ :  $P(A)=p$ . Выберем с помощью стандартного механизма случайное число  $R$  и будем считать, что если оно

меньше  $p$ , событие  $A$  произошло, если больше  $p$  – не произошло. Действительно: если  $R$  – случайное число от 0 до 1, то  $P(R < p) = \int_0^p f(r)dr$ , где  $f(r)=1$  при  $0 < r < 1$  или

$$P(R < p) = \int_0^p dr = p = P(A).$$

## 2. Какое из нескольких возможных событий появилось?

Пусть имеется полная группа несовместных событий:  $A_1, A_2, \dots, A_k$  с вероятностями  $p_1, p_2, \dots, p_k$ . Т.к. события несовместны и образуют полную группу, то  $p_1 + p_2 + \dots + p_k = 1$ . Разделим весь интервал от 0 до 1 на  $k$  участков длиной  $p_1, p_2, \dots, p_k$  (рис. 8.2).

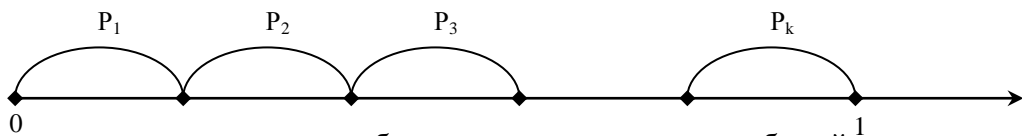


Рис. 8.2. Организация единичного жребия «какое из возможных событий произошло»

Если случайное число  $R$ , выданное стандартным механизмом, попало, например, на участок  $p_3$ , это означает, что появилось событие  $A_3$ .

## 3. Какое значение приняла случайная величина?

Пусть нам требуется «разыграть» значение случайной величины  $X$ , имеющей известный закон распределения. Случай, когда величина  $X$  дискретна (т.е. имеет отдельные значения  $x_1, x_2, \dots, x_k$  с вероятностями  $p_1, p_2, \dots, p_k$ ) рассматривать не будем, т.к. он сводится к предыдущему пункту 2. Рассмотрим случай, когда случайная величина  $X$  непрерывна и имеет заданную непрерывную функцию распределения  $F(x)$  (рис. 8.3).

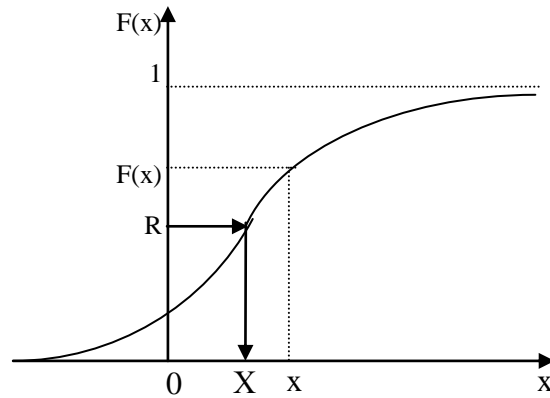


Рис. 8.3. Случайная величина  $X$  с непрерывной функцией распределения  $F(x)$

Докажем следующее утверждение: если взять на оси ординат случайное число  $R$  (от 0 до 1) и найти то значение  $X$ , при котором  $F(x)=R$ , то полученная случайная величина  $X$  будет иметь функцию распределения  $F(x)$ .

Действительно, возьмем случайную величину  $X$  и найдем её функцию распределения, т.е. вероятность  $P(X < x)$ . Из рисунка видно, что для того, чтобы выполнялось неравенство  $X < x$ , величина  $R$  должна принять значение, меньшее, чем  $F(x)$ .  $P(X < x) = P(R < F(x))$ . Но случайное число  $R$  имеет постоянную плотность распределения  $f(r)$ , равную 1 на отрезке  $(0,1)$ ; значит

$$P(X < x) = \int_0^{F(x)} f(r) dr = \int_0^{F(x)} 1 \cdot dr = F(x), \text{ что и требовалось доказать.}$$

Т.о. розыгрыш значения случайной величины  $X$  с заданной функцией распределения  $F(x)$  сводится к следующей процедуре: получить случайное число  $R$  от 0 до 1 и в качестве значения  $X$  взять  $X=F^{-1}(R)$ , где  $F^{-1}$  – функция, обратная по отношению к  $F$ .

Пример 1. Случайная величина  $X$  распределена по показательному закону с плотностью  $f(x)=\lambda e^{-\lambda x}$  ( $x>0$ ). Построить процедуру единичного жребия для получения значения  $X$ .

По заданной плотности  $f(x)$  находим функцию распределения:

$$F(x) = \int_0^x f(x)dx = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x} \quad (x > 0).$$

График  $F(x)$  дан на рис. 8.4.

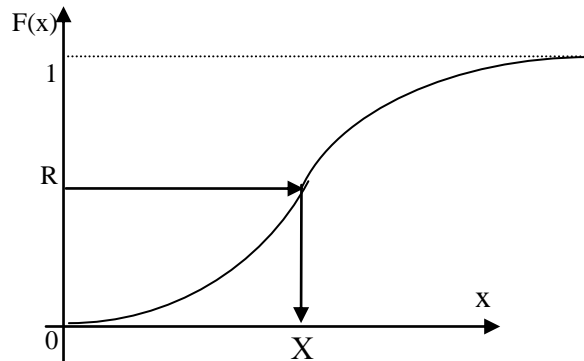


Рис. 8.4. Случайная величина  $X$  с показательной функцией распределения

Графически значение случайной величины  $X$  можно разыграть так: взять случайное число от 0 до 1 на оси ординат и найти соответствующее ему значение абсциссы  $X$ . Это же можно сделать расчетом, если написать:  $R = 1 - e^{-\lambda X}$  и решать это уравнение относительно  $X$  (т.е. найти обратную по отношению к  $F$  функцию). Имеем  $e^{-\lambda X} = 1 - R$ ,

$$-\lambda X = \ln(1 - R) \Rightarrow X = -\frac{1}{\lambda} \ln(1 - R).$$

Эту формулу можно упростить: вспомним, что если  $R$  – случайное число от 0 до 1, то  $(1 - R)$  – также случайное число от 0 до 1, поэтому можно взять  $X = -\frac{1}{\lambda} \ln(R)$ . Т.о. процедура розыгрыша сводится к следующему: взять случайное число от 0 до 1, прологарифмировать его при натуральном основании, изменить знак и разделить на  $\lambda$ .

Пример 2. Розыгрыш значения случайной величины, распределенной по нормальному закону (короче – «нормальной») с математическим ожиданием  $m_x$  и средним квадратичным отклонением  $\sigma_x$ . Плотность распределения случайной величины  $X$  имеет вид:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} .$$

Удобнее применить не общее правило, а поступить иначе: перейти от  $X$  к другой (т.н. «нормированной») случайной величине  $Z = \frac{x-m_x}{\sigma_x}$ , разыграть значение этой величины, а затем уже по ней найти  $X$ . Это удобно потому, что  $m_z=0$ ,  $\sigma_z=1$  и придется только один раз и навсегда найти обратную функцию. Легко показать, что значение нормальной случайной величины  $X$  с характеристиками  $m_x, \sigma_x$  разыгрывается по формуле:

$$X = \sigma_x \Phi^{-1}(R - 0.5) + m_x ,$$

где  $\Phi^{-1}$  – функция, обратная функции Лапласа. Есть и другой способ, основанный на центральной предельной теореме теории вероятностей. Согласно этой теореме, при сложении достаточно большого числа независимых случайных величин, сравнимых по своим дисперсиям, получается случайная величина, распределенная приближенно по нормальному закону, причем этот закон тем ближе к нормальному, чем больше случайных величин складывается (для большинства прикладных задач достаточно складывать 6 случайных величин от 0 до 1). В результате получается следующая процедура

$$X = \sigma_x \sqrt{2} \left( \sum_{i=1}^6 R_i - 3 \right) + m_x .$$

#### 4. Какую совокупность значений примет система случайных величин?

Пусть имеется система случайных величин:  $X_1, X_2, \dots, X_n$  с совместной плотностью распределения  $f(x_1, x_2, \dots, x_n)$ . Если случайные величины независимы, то  $f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n)$  и розыгрыш совокупности значений системы  $x_1, x_2, \dots, x_n$  сводится к тому, чтобы разыграть каждую из них в отдельности, т.е. организовать  $n$  единичных жребиев типа, описанно-

го в п.3. Если случайные величины зависимы, то  $f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2/x_1)f_3(x_3/x_1x_2)\dots$ , где каждая последующая плотность распределения берется условная, при условии, что предыдущие случайные величины приняли определенные значения. При розыгрыше последовательности значений случайных величин получается сначала значение  $x_1$  случайной величины  $X_1$ ; это значение берется в качестве аргумента в условной плотности  $f_2(x_2/x_1)$ ; разыгрывается значение  $x_2$  случайной величины  $X_2$ , оба значения  $x_1, x_2$  берутся в качестве аргументов в условной плотности  $f_3(x_3/x_1x_2)$  и т.д.

### **8.3. Приемы построения и эксплуатации имитационных моделей**

При создании имитационных моделей в настоящее время используется два подхода: дискретный и непрерывный. Выбор подхода в значительной мере определяется свойствами объекта-оригинала и характером воздействия на него внешней среды. Метод статистического моделирования (метод Монте-Карло) – можно рассматривать как частный случай дискретных вероятностных имитационных моделей. При использовании дискретного подхода к созданию имитационных моделей обычно применяются абстрактные системы (математические схемы) трех основных типов: автоматные системы, системы массового обслуживания и агрегативные системы. В случае непрерывного подхода моделируемый объект независимо от его природы формализуется в виде непрерывной абстрактной системы, между элементами которой циркулируют потоки той или иной природы. Структура такой системы представляется графически в виде диаграммы (схемы) потоков. Основными элементами непрерывной системы рассматриваемого

типа являются абстрактные «бункеры» (емкости, резервуары), а также элементы задержки.

Под имитационной моделью объекта – оригинала, в общем случае, мы можем понимать определенную систему, состоящую из отдельных подсистем (элементов, компонентов) и связей между ними, причем функционирование (изменение состояний) и внутреннее изменение всех элементов модели под действием связей может быть алгоритмизировано тем или иным образом, так же как и взаимодействие системы с внешней средой.

Тогда имитация функционирования системы сводится к пошаговому воспроизведению на ЭВМ процесса функционирования всех её элементов с учетом их взаимодействия и воздействий внешней среды. В имитационном моделировании могут быть алгоритмизированы и воспроизведены процессы функционирования и взаимодействия самых различных элементов абстрактной системы - дискретных и непрерывных, вероятностных и детерминированных, выполняющих функцию обслуживания, задержки и других. В качестве имитационной модели объекта при такой постановке выступает программа на ЭВМ (вместе с обслуживающими, сервисными программами), написанная либо на универсальных языках высокого уровня, либо с применением специализированных языков имитационного моделирования. До настоящего времени разработка и применение имитационных моделей все ещё в большей степени искусство, чем наука. Имитационное моделирование наиболее эффективно использовать на высоких уровнях иерархии, при рассмотрении взаимодействия во времени большого числа сложных объектов, что характерно для высшей степени технологии и производственных процессов.

### 8.3.1. Технология дискретной имитации

Цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы на основе результатов анализа наиболее существенных взаимосвязей между её элементами. Результаты исследования имитационной модели, как правило, представляют собой оценки функциональных характеристик той системы, поведение которой имитируется. Так, например, при имитационном моделировании любой СМО практический интерес могут представить такие показатели, как средняя продолжительность обслуживания заявки, средняя длина очереди, доля времени простоя и т.д.

Первый шаг к созданию имитационной модели состоит в описании реально существующей системы с использованием характеристик основных событий. *Событие* определяется как точка во времени, в которой происходят изменения характеристик системы. Обычно изменения имеют место в тех случаях, когда кончается один процесс (или несколько процессов) и начинаются другие. Для получения требуемых результатов моделирования достаточно наблюдать систему в те моменты, когда происходят события.

Для иллюстрации рассмотрим пример СМО с одним каналом (СМО с ожиданием). Оценка характера функционирования: среднее время пребывания заявки в очереди, средняя длина очереди и доля времени простоя системы. Эти характеристики могут менять свои значения либо в момент поступления дополнительного требования на обслуживание, либо при завершении обслуживания (возможны различные ситуации). Можно получить необходимую информацию, наблюдая различные условия, которые возникают при наступлении того или иного события.

Для эксплуатации любой имитационной модели необходимо выбрать единицу времени. В зависимости от природы моделируемой системы такой

единицей может быть минута, месяц и т.п. (для аэропорта крупного города – минута, небольшого города – час).

Допустим, что надо моделировать работу системы в течение  $T$  единиц времени. Работа начинается с данными, относящимися к нулевому моменту времени и отмечаются соответствующие события на шкале времени в хронологическом порядке (рис. 8.5). Таким образом, модель функционирует, перепрыгивая от одного события к другому, непосредственно за ним следующему. Каждое событие сопровождается корректировкой протокола, отражающей возможные изменения в показателях функционирования.

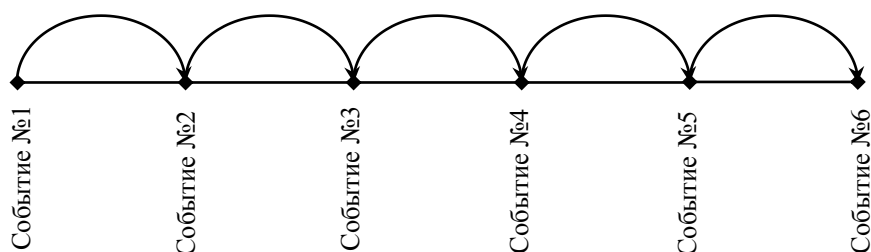


Рис. 8.5. Процесс эксплуатации дискретной модели

Резкие переходы (скачки), совершаемые моделью при переходе от одного события к другому, указывают на то, что процесс протекает в дискретном времени, откуда появилось название «дискретное моделирование».

В случае дискретного моделирования между реальным временем и временем работы модели нет ничего общего (время функционирования модели обычно значительно меньше реального).

Имитационная модель образуется взаимодействием следующих элементов:

- состояний;
- событий;
- датчиков случайных чисел;
- таймера;
- цепей событий;

- цели моделирования;
- счетчиков;
- блока инициализации;
- критерия остановки;
- методов обработки результатов.

Состояние системы должно быть определено со степенью детальности, необходимой и достаточной для вероятностного продолжения процесса моделирования. В частности, состояние системы массового обслуживания задается текущим числом заявок в ней, фазами текущего обслуживания (прибытия) и моментами наступления ближайших событий каждого вида.

Под событием модели понимается скачкообразное изменение ее состояния. События могут быть первичными (прибытие заявки, завершение обслуживания) и вторичными (по отношению к прибытию – прием заявки на обслуживание, продвижение очереди и т.п.), которые наступают как следствие первичных.

С помощью датчика случайных чисел (ДСЧ) в модели формируются ее очередные состояния (моменты наступления следующих первичных событий каждого вида и т.д.). Случайные величины генерируются в соответствии с заданными распределениями.

Имитируемый процесс развивается в модельном (системном) времени. Счетчик модельного времени называется таймером. Наиболее сложные процессы моделируются упрощенно: с постоянным шагом по оси времени. При этом отдельные стороны процесса (прибытие заявок, обслуживание) моделируются «суммарно» - через распределение числа событий данного рода за выбранный интервал. Подобное построение моделирующего алгоритма используется также при моделировании так называемых непрерывных систем, то есть систем, описываемых дифференциальными уравнениями (обыкновенных или в частных производных), когда невозможно предварительное выделение узловых событий, определяющих динамику системы, и приходится отслеживать всю траекторию.

Для наших целей наибольший интерес представляют чисто дискретные модели, состояние которых меняются скачкообразно при наступлении событий. Функционирование любого процесса разбивается на этапы (активные фазы), каждый из которых соответствует некоторому событию и реализуется в один момент системного времени. Между смежными активными фазами находится пассивная, в которой с данным процессом ничего не происходит, но может произойти любое число событий других

процессов. Событие может изменить значение атрибута, создать или уничтожить сущность, начать или прекратить активность. Моделирование требует программы, которая выстраивает последовательность событий в их взаимной зависимости.

Логика модели реализуется в процессе обработки цепей событий. В цепи текущих событий (ЦТС) находятся события, которые наступают в один момент модельного времени (уход из системы обслуженной заявки, продвижение очереди, выборка на обслуживание головной заявки очереди, формирование для нее момента завершения обслуживания). Ясно, что последовательность их обработки должна быть строго определенной. В цепи будущих событий (ЦБС) находятся события, запланированные посредством ДСЧ на последующие моменты системного времени (завершение обслуживания в других каналах, прибытие очередных заявок различных типов, поломка обслуживающего устройства, восстановление ранее отказавшего, уход из канала либо очереди нетерпеливой заявки и т.п.). В цепи задержанных событий находятся события, развитие которых заблокировано сложившимися в системе на данный момент модельного времени условиями (например, занятостью необходимых ресурсов). Могут использоваться и другие цепи событий.

Специализированные программные системы моделирования дискретных событий обычно содержат встроенные средства ведения цепей. Поэтому они позволяют программировать модели в потоковом (процессном) стиле, описывая траектории прохождения заявок через устройства системы и указывая узловые точки для сбора статистики. При работе на языках общего назначения организация и обработка цепей событий обеспечивается программистом.

Под инициализацией понимается приведение модели до начала прогона в исходное состояние. Простейший аспект инициализации – обнуление всех накапливающих счетчиков. Для обеспечения воспроизводимости результатов может потребоваться установка в исходное состояние генераторов псевдослучайных чисел.

Цель моделирования при построении модели трактуется в узком смысле – как определение показателей качества функционирования системы. Выбор цели существенно влияет на структуру модели через счетчики, необходимые для накопления результатов моделирования (например, для подсчета среднего времени ожидания начала обслуживания нужно иметь счетчики суммарного времени ожидания и количества выбранных на обслуживание заявок). Упомянутые показатели обычно связываются с некоторыми стационарными (установившимися при устремлении к бесконечности системного вре-

мени) характеристиками. Данные, накопленные за время переходного к такому режиму процесса, будут вносить погрешность в окончательные результаты. Инициализация модели, претендующей на повышенную точность, должна либо сразу задать начальные условия, близкие к стационарному режиму, либо обеспечить отбрасывание статистики переходного процесса.

Критерий останова определяет момент прекращения прогона модели. В простейшем случае прогон прекращается по достижению заданного значения таймера, счетчика числа обслуженных заявок и т.п. Однако правильнее управлять прогоном по достижению заданной точности одного из определяемых показателей. Обоснованный выбор правила остановки моделирования нетривиален, так как на этапе планирования эксперимента ни оценка определяемой величины, ни тем более ее дисперсия как правило не известны.

Обработка результатов моделирования состоит в сжатии получаемой информации, вычислении статистических (точечных и интервальных) оценок типа математических ожиданий и высших моментов для искомых показателей, оценке статистической значимости различия средних, построении гистограмм и статистических функций распределения. Другим аспектом обработки результатов является выдача их на печать в компактной и удобной для дальнейшего анализа форме. Соответственно в систему моделирования должны включаться необходимые подпрограммы – как правило широкого применения, слабо связанные со спецификой модели.

Пример. Пусть мы хотим моделировать работу одноканальной СМО с ожиданием. Поступление требований в этой системе подчинено пуассоновскому распределению со средним 3 клиента/час, а время обслуживания равно 0,2 часа с вероятностью 0,5 или 0,6 часа с вероятностью 0,5. Клиенты обслуживаются согласно дисциплине «первым пришел - первым обслуживается», длина очереди, а также источник поступления клиентов не ограничены. Предположим, что в начальный момент моделирования клиентов нет. Для пуассоновского входного потока промежутки времени между требованиями имеют экспоненциальное распределение и могут быть получены из формулы:

$$p = -(1/\lambda) \ln R = -(1/3) \ln R .$$

Время обслуживания определяется как:

$$q = \begin{cases} 0,2ч & \text{при } 0 \leq R \leq 0,5 \\ 0,6ч & \text{при } 0,5 < R \leq 1 \end{cases}.$$

В одноканальной системе возможны события только двух типов: поступление клиентов и их уход (окончание обслуживания). Действия, вызываемые этими событиями можно охарактеризовать следующим образом.

*Событие, связанное с поступлением клиента.*

1. Генерация момента времени, в который поступает следующее требование на обслуживание путем вычисления промежутка времени между требованиями  $r$  и добавления его к текущему времени моделирования (это действие необходимо для непрерывности процесса моделирования).

2. Проверка состояния системы (простоя или работа).

а) Если система простаивает, то начать обслуживание поступившего клиента, сгенерировать время обслуживания  $q$  и вычислить время окончания обслуживания (текущее время +  $q$ ); изменить состояние системы на рабочее и скорректировать протокол простоя системы.

б) Если система работает, поставить поступившего клиента в очередь и увеличить её длину на 1.

*Событие, связанное с окончанием обслуживания.*

Проверка состояния очереди (пустая или непустая).

а) Если очередь пуста, объявить простой системы.

б) Если очередь не пуста, то начать обслуживание первого по очереди клиента, уменьшить длину очереди на 1 и скорректировать протокол времени ожидания;

получить время обслуживания клиента  $q$  и вычислить время окончания обслуживания (текущее время +  $q$ ).

Последовательно во времени рассматриваем все события, происходящие в данной системе. Изменения в функциональных характеристиках

системы отражаем в трех протоколах: суммарном времени простоя, суммарном времени ожидания клиентов в очереди, длины очереди. Процедуры повторяются до тех пор, пока не будет промоделирован весь интервал  $(0, T)$ . После этого можно определить различные операционные характеристики, исходя из периода моделирования:

$$\left( \begin{array}{l} \text{Доля времени простоя} \\ \text{системы, \%} \end{array} \right) = \frac{\text{Сумм. время простоя}}{\text{Период моделир. } T} \times 100$$

$$\left( \begin{array}{l} \text{Среднее время ожидания} \\ \text{клиентом обслуживания} \end{array} \right) = \frac{\text{Суммарное время ожд. } W}{\text{Число поступ. клиентов}}.$$

Вычисление средней длины очереди осуществляется несколько иначе.

Строим гистограмму (рис. 8.6), а затем вычисляем

$$\bar{Q} = \frac{\text{Площадь } A}{\text{Моделируемый период } T}.$$

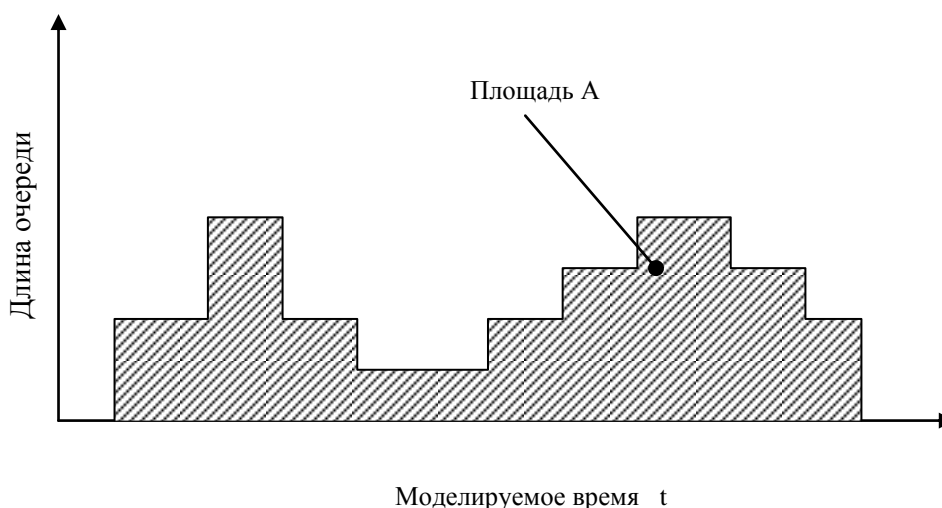


Рис. 8.6. Вычисление средней длины очереди

Моделирование дает и другую информацию, например, о распределении времени ожидания, которую можно восстановить с помощью соответствующих показателей, представленных в форме гистограммы. Более подробно данный пример рассматривается на практических занятиях. Вышеописанная процедура получения операционных характеристик имеет большое сходство с физическим экспериментом.

### 8.3.2. Определение характеристик стационарного случайного процесса по одной реализации

При имитационном моделировании нередко приходится встречаться со случаем, когда моделируемый случайный процесс является стационарным и протекает неограниченно долго, имея не зависящие от времени вероятностные характеристики.

Если речь идет об изучении начального, нестационарного периода функционирования системы, то моделирование производится обычным способом - разыгрывается множество реализаций процесса и необходимые вероятностные характеристики находятся обработкой «опытного» материала как статистические средние по множеству реализаций. Однако, когда речь идет об изучении не переходного, а стационарного, установившегося режима, достигаемого при  $t \rightarrow \infty$ , обстановка меняется. Дело в том, что при моделировании стационарных случайных процессов обычно можно пользоваться не множеством реализаций, а одной достаточно длинной реализацией. При этом вероятностные характеристики случайного процесса могут быть получены не как средние по множеству реализаций, а как средние по времени для одной достаточно длинной реализации.

Строго говоря, одной стационарности процесса для этого недостаточно. Процесс должен обладать еще так называемым эргодическим свойством. В элементарном истолковании это свойство состоит в том, что предельный режим, устанавливающийся в системе через некоторое время ее работы, не зависит от того, каковы были начальные условия и первоначальный период работы системы - каждая отдельная реализация является как бы «полномочным представителем» всего класса реализаций. Это значит, что какая бы реализация не выбрана, при  $t \rightarrow \infty$  получим процесс с одними и теми же характеристиками.

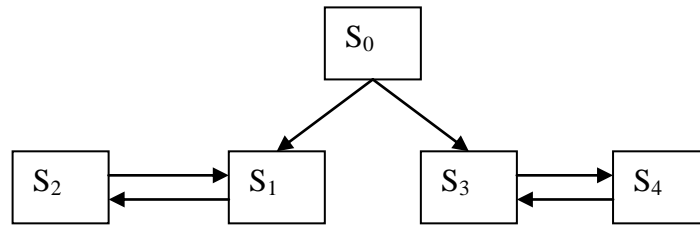


Рис. 8.7. Пример стационарного неэргодического процесса

Пример процесса стационарного, но не обладающего эргодическим свойством приведен на рис. 8.7.

Все потоки событий, переводящие систему из состояния в состояние, считаем стационарными. Пусть при  $t = 0$  система находится в  $S_0$ . Из него она может перейти либо в  $S_1$ , либо в  $S_3$ . Перейдя в  $S_1$ , система начнет «циркулировать» по состояниям  $S_1$  и  $S_2$ . Благодаря стационарности потоков событий, вызывающих эту циркуляцию, через достаточное время вероятности  $p_1(t)$  и  $p_2(t)$  станут постоянными, а процесс циркуляции стационарным:  $p_1(t) \rightarrow p_1 = \text{const}$ ;  $p_2(t) \rightarrow p_2 = \text{const}$ . Если же из  $S_0$  система перешла не в  $S_1$ , а в  $S_3$ , то она будет циркулировать по состояниям  $S_3$ ,  $S_4$  и  $p_3(t) \rightarrow p_3 = \text{const}$ ;  $p_4(t) \rightarrow p_4 = \text{const}$ . Таким образом, в этом примере процесс будет стационарным, но не эргодическим и его вероятностные характеристики существенно зависят от начального периода (начального поведения системы). Ясно, что моделирование такого процесса с помощью одной (хотя бы и очень длинной) реализации недостаточно для получения его вероятностных характеристик.

Эргодические случайные процессы на практике встречаются чаще, чем неэргодические. В частности, эргодическими оказываются процессы, протекающие в системах, граф состояний которых относится к схеме «гибели и размножения». Здесь система может через какое-то число шагов перейти из каждого состояния в каждое другое и «расщепление» процесса, подобного происходящему в системе с графом на рис. 8.7, не происходит. Дока-

зять существование предельного режима можно, строго говоря, только для марковской системы. Однако, с помощью косвенных рассуждений часто и в случае немарковских систем удается убедиться в существовании предельного режима.

Пример. Имеется двухканальная ( $n = 2$ ) СМО с очередью. Число мест в очереди  $m = 3$ . Поток заявок – не показательный, т.е. интервалы времени между заявками представляют собой независимые случайные величины, распределенные по некоторому (непоказательному) закону  $f(t)$ . Время обслуживания одной заявки – также случайная величина, распределенная по непоказательному закону  $\varphi(t)$ , отличному от  $f(t)$ . Требуется, моделируя работу СМО и располагая только одной длинной реализацией, оценить приближенно предельные характеристики системы ( $t \rightarrow \infty$ ):

- вероятности состояний (вероятности того, что будут заняты 0, 1, 2 канала; вероятности того, что в очереди будут находиться 0, 1, 2, 3 заявки);
- среднее время ожидания заявки в очереди; дисперсию времени ожидания заявки в очереди;
- среднее число занятых каналов;
- вероятность отказа (того, что заявка покинет СМО необслуженной).

Граф состояний системы имеет вид, аналогичный графу на рис. 5.5. Число состояний конечно; из каждого состояния можно перейти в каждое; потоки событий, переводящие систему из состояния в состояние, стационарны (хотя и непуассоновские); из этого заключаем, что система обладает эргодическим свойством и моделирование по одной реализации возможно. Предположим для простоты, что в начальный момент времени система находится в состоянии  $S_0$  (это не имеет значения, так как предельный режим не зависит от начального состояния). Начнем моделирование с того, что разыграем на оси  $0t$  поток заявок, т.е. ряд случайных точек  $t_1, t_2, t_3 \dots$  – моментов прихода соответствующих заявок (реализуем последовательность

операций:  $F(t) = \int_0^t f(t)dt$ ;  $R, T_1 = F^{-1}(R), t_1 = 0 + T_1$ ;  $R_1, T_2 = F^{-1}(R_1), t_2 = t_1 +$

$T_2$  и т.д.). Разумеется, цепочку моментов прихода заявок надо сделать достаточно длинной, разыграв, во всяком случае, не менее нескольких сот значений случайной величины  $T$ . Изобразим процедуру моделирования с помощью наглядной схемы (рис. 8.8). Вверху поместим ось времени (0) с отмеченными на ней моментами поступления заявок. Ниже поместим еще пять осей: (1), (2), (3), (4), (5). На осях (1) и (2) будем изображать состояния первого и второго каналов (жирная черта - «занят», тонкая - «свободен»). На осях (3), (4), (5) будем изображать состояния первого, второго и третьего мест в очереди (жирная черта - «занято», тонкая - «свободно»).

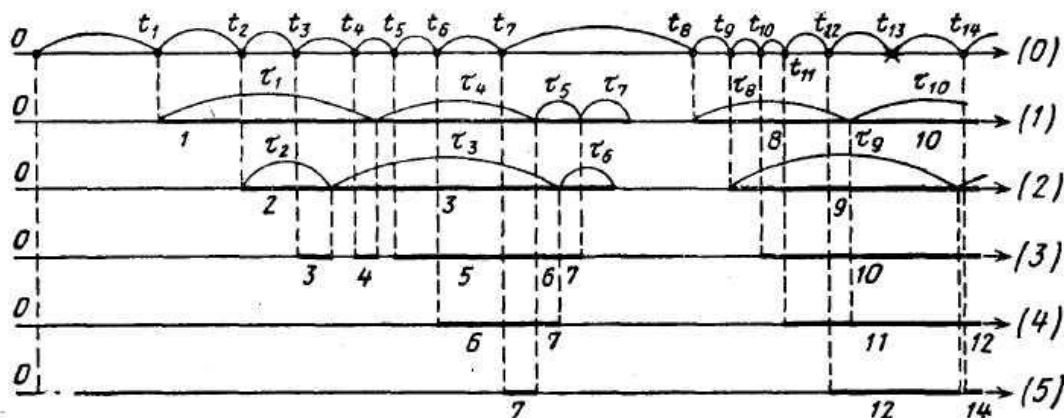


Рис. 8.8. Схема имитационного моделирования

Против каждого участка занятости канала (места в очереди) для удобства обработки проставлен номер заявки, занимающей это место; можно проследить, как заявка «путешествует» с последних мест в очереди на первые, затем - на обслуживание. Заявка, получившая отказ, помечается звездочкой. Предположим, что моделирование реализации продолжено достаточно долго (настолько долго, что влияние начальных условий перестает сказываться). Вероятности  $p_0, p_1, \tilde{p}_2$  того, что будут заняты 0, 1, 2 канала

( $\tilde{p}_2 = p_2 + p_3 + p_4 + p_5$ ), найдем следующим образом. Разделим всю ось  $0t$  на участки соответственно числу занятых каналов. Участки времени, на которых не занят ни один канал, отметим цифрой 0, один канал - цифрой 1, два канала - цифрой 2. На большом участке времени  $T$  сложим длины всех участков, помеченных нулем - получим  $T_0$ ; сумма длин всех участков, помеченных единицей, будет  $T_1$ , двойкой -  $T_2$ . Очевидно, что  $T = T_0 + T_1 + T_2$ . При большом  $T$  вероятности  $p_0, p_1, \tilde{p}_2$  будут приближенно равны отношениям соответствующих времен к общему времени:  $p_0 \approx T_0/T; p_1 \approx T_1/T; \tilde{p}_2 \approx T_2/T$ .

Заметим, что участок  $T$  целесообразно отсчитывать не с самого начала процесса, где еще сказывается влияние начальных условий, а от более удаленного по времени момента  $0'$ , где влияние начальных условий уже практически перестает сказываться.

Найдем вероятности  $\hat{p}_0, \hat{p}_1, \hat{p}_2, \hat{p}_3$  того, что в очереди будут стоять 0, 1, 2, 3 заявки. Складывая длины всех одинаково помеченных участков и деля суммы  $\hat{T}_i$  на  $T$ , получим:

$$\hat{p}_0 \approx \hat{T}_0/T; \hat{p}_1 \approx \hat{T}_1/T; \hat{p}_2 \approx \hat{T}_2/T; \hat{p}_3 \approx \hat{T}_3/T.$$

Среднее число занятых каналов  $\bar{z}$  получится обычным способом как математическое ожидание дискретной случайной величины  $Z$  - числа занятых каналов:

$$\bar{z} = 1 \cdot p_1 + 2 \cdot \tilde{p}_2 = p_1 + 2 \tilde{p}_2.$$

Среднее время ожидания заявки в очереди  $\bar{t}_{ож}$  находим следующим образом: рассмотрим ряд заявок, поступивших на большом участке времени  $T$  в моменты  $t_k, t_{k+1}, t_{k+2}, \dots, t_{k+i}, \dots, t_{k+N}$ , и для каждой из них непосредственно подсчитаем время ожидания в очереди  $t_{ож}^{k+i}$ , равное нулю, если  $(k+i)$ -я заявка была сразу принята к обслуживанию (или получила отказ), и сум-

ме времен ожидания этой заявки на разных осях ((3), (4), (5)), если она стояла в очереди. Среднее время ожидания заявки в очереди приближенно найдется как среднее арифметическое этих времен:

$$\bar{t}_{ож} \approx \frac{1}{N} \sum_{i=1}^N t_{ож}^{k+i}.$$

Дисперсия времени ожидания найдется аналогичным образом, как среднее арифметическое квадратов времен ожидания минус квадрат среднего времени ожидания:

$$D[T_{ож}] \approx \frac{1}{N} \sum_{i=1}^N (t_{ож}^{k+i})^2 - \bar{t}_{ож}^2.$$

Наконец, вероятность отказа найдется на большом участке времени  $T$  как отношение числа заявок  $N^*$ , помеченных звездочкой (получивших отказ), к общему числу  $N$  заявок, поступивших за это время:

$$P_{отк} \approx N^*/N.$$

### 8.3.3. Методы сбора статистических данных

Рассмотрим теперь важный вопрос, касающийся получения наблюдений при моделировании. Поскольку моделирование представляет собой эксперимент, получаемые наблюдения должны быть статистически независимы и одинаково распределены, с тем, чтобы была обеспечена возможность правильной статистической интерпретации моделируемой системы.

В любом физическом эксперименте оценка результата обычно основывается на среднем значении  $N$  независимых наблюдений. Величина  $N$  выбирается таким образом, чтобы был гарантирован определенный доверительный уровень. При имитационном моделировании оценка операционной характеристики системы также должна основываться на  $N$  наблюдениях. Тем не менее получение независимых наблюдений при мо-

делировании намного сложнее, чем при обычном лабораторном эксперименте. Многочисленные примеры применения метода имитационного моделирования показывают, что первоначально результаты моделирования имеют неустойчивый характер (переходное состояние), а устойчивость (стационарность) обычно достигается при достаточно продолжительном прогоне модели. Таким образом, следует не начинать наблюдения слишком рано, поскольку полученные при этом данные характеризуются значительным разбросом и поэтому не могут давать представление о подлинном поведении системы. Представляет интерес в основном получение наблюдений после того, как достигнуто стационарное состояние, так как в этом случае выборочная ошибка (измеряемая средним квадратичным отклонением) уменьшается, и, следовательно, результаты становятся более точными.

При дискретном моделировании достижении стационарного состояния зависит от начальных условий системы, а также от параметров системы. Поскольку основная цель состоит в получении наблюдений с возможно меньшей ошибкой, этого можно достичь с помощью:

- 1) очень длительных прогонов модели, позволяющих увеличить вероятность достижения стационарного состояния;

- 2) повторения прогонов модели с различными последовательностями случайных чисел, каждый из которых дает одно наблюдение. Использование различных последовательностей случайных чисел приводит к желаемой независимости получаемых наблюдений. Выборочная ошибка уменьшается, если наблюдения получены в стационарных условиях, но ее можно сделать еще меньше, взяв среднее этих наблюдений, поскольку среднее квадратичное отклонение среднего  $N$  наблюдений составляет  $1/\sqrt{N}$  среднего квадратичного отклонения отдельных наблюдений.

Следует обратить внимание на усилия, необходимые для получения наблюдений. Другими словами, хотя уменьшение выборочной ошибки важно, нельзя добиваться этой цели любой ценой.

На практике при получении наблюдений при моделировании необходимо иметь в виду два следующих обстоятельства:

1) затраты на моделирование могут существенно зависеть от продолжительности прогонов модели;

2) выборочную ошибку можно уменьшить за счет использования улучшенных методов получения выборок, направленных на уменьшение статистической ошибки.

Вполне естественно, что нельзя получить что-то из ничего. Продолжительность прогонов модели можно уменьшать, либо получая выборки в переходном состоянии, либо достигая устойчивого состояния, но жертвуя при этом некоррелированностью наблюдений. Методы уменьшения выборочной ошибки, называемые методами уменьшения дисперсии, полезны, однако их реализация при построении имитационной модели связана с рядом трудностей.

Имитационное моделирование представляет собой статистический эксперимент. Его результаты должны основываться на соответствующих статистических проверках (с использованием, например, доверительных интервалов и методов проверки гипотез). Для выполнения этой задачи получаемые наблюдения и имитационный эксперимент должны удовлетворять трем требованиям.

1. Наблюдения имеют стационарные распределения, т.е. распределения не изменяются во время проведения эксперимента.

2. Наблюдения подчиняются нормальному распределению.

3. Наблюдения независимы.

Иногда на практике результаты имитационного моделирования не удовлетворяют ни одному из этих требований. Тем не менее, их выполнение гарантирует наличие корректных способов сбора наблюдений над имитационной моделью.

Рассмотрим сначала вопрос о стационарности распределений (первое требование). Нестационарность процесса имитации можно преодолеть путем использования выборки достаточно большого объема. Результаты наблюдений над моделью зависят от продолжительности периода имитации. Начальный период неустойчивого поведения модели (системы) обычно называется переходным. Когда результаты имитационного эксперимента стабилизируются, говорят, что система работает в установившемся режиме. Продолжительность переходного периода определяется в значительной степени начальными характеристиками модели и невозможно предсказать, когда наступит установившийся режим. В общем случае, чем длиннее продолжительность прогона модели, тем выше шанс достижения установившегося состояния.

Рассмотрим теперь второе требование, состоящее в том, что наблюдения над имитационной моделью должны иметь нормальное распределение. Это требование можно выполнить, если привлечь центральную предельную теорему, утверждающую, что распределение среднего выборки является асимптотически нормальным независимо от распределения генеральной совокупности, из которой взята выборка. Центральная предельная теорема, таким образом, есть главное средство удовлетворения требования о нормальности распределения.

Третье требование касается независимости наблюдений. Природа имитационного эксперимента не гарантирует независимости между последовательными наблюдениями над моделью. Однако использование выборочных средних для представления отдельных наблюдений позволяет смягчить проблему, связанную с отсутствием независимости. Для этого, в частности, следует увеличить интервал времени имитации для получения выборочного среднего.

Понятие переходного и установившегося состояний имеют силу в ситуациях, именуемых незаканчивающейся имитацией, т.е. имитацией, применяемой к системам, которые функционируют бесконечно долго. В случае заканчивающейся имитации (например, работа банка, если он обычно работает восемь часов в день) переходное поведение является частью нормального функционирования системы и, следовательно, не может игнорироваться. Единственным выходом в такой ситуации является увеличение, насколько это возможно, числа наблюдений.

Рассмотрим теперь три наиболее общих метода сбора информации в процессе имитационного моделирования.

*Метод подинтервалов.* Предположим, что имитация длится на протяжении  $T$  единиц времени (т.е. длина прогона модели равна  $T$ ) и требуется получить  $N$  наблюдений. В соответствии с методом подинтервалов необходимо сначала «обрезать» информацию, относящуюся к переходному периоду, а затем разделить остаток результатов имитации на  $N$  равных подинтервалов (групп). Среднее значение искомой величины внутри каждого подинтервала используется затем в качестве единственного наблюдения. Отбрасывание начального переходного периода означает, что статистические данные, собранные на протяжении этого периода, не используются. Преимущество данного метода состоит в том, что влияние переходных (нестационарных условий) уменьшается, в частности, на те данные, которые собраны в конце времени имитации. Недостаток заключается в том, что последовательные группы с общей границей являются коррелированными, что приводит к невыполнению предположения о независимости. Влияние корреляции может быть уменьшено путем увеличения интервала времени для каждой группы.

*Метод повторения.* В данном методе каждое наблюдение представляется независимым прогоном модели, в котором переходной период не учитывается. Вычисление средних величин выборки для каждой группы проводится точно так, как и в методе подинтервалов. Преимуществом этого метода является то, что каждый имитационный прогон модели определяется своей последовательностью случайных чисел из интервала  $[0,1]$ , что действительно обеспечивает статистическую независимость получаемых наблюдений. Недостаток состоит в том, что все наблюдения могут оказаться под сильным влиянием начальных переходных условий. Этот недостаток можно смягчить в результате увеличения длины прогона модели.

*Метод циклов.* Этот метод можно рассматривать как расширенный вариант метода подинтервалов. Мотивацией данного метода является попытка

ка уменьшить влияние автокорреляции, которая характерна для метода подинтервалов, путем выбора групп таким образом, чтобы обеспечить одинаковые начальные условия для каждой из них. В отличие от метода подинтервалов, в методе циклов длины интервалов каждой группы могут оказаться различными. Недостатком метода является меньшее по сравнению с методом подинтервалов число получаемых наблюдений при заданной длине прогона модели. Это следует из того, что нельзя заранее сказать, когда новая группа (цикл) начинается и какова продолжительность каждого цикла. Однако можно ожидать, что в стационарных условиях начальные точки последовательных циклов будут расположены более или менее равномерно. Вычисление среднего для цикла  $i$  в рассматриваемом методе определяется в виде отношения двух случайных величин  $a_i$  и  $b_i$ , т.е. в виде  $x_i = a_i/b_i$ . Определение величин  $a_i$  и  $b_i$  зависит от вычисляемой переменной. Например, если переменная является функцией времени, то  $a_i$  представляет площадь под кривой, а  $b_i$  – длительность соответствующего интервала времени. Если же переменная является функцией количества событий, то  $a_i$  – общая сумма наблюдений этой величины в пределах цикла  $i$ , а  $b_i$  – общее число событий внутри соответствующего цикла.

Прямым методом уменьшения выборочной ошибки является увеличение размера выборки и использование продолжительного прогона модели. Такая процедура вряд ли будет экономичной. Поэтому разработаны другие методы, которые, не увеличивая размера выборки, уменьшают выборочную ошибку. Такие методы получили название *методов уменьшения дисперсии*.

Наиболее широко используемым методом уменьшения дисперсии является метод дополнения. Он основан на том, что если  $R$  – полученное случайное число, то его дополнение  $1 - R$  также случайное число. Идея метода дополнения связана с осуществлением двух прогонов, в одном из которых

используется последовательность  $R_1, R_2, \dots, R_k$ , а в другом  $(1 - R_1), (1 - R_2), \dots, (1 - R_k)$ . Пусть  $x_1, x_2, \dots, x_k$  и  $y_1, y_2, \dots, y_k$  - соответствующие результаты двух таких прогонов. Между двумя последовательностями  $(x_1, x_2, \dots, x_k)$  и  $(y_1, y_2, \dots, y_k)$  имеет место отрицательная корреляция, поскольку как только получается малое (случайное) число  $x_i$ , сразу же генерируется большое (случайное) число  $y_i$ , и наоборот. Кроме того, обе последовательности имеют одинаковое среднее  $\mu$ . В данном случае оценка  $\mu$  определяется как

$$\bar{z} = (1/N) \sum_{i=1}^N (x_i + y_i) / 2.$$

Оценка  $\mu$  с помощью такого  $\bar{z}$  эквивалентна использованию выборки размером  $2N$ . Нетрудно доказать, что  $\bar{z}$  обладает меньшей дисперсией, чем оценка, получаемая из выборки размером  $2N$ , в которой наблюдения не коррелированы.

К методам уменьшения дисперсии также относят: выбор оценок с наименьшей дисперсией, расслоенные выборки, значимые выборки, контрольные переменные, параллельные выборки, условная имитация.

*Выбор оценок с наименьшей дисперсией.* Установлено, что усреднение по модельному времени дает оценки с меньшей дисперсией, чем усреднение по числу наблюдений. В частности, стационарную вероятность иметь в системе  $k$  заявок лучше оценивать по формуле

$$\hat{p}_k = \sum_i t_i(k) / T,$$

чем согласно

$$\hat{p}'_k = N(k) / N.$$

Здесь  $t_i(k)$  - длина  $i$  - того отрезка времени, проведенного системой в  $k$  - том состоянии,

$T$  - системное время моделирования,

$N(k)$  - число наблюдений системы в  $k$  - том состоянии,

$N$  - полное число наблюдений.

Также установлено, что для определения средних характеристик СМО посредством имитационного моделирования целесообразно оценить среднее время ожидания, а остальные оценки вычислить по формулам связи со средним временем ожидания. Использование этих рекомендаций, как отмечалось, давало уменьшение дисперсии оценок до 200 раз. Однако отметим, что расчет, например, средней длины очереди через среднее время ожидания заявки требует моделирования очереди: запоминание моментов прибытия заявок, их продвижения и т.п. Прямой расчет упомянутого показателя эти затраты исключает. Теперь ясно, чем оплачивается повышение точности оценок через временные показатели.

#### **8.3.4. Структурная схема имитационной модели для многоканальной СМО с ожиданием**

На рис. 8.9 дается детальная структурная схема имитационной модели многоканальной СМО с ограниченным ожиданием (ограничение на число заявок в системе), с помощью которой определяются стационарное распределение  $p[0: R]$  числа заявок в системе и начальные моменты распределения времени ожидания. В этой модели используются два класса событий: прибытие заявки и завершение обслуживания. Поток заявок моделируется как рекуррентный (момент прибытия очередной заявки получаем добавлением случайного интервала к предыдущему), моменты освобождения каналов — добавлением к текущему моменту случайной длительности обслуживания. Упомянутые интервалы формируются посредством датчиков псевдослучайных чисел, настроенных на требуемые законы распределения.

Управление прогоном производится по числу обслуженных заявок. Схема рассчитана на выполнение 10 прогонов с выдачей результатов в конце каждого из них. Приведем словарь этой модели:

T — таймер;

T1 — момент предыдущего события.

$e[l:n]$  — моменты освобождения каналов от текущего обслуживания;  
 $E$  — наиболее ранний из них;  
 $\tau[n+1:R]$  — моменты прибытия находящихся в очереди заявок;  
 $Z$  — момент прибытия очередной заявки;  
 $p[0:R]$  — счетчики стационарных вероятностей;  
 $r$  — текущее число заявок в системе;  
 $nz$  — общее число выбранных на обслуживание заявок (ограничено 5000):  
 $n1$  — текущее число заявок в данной серии (ограничено 500);  
 $d$  — количество отказов в приеме на обслуживание;  
 $s[1:4]$  — счетчики для моментов распределения времени ожидания начала обслуживания;  
 $w[1:4]$  — искомые моменты.

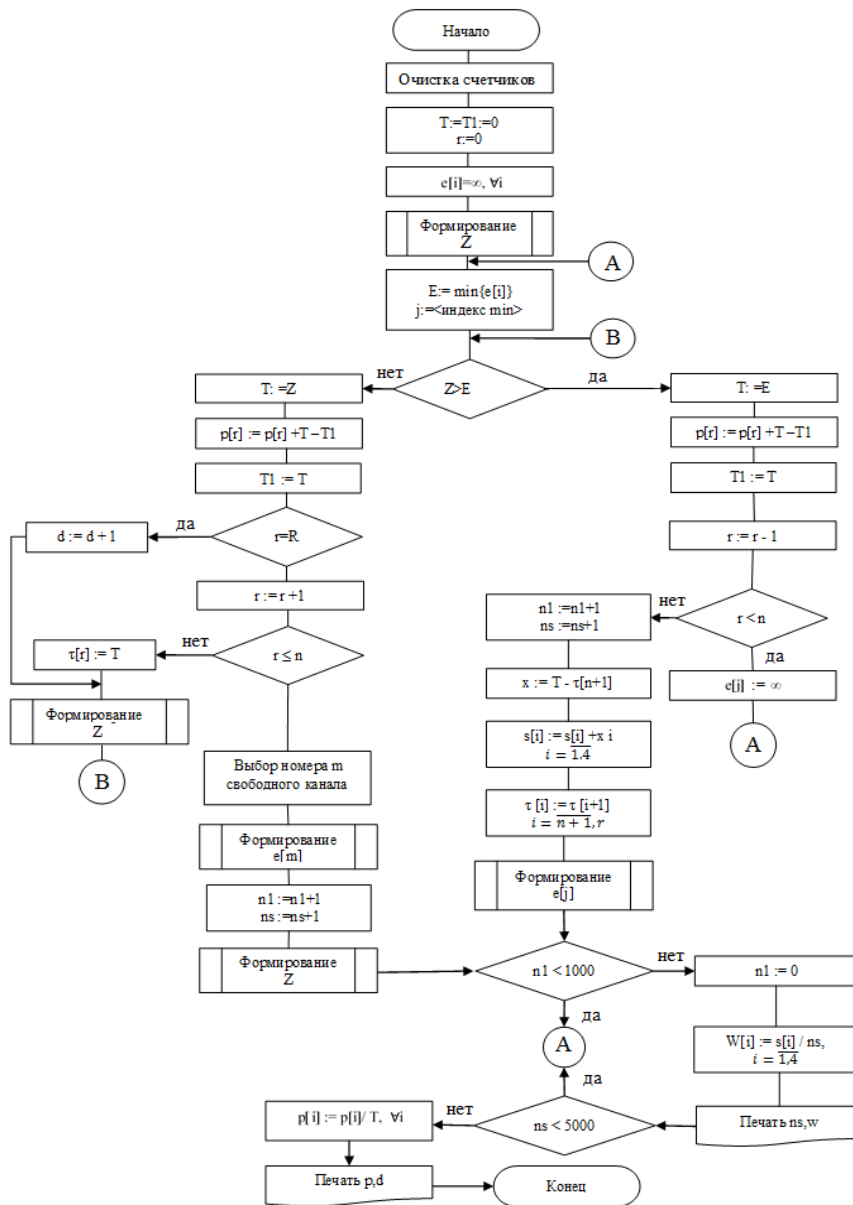


Рис. 8.9. Структурная схема имитационной модели

Работа модели начинается с установки в нуль таймера (счетчика модельного времени)  $T$ , момента предыдущего события  $T1$  и всех остальных накапливающих счетчиков. Моменты  $\{e[i]\}$  выхода из системы обслуженных заявок полагаем равными бесконечности (точнее — числу, заведомо превышающему предельное значение таймера в процессе прогона). Далее формируется случайный момент  $Z$  прибытия заявки и выбираются  $E$  (минимальный из моментов освобождения каналов) и индекс минимума  $j$ .

Вариант обработки текущего события определяется соотношением между  $Z$  и  $E$ . Если  $Z < E$ , то ближайшим событием в системе является прибытие заявки. Соответственно продвигается таймер  $T$ , к счетчику  $p[r]$  времени пребывания системы в состоянии с  $r$  заявками добавляется разность  $T - T1$  и обновляется значение  $T1$ . Если новая заявка застает в системе предельное число  $R$  ранее прибывших, то к счетчику отказов  $d$  добавляется единица. Затем формируется момент  $Z$  прибытия новой заявки. При наличии в системе свободного места  $r$  увеличивается на единицу и проверяется, есть ли свободные каналы. Если  $r > n$ , то их нет, для прибывшей заявки запоминается момент  $\tau[r]$  прибытия и формируется очередная заявка. В обоих рассмотренных случаях мы выходим на метку  $B$  — сравнение  $E$  с обновленным  $Z$ .

При  $r < n$  в системе имеется хотя бы один свободный канал. Тогда определяется его номер  $m$ ; в элементе массива  $e[m]$  формируется момент освобождения  $m$ -го канала; корректируются счетчики числа случаев ожидания  $n1$  и  $n2$  (длительность ожидания в данном случае нулевая, и соответствующие счетчики времени ожидания не изменяются); опять же формируется момент прибытия новой заявки; при  $n1 < 500$  происходит возврат к выбору  $\min_i \{e[i]\}$ , поскольку один из элементов этого массива обновляется.

В случае  $Z > E$  очередным событием оказывается завершение начатого обслуживания. Здесь также продвигается таймер, пересчитывается счетчик  $p[r]$  времени пребывания в состоянии  $r$ , обновляется момент последнего изменения состояния  $T1$ . Затем число  $r$  заявок в системе уменьшается на единицу. Если  $r < n$ , то очереди в системе нет, в массив  $e[i]$  для освобожденного  $j$ -го канала записывается «бесконечность» и выполняется переход на выбор минимальной компоненты массива  $e[i]$ .

При  $r > n$  в системе существует очередь, так что в освободившийся канал будет выбрана на обслуживание новая заявка. Соответственно увеличиваются на единицу счетчики  $n1$  и  $n2$  случаев выбора на обслуживание, вычисляется время  $x$  ожидания начала обслуживания головной заявкой очереди. Далее этот  $x$  и его последовательные степени добавляются к счетчикам  $\{s[i]\}$ , накапливающим данные для расчета моментов распределения времени ожидания. Затем продвигается очередь (точнее, записанные в ней моменты  $\{\tau[i]\}$  прибытия в систему ожидающих заявок) и формируется время  $e[j]$  освобождения вновь занимаемого  $j$ -го канала.

При счетчике  $n1 < 500$  модель возвращается на выбор  $\min_i \{e[i]\}$ . При  $n1=500$  прогон серии заканчивается, счетчик  $n1$  обнуляется, рассчитываются и выводятся на печать оценки  $\{w[i]\}$  моментов распределения времени ожидания вместе  $i$  текущим значением  $n2$ . Указанные промежуточные выдачи нужны для контроля темпа прохождения модели (чтобы рассчитать потребное машинное время в случае снятия задачи до полного завершения моделирования), а также для оценки влияния на результаты начальных условий и выполненного числа испытаний. После достижения  $n2=5000$  на печать выдаются стационарные вероятности  $\{p[r]\}$ , определяемые как доля общего времени  $T$ , в течение которого в системе находилось ровно  $r$  заявок, а также количество  $d$  отказов в приеме на обслуживание. На этом моделирование завершается.

#### **8.4. Моделирование на универсальных и специализированных языках**

Любой алгоритм моделирования может быть записан на универсальном процедурно ориентированном языке. Преимуществами подобной организации являются:

- возможность работы на практически любой ЭВМ;

- отсутствие необходимости в специальном программном обеспечении и его изучении;

- «прозрачность» логики модели не только для разработчика, но и для широкого круга программистов;

- доступность встроенных функций, библиотечных подпрограмм, средств статистического анализа, редактирования вывода и других возможностей, предоставляемых штатными системами программирования широкого назначения;

- отсутствие ограничений на состав и логику модели, исключая налагаемые конфигурацией моделирующей вычислительной установки и режимом ее использования;

- возможность управления экспериментом и реализация методов уменьшения дисперсии результатов.

Опыт программирования на универсальных языках показал, что составление имитационных моделей на них весьма трудоемко по следующим обстоятельствам:

- постоянно возрастает количество элементов и отслеживаемых процессов в моделируемых системах;

- имитационные модели сложны в отладке из-за богатой логики, взаимной зависимости ветвей алгоритма, большой трудоемкости прогонов и низкой точности результатов, способной в ряде случаев замаскировать ошибку;

- незначительные изменения моделируемой системы требуют внесения в программу многочисленных поправок и проведения отладки практически заново;

- достигнутая на уровне блок-схемы наглядность модели теряется после составления программы и фактически требует от программиста почти такого же знания системы, как от разработчика.

Реализация имитационных моделей влечет за собой два различных типа вычислений: 1) манипуляции регистрацией, которые имеют дело с хронологическим накоплением и обработкой событий модели и 2) вычисления, связанные с генерированием случайных чисел и сбором статистических данных, относящихся к модели. Вычисления первого типа основываются на различных логических методах обработки списков, а вычисления второго типа обычно очень громоздки и занимают много времени. Природа этих вычислений делает компьютер важным инструментом в реализации имитационных моделей и, в свою очередь, стимулирует создание специализированных языков программирования, что позволяет выполнять эти вычисления более удобным и эффективным способом.

Трудности моделирования на общецелевых языках программирования снимаются при использовании некоторых стандартных схем в качестве типовых моделей. Такой язык приближает ЦВМ к разработчику и разработчика – к ЦВМ. Введенные в языках понятия задают рамки, гарантирующие возможность машинной реализации и в то же время ограничивающие класс описываемых моделей. Возможность создания языков моделирования базируется на двух обстоятельствах:

Даже в сложных моделях с тысячами элементов количество видов действий остается весьма ограниченным. Такие действия естественно описывать как подпрограммы – при необходимости с дополнительными параметрами.

Имитационные модели достаточно широких классов реальных процессов имеют сходную логику, определяемую общей математической моделью.

Эти обстоятельства в условиях постоянно растущих потребностей в имитационных моделях привели к созданию языков моделирования на ЭЦВМ, которые существенно ускорили написание и в особенности отладку моделирующих программ. В общем случае моделирование обеспечивается входным языком, компилятором, системой интерпретации отдельных конструкций языка, библиотекой стандартных подпрограмм, банком дан-

ных (для проблемно ориентированных систем), средствами отладки и т.д. Поэтому правильнее пользоваться термином «система моделирования».

Языки моделирования позволяют писать имитационные программы в такой форме, которая напоминает описание моделируемой системы, и при этом так, что малым изменениям описания системы соответствуют малые изменения в программе. Системы моделирования упрощают описание модели, позволяют строить модель и работать с ней в диалоговом режиме, зачастую включают в себя банки данных по предметным областям.

Преобразование состояний элементов системы обычно связано в выполнении арифметических операций над соответствующими переменными. Поэтому язык моделирования должен включать в себя базисные вычислительные средства (как правило, подмножество какого-либо алгоритмического языка). Специфические модельные средства должны обеспечить удобство описания реальных объектов, имитации событий, параллельного выполнения процессов и их взаимодействия. Наиболее характерной особенностью систем моделирования является наличие выделенной переменной – системного времени. Сервисные средства должны обеспечить удобный интерфейс, средства отладки и обработку статистического материала.

Логика моделирования дискретных процессов в основном встроена в универсальный интерпретатор и при написании программы моделирования подразумевается. Естественно, что программа модели оказывается короче, чем на универсальном языке, поскольку в последнем случае моделирующая программа является синтезом описания модели и его интерпретатора. Компиляторы обеспечивают контроль синтаксической правильности программы. Это обеспечивает быстроту и правильность составления программ моделирования, в особенности при возможности интерактивной работы с моделью.

В настоящее время существует большое количество языков имитационного моделирования и перед разработчиком машинной модели возникает проблема выбора языка, наиболее эффективного для целей моделирования конкретной системы. Языки моделирования заслуживают пристального внимания, так как, во-первых, число существующих языков и систем моделирования превышает несколько сотен и необходимо научиться ориентироваться в них, а во-вторых, почти каждый новый язык моделирования не только является средством, облегчающим доведение концептуальной модели до готовой машинной моделирующей программы, но и представляет собой новый способ “видения мира”, т. е. построения моделей реальных систем. Основными моментами, характеризующими качество языков моделирования, являются: удобство описания процесса функционирования системы; удобство ввода исходных данных моделирования и варьирования структуры, алгоритмов и параметров модели; реализуемость статистического моделирования; эффективность анализа и вывода результатов моделирования; простота отладки и контроля работы моделирующей программы; доступность восприятия и использования языка.

Имеющиеся ЯИМ можно разбить на три основные группы, соответствующие трем типам математических схем: непрерывные, дискретные и комбинированные.

*Непрерывное* представление системы сводится к составлению уравнений, с помощью которых устанавливается связь между эндогенными и экзогенными переменными модели. Примером такого непрерывного подхода является использование дифференциальных уравнений. Причем в дальнейшем дифференциальные уравнения могут быть применены для непосредственного получения характеристик системы, это, например, реализовано в языке MIMIC. А в том случае, когда экзогенные переменные модели

принимают дискретные значения, уравнения являются разностными. Такой подход реализован, например, в языке DYNAMO.

Представление системы в виде типовой схемы, в которой участвуют как непрерывные, так и дискретные величины, называется *комбинированным*. Примером языка, реализующего комбинированный подход, является GASP, построенный на базе языка FORTRAN.

*Языки моделирования дискретных систем.* Доступные языки дискретного имитационного моделирования делятся на две большие категории.

Языки, ориентированные на планирование событий.

Языки, ориентированные на обработку процессов (процедур).

При использовании языков, ориентированных на планировании событий, пользователю необходимо указать действия, связанные с каждым событием, происходящим в системе. Основная роль программы в этом случае сводится к автоматизации процесса получения случайных значений, имеющих соответствующее распределение, хронологическому накоплению, обработке событий и сбору данных, относящихся к модели.

Процедурно-ориентированные языки используют блоки (или узлы), которые можно соединять для формирования сети, которая описывает движение транзакций или объектов (т.е. клиентов) в системе. Например, наиболее известными типами узлов в любом языке имитационного моделирования являются источник, в котором транзакции создаются, очередь, где при необходимости они могут ожидать, и сервиса, где выполняется обслуживание. Каждый из этих узлов при его определении обеспечивается всей необходимой информацией, позволяющей выполнять имитацию автоматически. Например, если время между поступлениями заказов для источника определено, процедурно-ориентированный язык автоматически «знает», когда будут иметь место события, связанные с прибытием заказа. В дейст-

вительности каждый узел имеет установленные инструкции, т.е. точно определяет как и когда транзакции перемещаются по имитационной сети.

Процедурно-ориентированные языки управляются теми же действиями, что и языки, ориентированные на планирование событий. Отличие состоит в том, что эти действия автоматизированы для освобождения пользователя от утомительных вычислительных и логических деталей. В некотором отношении можно рассматривать процедурно-ориентированные языки как основанные на концепции «черного ящика», имеющего заданный «вход» и «выход». Это означает, что процедурно-ориентированные языки просты и легки в использовании благодаря гибкости процесса моделирования.

Наиболее известными языками программирования, ориентированными на планирование событий, являются SIMSCRIPT, SLAM, SIMAN. Развитие этих языков на протяжении многих лет привело к тому, что они включают и возможности процедурно-ориентированных языков. Все эти языки позволяют пользователю создавать модели (или их отдельные части) на языках высокого уровня, таких как FORTRAN и С. Это необходимо для того, чтобы дать возможность пользователю программировать сложные логические операции, которые невозможно или трудно осуществить обычными средствами этих языков. Главной причиной этого является ограничительная и, возможно, запутанная процедура, с помощью которой данные языки перемещают транзакции между очередью и сервисами, присутствующими в модели.

Первым процедурно-ориентированным языком был GPSS. Этот язык, первая версия которого появилась в начале 60-х годов, совершенствовался на протяжении нескольких лет, чтобы удовлетворить новым требованиям, связанным с моделированием сложных систем. Чтобы эффективно исполь-

зовать этот язык, пользователю необходимо настроить примерно 80 различных блоков.

Другой процедурно-ориентированный язык программирования, именуемый SIMNET II, разработан для непосредственного моделирования сложных ситуаций. SIMNET II использует три типа узлов: источник, который генерирует заказы, очередь, где заказы могут ожидать, и средство обслуживания, где выполняется обслуживание. В этом языке имеется еще четвертый дополнительный тип узла, который рассматривается как средство обслуживания неограниченной емкости, что призвано усилить моделирующие возможности языка.

Поскольку имитационная модель — это программа для ЭВМ, то при ее создании этап программирования является одним из основных. Он может быть выполнен тремя способами.

1. Использование универсальных алгоритмических языков программирования Паскаль, С++, Фортран, PL/1, Ада и др. В этом случае программист имеет практически неограниченные возможности по созданию эффективной имитационной модели, наилучшим образом использующей ресурсы ЭВМ, особенности операционной системы, обладающей высоким быстродействием, и т. д. Создание модели таким способом требует больших трудозатрат, работы программистов высокой квалификации, взаимодействия специалистов различного профиля (системных программистов, экспертов проблемной области, исследователей и др.). Имитационная модель получается узконаправленной на решение конкретной задачи и, как правило, не может быть использована для других приложений.

2. Создание и использование специализированных языков ИМ. Событийный подход реализуют языки SLAM II, GASP IV, SIMASCRIPT II, ПСИМПАК, СИМКОМ. Языки CSL, DRAFT, HOCUS, ADLANDS реализуют подход сканирования активностей.

Из всех процессно-ориентированных языков наиболее часто употребляют языки GPSS, SLAM II, СИМУЛА, SOL, Q-GERT, SIMAN, PAWS, QNAR. Написанная на таком языке программа работает так же, как несколько программ на языке, ориентированном на события.

Языки ИМ за счет снижения гибкости и универсальности позволяют создавать имитационную модель на несколько порядков быстрее и не требуют работы системных программистов. Они обладают двумя важными достоинствами: удобством программирования и концептуальной выразительностью. Последнее достоинство позволяет четко и ясно описывать различные понятия, что наиболее важно на стадии моделирования и для определения общего подхода к изучению исследуемой системы.

Основным недостатком языков ИМ является их относительно ограниченная гибкость с точки зрения многообразия исследуемых систем и возможного разнообразия задач их анализа и синтеза. Это приводит к необходимости разработки новых моделей для альтернативных вариантов не только структуры моделируемой системы, но и управляющих элементов сложной системы. При использовании специализированных языков ИМ для подготовки имитационной модели требуется описывать структуру и физические процессы объекта абстрактными понятиями в терминах используемого языка, как правило, далекими от понятий предметной области.

3. Создание и использование проблемно-ориентированных систем моделирования, например ПОДСИМ (МГТУ им. Н.Э.Баумана), ДСИМ (ЭНИМС), АСИМПТОТА, DOSIMIS-3, Process Charter 1.0.2 (компании Scitor), Power-sim 2.01 (фирмы Modell Data AS), Ithink 3.0.61 (производства High Performance Systems), Extend+BPR 3.1 (компании Imagine That), ARENA (фирмы Systems Modeling); ProModel (фирмы ProModel); ReThink (фирмы Gensym) и ряда других.

В этих системах исследователь не пишет модель на языке, она генерируется автоматически в процессе диалога с исследователем, который происходит в понятиях языка предметной области. Системы обеспечивают моделирование с помощью графических средств, библиотек специализированных программ и языков.

Проблемно-ориентированные системы ИМ позволяют быстро и эффективно создавать имитационные модели и проводить исследования вообще без программирования, и поэтому с ними непосредственно работает исследователь. Однако это достигается еще большими ограничениями на класс моделируемых систем и снижением гибкости процесса имитации. Эти системы имеют большую стоимость и, как правило, ориентированы на выполнение крупных проектов в больших фирмах и организациях.

## **8.5. Современное содержание терминов «имитация», «имитационная модель»**

До сих пор термины «имитация», «имитационное моделирование», «имитационная модель», «имитационный эксперимент» охарактеризовывались с разных точек зрения. Соберем воедино все эти характеристики и постараемся выработать целостное представление о содержании этих терминов. Итак, термины «имитация» и «имитационный эксперимент» появились сначала в теории вероятностей и математической статистике как способ вычисления статистических характеристик интересующих нас случайных величин посредством воспроизведения реализаций соответствующего случайного процесса с помощью его математической модели. Воспроизведение реализаций случайного процесса и есть то, что естественно называть имитационным экспериментом, поскольку реальные эксперименты с измерением интересующих нас случайных величин как бы заменяются их имитацией с помощью математической модели данного процесса.

Вскоре после начала использования методов прикладной математики в управлении экономикой, планировании, исследовании операций, проектировании термины «имитация», «имитационный эксперимент» приобрели в этих областях смысл, не совпадающий с их первоначальной трактовкой. Этими терминами стали обозначать способ выбора рационального управления сложным процессом (рационального плана, рациональной конструкции проектируемого изделия), состоящий в следующем. Некоторым образом разрабатываются варианты управлений (планов, конструкций). Затем эти варианты сравниваются. Для этого при каждом таком варианте процесс (функционирование проектируемого изделия) воспроизводится с помощью его математической модели. Сравнение может происходить по некоторым формальным критериям, а может носить неформальный характер, причем

чем сложнее используемая модель, чем больше она содержит реальных факторов, влияющих на принятие решений, тем более естественна неформальная оценка сравниваемых результатов. Математические модели, ориентированные на такое их использование, получили название имитационных, процесс их составления стал называться имитационным моделированием, а каждая акция воспроизведения процесса (функционирования проектируемого изделия) - имитационным экспериментом.

Если изучаемый процесс является случайным (если процесс достаточно сложен, то почти неизбежно он случаен), для сравнения, о котором шла речь выше, необходимо выполнять то, что в теории вероятностей и математической статистике называется имитацией, т. е. вычислять статистические характеристики этого случайного процесса путем набора необходимого для этого количества реализаций (предполагается, что аналитическими средствами вычислить статистические характеристики нельзя) и именно эти статистические характеристики и сравнивать. Даже если дело обстоит именно таким образом, специалисты в области управления, планирования, проектирования, исследования операций, произнося слово «имитация», будут иметь в виду не способ вычисления характеристик случайных процессов путем набора статистики (для них это некоторая необходимая техническая деталь), а то, что альтернативные варианты управлений (планов, конструкций проектируемого объекта) являются внешними по отношению к модели процесса, задаются «извне» ее, а не являются ее продуктом. Они будут называть воспроизведения процесса имитацией и тогда, когда процесс детерминирован. Для них термин «имитация» несет смысловую нагрузку противопоставления термину «оптимизация», в то время как для специалистов в области теории вероятностей и математической статистики термин «имитация» несет оттенок противопоставления аналитическим методам расчета статистических характеристик случайного процесса.

Сопоставление оптимизационных задач с реальным содержанием задач планирования, управления, проектирования приводило к попыткам улучшить модели, лежащие в основе оптимизационных задач, что влекло за собой их усложнение, появление вместо одного критерия оптимальности нескольких или же вообще отказ от оптимизации в рамках усложнившихся моделей и использование их в режиме вариантных расчетов с задаваемыми извне модели вариантами планов (управлений, конструкций проектируемого изделия). Поскольку в этих усложненных моделях присутствовали, как правило,

случайные факторы, то получение обоснованных результатов требовало вычисления статистических характеристик, т. е. имитации в том смысле, в котором ее понимают в теории вероятностей и математической статистике. Очень скоро специалисты в области использования математических методов в планировании, управлении, исследовании операций, проектировании присвоили новое содержание терминам «имитация», «имитационная модель», «имитационный эксперимент» — то, о котором говорилось выше. Конечно, содержание этих гуманитарных терминов нельзя очертить достаточно четко: каждый специалист имеет право понимать под ними то, что ему заблагорассудится. Тем не менее, суммируя все то, о чем говорилось, сопоставляя различные трактовки этих терминов, даваемые специалистами в разных сферах, можно сказать то, что уже говорилось ранее. Именно, имитационная модель — это модель, обладающая качествами из следующего набора: «сложность» модели, наличие в ней случайных факторов, описание процесса, развивающегося во времени, невозможность получения результатов без ЭВМ, предназначенность модели для использования ее в режиме вариантных расчетов, т. е. для сравнения путем выполнения имитационных экспериментов, заданных заранее, «извне модели» вариантов планов, управлений, конструкций. Ни одно из перечисленных качеств не является обязательным для того, чтобы именовать модель имитационной. Например, специалист по оптимизации назовет даже очень простую детерминированную модель имитационной, если в ее рамках никакая оптимизационная задача не решается, а она используется в режиме вариантных расчетов для сравнения заранее сформированных альтернативных вариантов управлений. Тем не менее, перечисленные качества в совокупности дают целостное представление о том, какой смысл имеет понятие «имитационная модель» в современной русскоязычной научной литературе.

В англоязычной литературе терминам «имитация», «имитационная модель», «имитационный эксперимент» приблизительно соответствует термин «simulation». Наиболее известна трактовка термина «имитация», которую дал Р.Шеннон, определив ее как «процесс конструирования модели реальной системы и постановки экспериментов на этой модели с целью понять поведение системы либо оценить (в рамках ограничений, накладываемых некоторым критерием или совокупностью критериев) различные стратегии, обеспечивающие функционирование данной системы».

Аналогичное определение этому термину дает Т. Нейлор: «численный метод проведения на цифровых вычислительных машинах экспериментов с математическими моделями, описывающими поведение сложных систем в течение продолжительных периодов времени». Обе приведенные трактовки термина «simulation» укладываются в то содержание терминов «имитация», «имитационная модель», «имитационный эксперимент», которое было очерчено выше.

### **8.6. Вопросы для самопроверки**

1. В чем состоит сущность метода статистических испытаний?
2. В каких случаях оправдано применение метода статистических испытаний?
3. Приведите описание основных способов организации единичного жребия
4. Назовите приемы построения и эксплуатации дискретных имитационных моделей.
5. Укажите методы сбора информации в процессе имитационного моделирования и требования к наблюдениям в имитационном эксперименте.
6. Назовите достоинства и недостатки способов создания компьютерных имитационных моделей на универсальных и специализированных языках?
7. Каково современное содержание терминов «имитация», «имитационная модель»?

## ГЛАВА 9. НЕЧЕТКОЕ ОПИСАНИЕ СИТУАЦИЙ

Все известные измерительные шкалы (номинальная, порядковая, интервалов, отношений) основаны на справедливости отношения эквивалентности. В действительности встречаются случаи, когда тождество или различие двух состояний и/или наблюдений нельзя утверждать с полной уверенностью, например высокий молодой человек, довольно тяжелый сверток. Почти каждое наше слово обозначает некоторое не вполне определенное множество. Неопределенность смысла языковых конструкций – одна из основных трудностей автоматизации анализа и синтеза речи.

Роль человека в современных системах управления является определяющей: он выступает генератором целей системы и альтернативных путей её развития, определяет реальную структуру системы и формирует её поведение. Наиболее сложным и ответственным этапом деятельности человека в системах управления и главным фактором всякого руководства и управления считается принятие решений. Моделирование процессов принятия решений сегодня становится центральным направлением автоматизации деятельности лица, принимающего решение (ЛПР). Задачи принятия сложных решений ЛПР формирует и обсуждает на профессиональном языке, отражающем специфику задач. Следствием этого является использование в процессе поиска наилучшего решения качественных элементов: понятий и отношений с нечеткими границами, высказываний с многозначной шкалой истинности. Построение моделей принятия решений для задач, имеющих нечеткое словесное описание, оказалось возможным благодаря введению понятий нечеткого множества и лингвистической переменной.

Объект может принадлежать к классу, описываемому данным понятием, отношением или высказыванием, может, не относится к нему, но возможно и промежуточные градации принадлежности. Понятия и отноше-

ния, описывающие такие классы, будем называть нечеткими. Обобщение классического понятия множества для более корректного и полного использования нечетких описаний процессов принятия решений привело к понятию нечеткое множество. Теория нечетких множеств является средством формализации нечетких понятий и отношений.

### **9.1. Основные понятия теории нечетких множеств.**

Нечеткое множество (fuzzy set) представляет собой совокупность элементов произвольной природы, относительно которых нельзя с полной уверенностью утверждать – принадлежит ли тот или иной элемент рассматриваемой совокупности данному множеству или нет. Другими словами, нечеткое множество отличается от обычного множества тем, что для всех или части его элементов не существует однозначного ответа на вопрос: «Принадлежит или не принадлежит тот или иной элемент рассматриваемому нечеткому множеству?» Можно этот вопрос задать и по-другому: «Обладают или нет элементы некоторым характеристическим свойством, которое может быть использовано для задания этого нечеткого множества?»

Для построения нечетких моделей систем само понятие нечеткого множества следует определить более строго, чтобы исключить неоднозначность толкования тех или иных его свойств. Оказалось, что существуют несколько вариантов формального определения нечеткого множества, которые по сути отличаются между собой способом задания характеристической функции данных множеств. Среди этих вариантов наиболее естественным и интуитивно понятным является задание области значений подобной функции как интервала действительных чисел, заключенных между 0 и 1 (включая и сами эти значения).

Формально нечеткое множество  $A$  определяется как множество упорядоченных пар или кортежей вида:  $\langle x, \mu_A(x) \rangle$ , где  $x$  является элементом некоторого универсального множества или универсума  $X$ , а  $\mu_A(x)$  – функция принадлежности, которая ставит в соответствие каждому из элементов  $x \in X$  некоторое действительное число из интервала  $[0, 1]$ , т. е. данная функция определяется в форме отображения:

$$\mu_A: X \rightarrow [0, 1].$$

При этом значение  $\mu_A(x)=1$  для некоторого  $x \in X$  означает, что элемент  $x$  определенно принадлежит нечеткому множеству  $A$ , а значение  $\mu_A(x)=0$  означает, что элемент  $x$  определенно не принадлежит нечеткому множеству  $A$ .

Формально конечное нечеткое множество будем записывать в виде:

$A = \{ \langle x_1, \mu_A(x_1) \rangle, \langle x_2, \mu_A(x_2) \rangle, \dots, \langle x_n, \mu_A(x_n) \rangle \}$ , а в общем случае – в виде:

$$A = \{ \langle x, \mu_A(x) \rangle \}.$$

В теории нечетких множеств сохраняют свой смысл некоторые специальные классические множества. Так, например, пустое нечеткое множество или множество, которое не содержит ни одного элемента, по-прежнему обозначается через  $\emptyset$  и формально определяется как такое нечеткое множество, функция принадлежности которого тождественно равна нулю для всех без исключения элементов:  $\mu_{\emptyset} = 0$ .

Что касается другого специального множества, то так называемый универсум, обозначаемый через  $X$ , уже был использован выше в качестве обычного множества, содержащего в рамках некоторого контекста все возможные элементы. Формально удобно считать, что функция принадлежности универсума как нечеткого множества тождественно равна единице для всех без исключения элементов:  $\mu_X = 1$ .

Носителем нечеткого множества  $A$  называется обычное множество  $A_S$ , которое содержит те и только те элементы универсума, для которых значения функции принадлежности соответствующего нечеткого множества от-

личны от нуля. Математически носитель нечеткого множества определяется следующим условием:

$$A_s = \{x \in X \mid \mu_A(x) > 0\} \quad \forall x \in X. \quad (9.1)$$

В зависимости от количества элементов в нечетком множестве по аналогии с обычными множествами можно определить конечные и бесконечные нечеткие множества.

Нечеткое множество называется конечным, если его носитель является конечным множеством. При этом вполне уместно говорить, что такое нечеткое множество имеет конечную мощность, которая численно равна количеству элементов его носителя как обычного множества.

Аналогичным образом можно определить и бесконечные нечеткие множества как такие нечеткие множества, носитель которых не является конечным множеством. Чтобы привести некоторые примеры нечетких множеств и приступить к определению их основных свойств, следует рассмотреть основные способы, которыми формально могут быть заданы произвольные нечеткие множества.

Нечеткие множества могут быть заданы двумя основными способами:

1. В форме *списка* с явным перечислением всех элементов и соответствующих им значений функции принадлежности, образующих рассматриваемое нечеткое множество. При этом зачастую элементы с нулевыми значениями функции принадлежности просто не указываются в данном списке. Этот способ подходит для задания нечетких множеств с конечным дискретным носителем и небольшим числом элементов. В этом случае нечеткое множество удобно записывать в виде:  $A = \{ \langle x_1, \mu_A(x_1) \rangle, \langle x_2, \mu_A(x_2) \rangle, \dots, \langle x_n, \mu_A(x_n) \rangle \}$ , где  $n$  – рассматриваемое число элементов нечеткого множества  $A$  (его носителя). Например, возьмем в качестве универсума  $X = \{1, 2, 3, \dots\}$  – множество натуральных чисел. Тогда нечеткое множество  $A$ , представляющее в некотором контексте «небольшое натуральное чис-

ло», можно задать следующим образом:  $A = \{ \langle 1, 1.0 \rangle, \langle 2, 1.0 \rangle, \langle 3, 0.9 \rangle, \langle 4, 0.8 \rangle, \langle 5, 0.6 \rangle, \langle 6, 0.5 \rangle, \langle 7, 0.4 \rangle, \langle 8, 0.2 \rangle, \langle 9, 0.1 \rangle \}$ . При этом элементы, для которых  $\mu_A(x) = 0$  отсутствуют в этом списке.

В литературе по теории нечетких множеств, которая исчисляется огромным количеством работ, можно встретить не только различные определения, но и разнообразные обозначения для нечетких множеств. Мы также будем в дальнейшем для записи дискретного нечеткого множества применять следующую форму:  $A = \left\{ \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n} \right\}$ . При этом горизонтальная черта служит просто разделителем, а знак «+» обозначает не арифметическую сумму, а теоретико-множественное объединение отдельных элементов.

2. *Аналитически* в форме математического выражения для соответствующей функции принадлежности. Этот способ может быть использован для задания произвольных нечетких множеств как с конечным, так и с бесконечным носителем. В этом случае нечеткое множество удобно записывать в виде:  $A = \{ \langle x, \mu_A(x) \rangle \}$  или  $A = \{ x, \mu_A(x) \}$ , где  $\mu_A$  – некоторая функция, заданная аналитически в форме математического выражения  $f(x)$  или графически в форме некоторой кривой.

Для формальной строгости при задании нечетких множеств необходимо явно указать соответствующий универсум  $X$  элементов, из которых формируется то или иное конкретное нечеткое множество. В общем случае никаких предположений относительно элементов этого множества не делается. Однако с практической точки зрения целесообразно ограничить универсум элементами рассматриваемой предметной области или решаемой задачи. Поскольку при построении нечетких моделей систем используются количественные переменные, то наиболее часто в качестве универсума  $X$  используется некоторое подмножество действительных чисел, например, множество неотрицательных действительных чисел  $R^+$  или натуральных чисел  $N$ .

В качестве конкретного примера нечеткого множества рассмотрим типичную бытовую ситуацию, с которой сталкиваются многие из нас при попытке дать характеристику температуры того или иного напитка. Подобная характеристика обычно основывается исключительно на субъективных ощущениях, например, горячий кофе или чай, холодный квас или кола. Хотя в этом случае неявно используется некоторая шкала температуры, при этом, как правило, не применяется никаких измерительных инструментов.

Применительно к данной ситуации рассмотрим нечеткое множество  $B$ , которое будет характеризовать «горячий кофе». В этом случае в качестве универсума естественно взять шкалу температуры, измеренной в градусах Цельсия и заключенной в открытом интервале  $(0^{\circ}\text{C}, 100^{\circ}\text{C})$ , т. е.  $X = \{x | 0^{\circ}\text{C} < x < 100^{\circ}\text{C}\}$ . Выбор этого интервала вполне оправдан с физической точки зрения, поскольку именно в этом диапазоне температур кофе может потенциально существовать как напиток. Очевидно, что отдельная чашка кофе, скажем  $x_1$ , с температурой  $10^{\circ}\text{C}$  не может быть признана горячей, поэтому для нее значение функции принадлежности рассматриваемому множеству  $B$  будет равно 0, т. е.  $\mu_B(x_1) = 0$ . С другой стороны, другая чашка кофе  $x_2$  с температурой  $90^{\circ}\text{C}$  вполне может быть признана горячей, поэтому для нее значение функции принадлежности рассматриваемому множеству  $B$  будет равно 1, т. е.  $\mu_B(x_2) = 1$ . Что касается значений температур, заключенных между этими крайними значениями, то ситуация представляется уже не столь однозначной. Более того, она по своей сути является исключительно субъективной и неопределенной, поскольку чашка кофе с температурой  $55^{\circ}\text{C}$  для одного индивидуума может оказаться горячей, а для другого – не слишком горячей. Именно в этом и проявляется нечеткость задания соответствующего множества. Тем не менее, мы можем быть вполне уверены в общем виде функции принадлежности, а именно – в том, что со-

ответствующая функция принадлежности является монотонно возрастающей (или более строго – монотонно неубывающей).

Таким образом, в качестве множества  $V = \{x, \mu_V(x)\}$ , описывающего горячий кофе, можно рассматривать, например, такое нечеткое множество, для которого функция принадлежности имеет следующий вид (см. рис. 9.1, а и/или рис. 9.1, б).

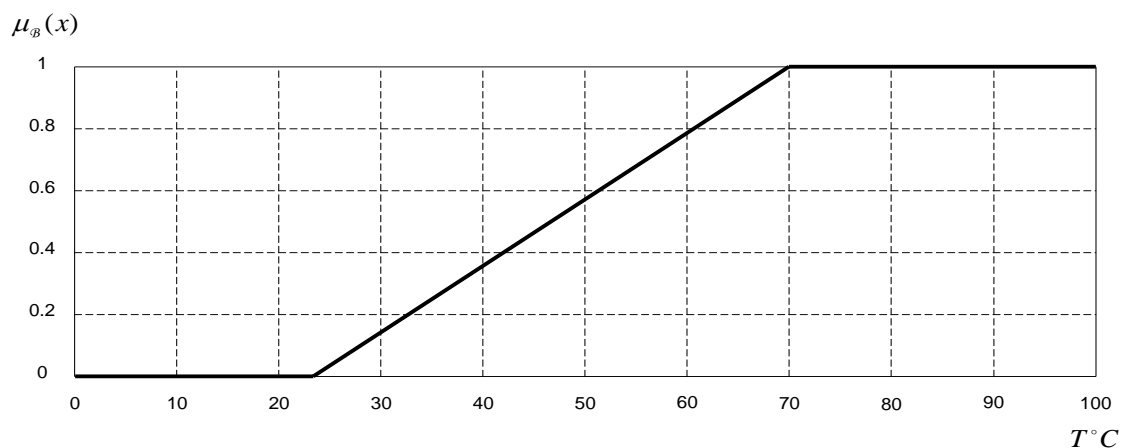
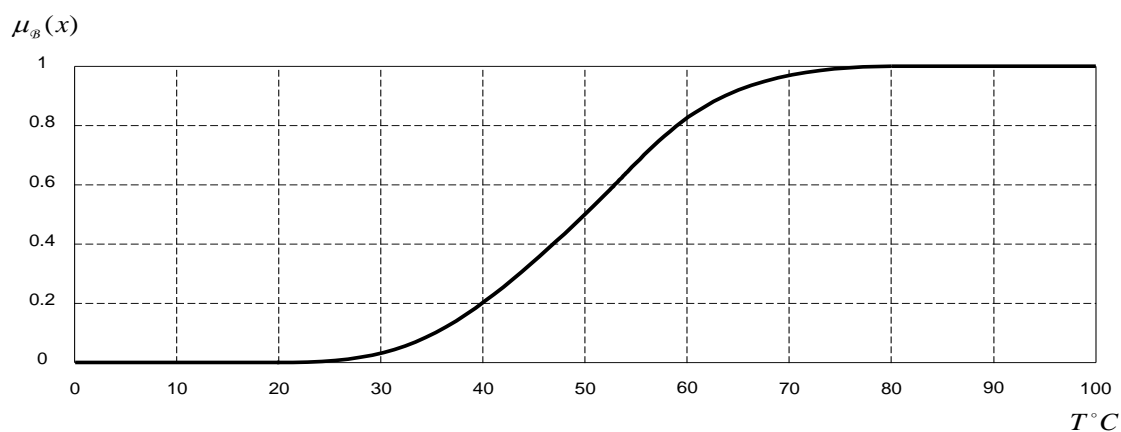


Рис. 9.1. Графики вариантов функций принадлежности для нечеткого множества  $V$ , описывающего «горячий кофе»

Рассмотренный пример допускает обобщения на другие ситуации, связанные с представлением аналогичной нечеткой информации. В частности, целый ряд свойств технических устройств, бытовых приборов и социальных явлений могут инициировать похожие нечеткие множества. Например, такие фразы, как «скоростной автомобиль», «высокооплачиваемая работа», «благоустроенная квартира», «щедрый чаевые», «престижный район»,

«вкусный ужин» порождают нечеткие множества, аналогичные рассмотренному в примере. При этом общий вид функций принадлежности таких множеств будет подобен изображенным на рис. 9.1, а, б.

Характерные аспекты неопределенности, которые встречаются в практике нечеткого моделирования. Во-первых, каждое из нечетких множеств допускает в общем случае неоднозначное представление, что отражает субъективную точку зрения на моделирование соответствующих практических ситуаций. Во-вторых, концептуальное различие между теорией нечетких множеств и теорией вероятностей, поскольку рассмотренные варианты неопределенности имеют не стохастический характер. И, наконец, в-третьих, выбор аналитической функции или вида кривой для той или иной функции принадлежности с целью задания соответствующего нечеткого множества зачастую определяется соображениями удобства и простоты.

Перейдем к рассмотрению основных характеристик нечетких множеств, которые используются для их более детального описания и систематического изучения.

Пусть  $A = \{x, \mu_A(x)\}$  – произвольное нечеткое множество (конечное или бесконечное) с элементами из универсума  $X$  и функцией принадлежности  $\mu_A(x)$ .

Обобщением носителя нечеткого множества является понятие *множества  $\alpha$ -уровня*, под которым понимается обычное множество  $A_\alpha$ , удовлетворяющее следующему условию:  $A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$ , где  $\alpha$  – некоторое действительное число из интервала  $[0, 1]$ , т. е.  $\alpha \in [0, 1]$ .

Величина  $h_A = \sup\{\mu_A(x)\}$ , где супремум берется по всем значениям функции принадлежности для  $x \in X$ , называется высотой нечеткого множества  $A$ . Согласно этому определению, нечеткое множество  $A$  пусто, если его высота в точности равна 0, т. е.  $h_A = 0$ .

Особенность определения высоты заключается в том, что высота нечеткого множества всегда существует и равна некоторому действительному числу из интервала  $[0, 1]$ , которому может соответствовать несколько элементов универсума.

При определении высоты нечеткого множества использована специальная функция  $y=\sup(f)$ , которая получила свое название от латинского *supremum* – наивысшее и называется верхней гранью (или наименьшей верхней гранью). Как будет видно из дальнейших рассуждений, эта функция отличается от похожей на нее функции  $\max(f)$ . Формально функция  $y=\sup(f)$  определяется для обычных множеств следующим образом. Рассмотрим произвольное отображение  $f: D \rightarrow R$ , где  $D$  – область определения этого отображения ( $D \subseteq X$ ). Отображение  $f$  называется ограниченным сверху (снизу) на множестве  $D$ , если существует конечное число  $k \in R$ , такое что выполняется условие:  $f(x) \leq k$  (соответственно,  $f(x) \geq k$ ) ( $\forall x \in D$ ). При этом отображение  $f(x)$  называется ограниченным на множестве  $D$ , если оно одновременно ограничено на  $D$  сверху и снизу. Далее рассмотрим некоторое ограниченное сверху отображение  $f(x)$ , для которого  $D \subseteq R$ , т.е. ограниченную сверху функцию  $f(x)$ . В общем случае числовые значения  $k \in R$ , для которых выполняется условие:  $f(x) \leq k, \forall x \in D$ , образуют некоторое числовое подмножество  $U \subseteq R$ , при этом очевидно, что  $U \neq \emptyset$ . Если среди всех  $k \in U$  найдется некоторое наименьшее значение, обозначим его через  $k_s$ , то оно называется наименьшей верхней гранью функции  $f(x)$  на множестве  $D \subseteq R$  и обозначается через  $k_s = \sup(f)$  (читается «супремум  $f$  на множестве  $D$ »). Очевидно, если  $k_s \in \text{Im}_f$ , то это значение одновременно является и максимумом функции  $f(x)$  на множестве  $D \subseteq R$ , т. е.  $k_s = \max(f)$  и тогда значения этих двух функций  $\sup(f)$  и  $\max(f)$  совпадают. С другой стороны, может оказаться, что в множестве  $\text{Im}_f$  не существует такого  $k_s$ , для которого выполнялось бы условие:  $f(x) \leq k_s, \forall x \in D$ , т.е.  $k_s \notin \text{Im}_f$ . Введение в рассмотрение функции  $\sup(f)$  всегда гарантирует существование такого  $k_s$  в множестве действительных чисел  $R$ , поскольку последнее является непрерывным и вполне упорядоченным множеством. Следует также отметить, что функции  $\sup(f)$  и  $\max(f)$  всегда можно определить на множестве значений рассматриваемых функций  $f$ , т. е. на  $\text{Im}_f$ . Таким образом, поскольку область значений любой функции принадлежности ограничена интервалом  $[0, 1]$ , высота произвольного нечеткого множества всегда существует и это числовое значение принадлежит интервалу  $[0, 1]$ .

Нечеткое множество  $A$  называется *нормальным*, если максимальное значение его функции принадлежности равно 1. Формально это означает, что для нормального нечеткого множества необходимо выполнение следующего условия:

$$\mu_A(x) = 1, (\exists x \in X). \quad (9.2)$$

Если высота нечеткого множества равна единице ( $h_A=1$ ), но условие (9.2) не выполняется, то такое нечеткое множество будем называть субнормальным.

Нечеткое множество  $A$  называется унимодальным (строго унимодальным), если его функция принадлежности  $\mu_A(x)$  является унимодальной (строго унимодальной).

В свою очередь произвольная функция принадлежности  $\mu(x)$  называется унимодальной на интервале  $[a, b] \subset \mathbb{R}$ , если она непрерывна на  $[a, b]$ , а также существует некоторый непустой  $[c, d] \subset [a, b]$ , такой что  $a \leq c \leq d \leq b$  и выполняются следующие условия:

- функция  $\mu(x)$  строго монотонно возрастает на интервале  $[a, c]$  при  $a < c$ ;
- функция  $\mu(x)$  строго монотонно убывает на интервале  $[d, b]$  при  $d < b$ ;
- функция  $\mu(x)$  принимает свое максимальное значение на интервале  $[c, d]$ , т.е. любая точка  $x_m \in [c, d]$  является точкой максимума функции принадлежности относительно интервала  $[a, b]$ :

$$x_m = \arg \max_{x \in [a, b]} \{\mu(x)\}. \quad (9.3)$$

В этом случае любая точка  $x_m \in A$  нечеткого множества  $A$ , удовлетворяющая условию (9.3), называется модальным значением или модой нечеткого множества  $A$ .

Если в этом определении интервал  $[c, d]$  вырождается в точку, т. е.  $c=d$ , то соответствующая функция принадлежности называется строго унимодальной на интервале  $[a, b]$ .

Функция принадлежности  $\mu_A(x)$  называется унимодальной (строго унимодальной), если она унимодальна (строго унимодальна) на носителе соответствующего нечеткого множества  $A$ .

Ядром нечеткого множества  $A$  называется такое обычное множество  $A_1$ , элементы которого удовлетворяют условию:  $A_1 = \{x \in X \mid \mu_A(x) = 1\}$ .

Нетрудно заметить, что если произвольное нечеткое множество не является нормальным, то ядро такого нечеткого множества будет пустым. Таким образом, имеет место следующая фундаментальная теорема. Для то-

го, чтобы некоторое нечеткое множество было нормальным, необходимо и достаточно, чтобы оно имело непустое ядро.

Поскольку, как было показано выше, высота нечеткого множества всегда существует, то произвольное непустое нечеткое множество  $A$  всегда можно преобразовать по меньшей мере к субнормальному нечеткому множеству  $A'$  по следующей формуле:

$$\mu_{A'}(x) = \frac{\mu_A(x)}{h_A} . \quad (9.4)$$

Более того, если в исходном нечетком множестве  $A$  найдется хотя бы один элемент  $x \in A$ , для которого значение функции принадлежности равно высоте этого нечеткого множества, т. е.  $\mu_A(x)=h_A$ , то полученное после преобразования (9.4) нечеткое множество  $A'$  будет нормальным.

Границами нечеткого множества называются такие элементы универсума, для которых значения функции принадлежности отличны от 0 и 1. Другими словами, границы нечеткого множества  $A = \{x, \mu_A(x)\}$  включают те и только те элементы универсума  $x \in X$ , для которых выполняется условие:  $0 < \mu_A(x) < 1$ .

Элементы нечеткого множества  $y \in A$ , для которых выполняется условие:  $\mu_A(y)=0.5$ , называются точками перехода этого нечеткого множества  $A$ .

Для характеристики нечетких множеств используют также понятие выпуклости, которое ассоциируется с соответствующим графическим изображением функции принадлежности.

Нечеткое множество  $A = \{x, \mu_A(x)\}$  с универсумом  $X$  называют выпуклым, если его функция принадлежности  $\mu_A(x)$  удовлетворяет следующему неравенству:  $\mu_A(x) \geq \min \{\mu_A(a), \mu_A(b)\}$  для любых значений  $x, a, b \in X$ , при которых  $a < x < b$  и  $a \neq b$ .

## 9.2. Операции над нечеткими множествами

По аналогии с обычными множествами, прежде всего, определим два простейших обычных отношения, которые могут иметь место между двумя произвольными нечеткими множествами  $A$  и  $B$ , заданными на одном и том же универсуме  $X$ . Первое из них – равенство двух нечетких множеств.

*Равенство нечетких множеств.* Два нечётких множества  $A = \{x, \mu_A(x)\}$  и  $B = \{x, \mu_B(x)\}$  считаются равными, если функции принадлежности принимают равные значения на всем универсуме  $X$ :

$$\mu_A(x) = \mu_B(x) \text{ для любого } x \in X. \quad (9.5)$$

Равенство множеств в данном случае записывается как  $A=B$ .

Следующим простейшим отношением является понятие нечеткого подмножества (или нечеткого доминирования) произвольных нечетких множеств. Формально это определение также записывается с помощью соответствующих функций принадлежности.

*Нечеткое подмножество.* Нечеткое множество  $A = \{x, \mu_A(x)\}$ , является нечётким подмножеством нечеткого множества  $B = \{x, \mu_B(x)\}$  (записывается как  $A \subseteq B$ ) тогда и только тогда, когда значения функции принадлежности первого не превосходят соответствующих значений функции принадлежности второго, т.е. выполняется следующее условие:

$$\mu_A(x) \leq \mu_B(x) \quad (\forall x \in X). \quad (9.6)$$

Также как и для обычных множеств, для обозначения нечёткого подмножества используется символ " $\subseteq$ ". При этом в случае  $A \subseteq B$  часто говорят, что нечеткое множество  $B$  доминирует нечеткое множество  $A$ , а нечеткое множество  $A$  содержится в нечетком множестве  $B$ .

По аналогии с классическими множествами среди нечетких множеств можно различать два различных варианта доминирования. Рассмотренное выше определение характерно для таких называемого несобственного подмножества, когда не исключается

случай возможного равенства двух нечётких множеств  $A$  и  $B$ . Если же в определении нечеткого подмножества исключается равенство соответствующих нечётких множеств в форме (9.5), то в этом случае  $A$  называется собственным нечетким подмножеством  $B$  и обозначается:  $A \subset B$ . При этом часто говорят, что нечеткое множество  $B$  строго доминирует нечеткое множество  $A$ , а нечеткое множество  $A$  строго содержится в нечетком множестве  $B$ .

Если для двух нечетких множеств  $A$  и  $B$ , заданных на одном универсуме, не выполняется ни отношение  $A \subseteq B$ , ни отношение  $B \subseteq A$ , то в этом случае говорят, что нечеткие множества  $A$  и  $B$  несравнимы.

Нечеткое доминирование или факт включения элементов одного нечеткого множества в другое нечеткое множество можно изобразить графически в декартовой системе координат на плоскости. С этой целью изобразим прямоугольную систему координат, на оси абсцисс которой в том или ином порядке расположим элементы универсума  $X$ , а на оси ординат – соответствующие им значения функции принадлежности рассматриваемого нечеткого множества.

Для случая  $B \subseteq A$  график функции принадлежности нечеткого множества  $B$  будет расположен по вертикальной оси не выше графика функции принадлежности нечёткого множества  $A$  (рис. 9.2 (а, б)). Более того как на рис. 9.2 (а), так и на рис. 9.2 (б), изображены случаи строгого доминирования  $B \subset A$ .

Пусть  $A$  и  $B$  – произвольные (конечные или бесконечные) нечеткие множества, заданные на одном, и том же универсуме  $X$ .

*Пересечение.* Пересечением двух нечетких множеств  $A$  и  $B$  будем называть некоторое третье нечеткое множество  $C$ , заданное на этом же универсуме  $X$ , функция принадлежности которого определяется по следующей формуле:

$$\mu_C(x) = \min\{\mu_A(x), \mu_B(x)\} (\forall x \in X). \quad (9.7)$$

Операция пересечения нечетких множеств по аналогии с обычными множествами обозначается знаком " $\cap$ ". В этом случае результат операции пересечения двух нечетких множеств записывается в виде:  $C=A \cap B$ .

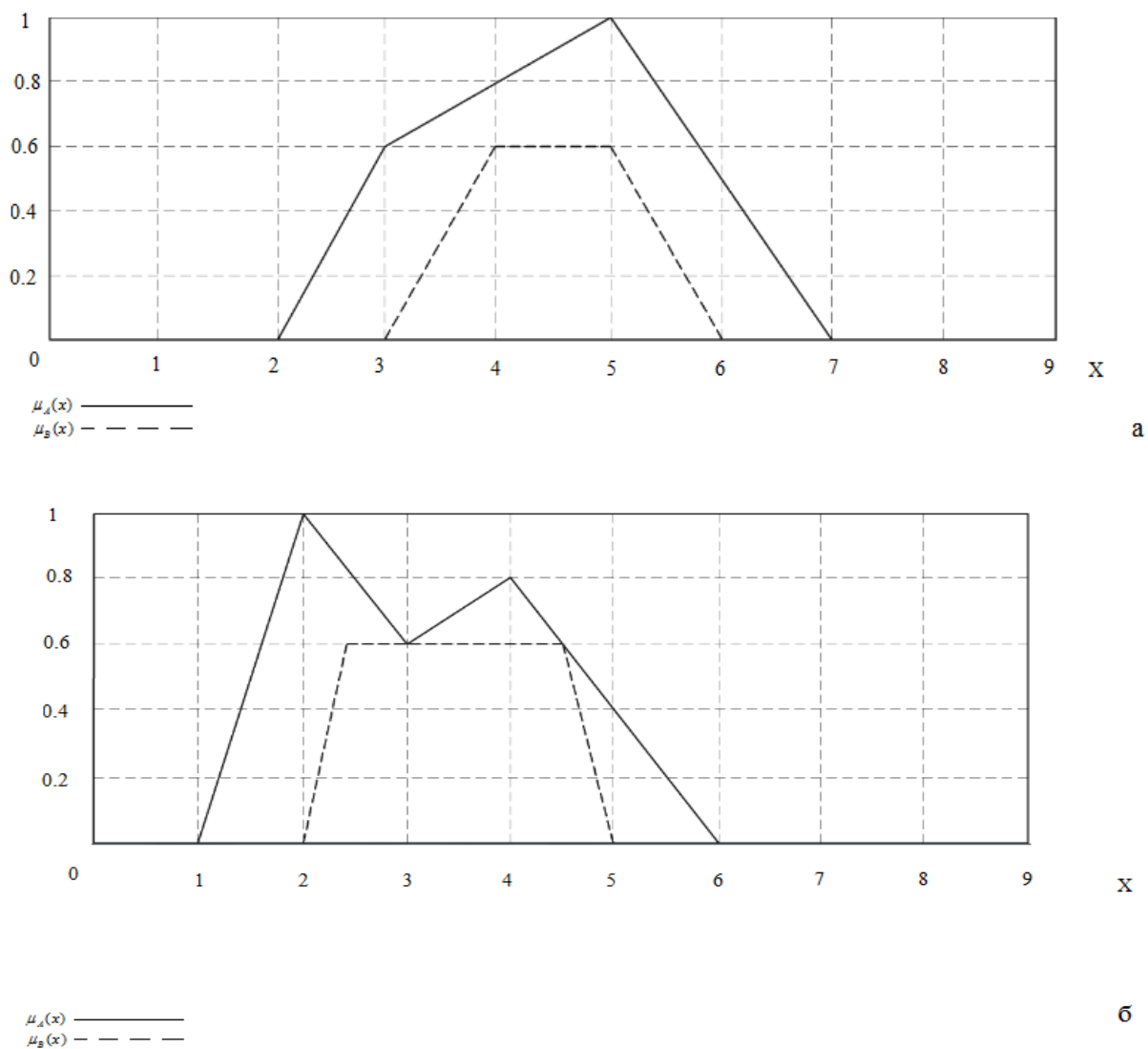


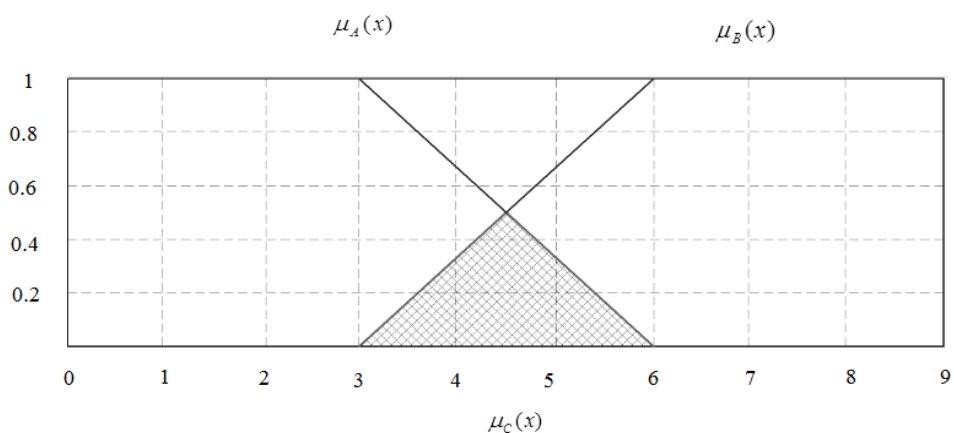
Рис. 9.2. Различные варианты отношения доминирования  $B \subseteq A$  двух нечетких множеств

Как не трудно заметить, пересечение  $A \cap B$  есть наибольшее нечеткое подмножество  $C$ , которое содержится одновременно в нечетких множествах  $A$  и  $B$ . Операцию пересечения нечетких множеств в смысле (9.7) иногда называют  $\min$  – пересечением или  $\wedge$ - пересечением. Последнее обо-

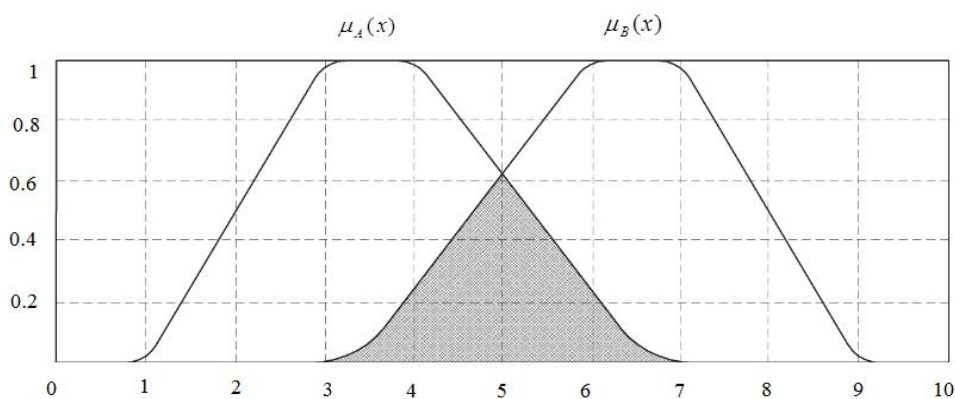
значение связано с определением логической операции "И" – которая в математической логике обозначается знаком " $\wedge$ ".

Результат операции пересечения двух и большего числа нечетких множеств, заданных на одном и том же универсуме  $X$ , также можно изобразить графически в декартовой системе координат на плоскости. Этот способ особенно удобен для визуализации операций с бесконечными нечеткими множествами. В данном случае каждое из нечетких множеств изображается соответствующей функцией принадлежности, а функция принадлежности результата операции пересечения изображается утолщенной линией. Для дополнительной наглядности область, расположенная ниже значений результирующей функции принадлежности, изображается затемненной.

Для случая пересечения двух нечетких множеств  $A \cap B$ , заданных различными функциями принадлежности, результат операции изображен на рис. 9.3 (а, б).



а



б

Рис. 9.3. Графическое представление операции пересечения двух нечетких множеств А и В, заданных линейными Z-образной и S-образной функциями принадлежности (а) и П-образными (б) - функциями принадлежности

*Объединение.* Объединением двух нечетких множеств А и В называется некоторое третье нечеткое множество D, заданное на этом же универсуме X, функция принадлежности которого определяется по следующей формуле:

$$\mu_D(x) = \max\{ \mu_A(x), \mu_B(x) \} (\forall x \in X). \quad (9.8)$$

Операция объединения нечетких множеств по аналогии с обычными множествами обозначается знаком " $\cup$ ". В этом случае результат операции, объединения двух нечетких множеств записывается в виде:  $D = A \cup B$ .

Как нетрудно заметить, объединение  $A \cup B$  есть наименьшее нечеткое множество D, которое доминирует одновременно как А, так и В. Операцию объединения нечетких множеств в смысле (9.8) иногда называют  $\max$  – объединением или  $\vee$ -объединением. Последнее обозначение связано с определением логической операции "ИЛИ" (неисключающего ИЛИ), которая в математической логике обозначается знаком " $\vee$ ".

Результат операции объединения двух и большего числа нечетких множеств, заданных на одном и том же универсуме X, можно изобразить графически в декартовой системе координат на плоскости. Для случая

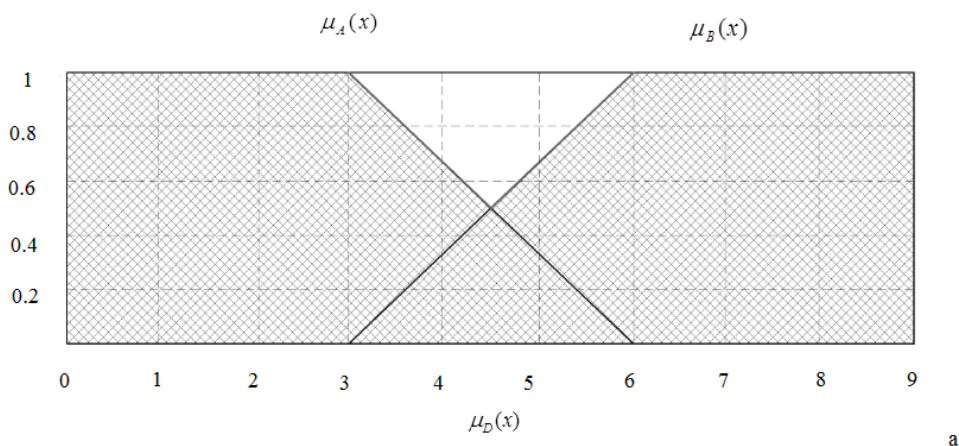
объединения двух нечетких множеств  $A \cup B$ , заданных различными функциями принадлежности, результат операции изображен на рис. 9.4 (а, б).

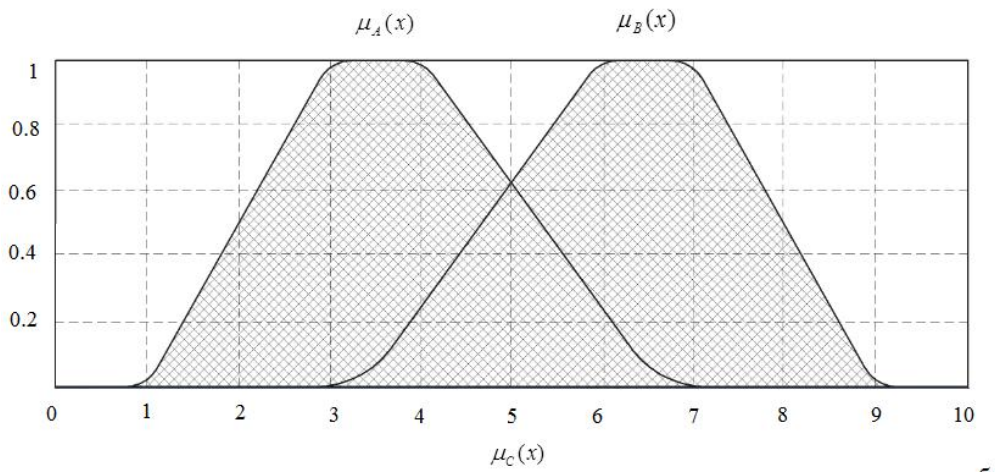
*Разность.* Разностью двух нечетких множеств  $A$  и  $B$  называется некоторое третье нечеткое множество  $E$ , заданное на этом же универсуме  $X$ , функция принадлежности которого определяется по следующей формуле:

$$\mu_E(x) = \max\{\mu_A(x) - \mu_B(x), 0\} (\forall x \in X), \quad (9.9)$$

где под знаком максимума используется обычная операция арифметической разности двух чисел. Операция разности двух нечетких множеств по аналогии с обычными множествами обозначается знаком "\". В этом случае результат операции разности двух нечетких множеств можно записать в виде:  $E=A \setminus B$ .

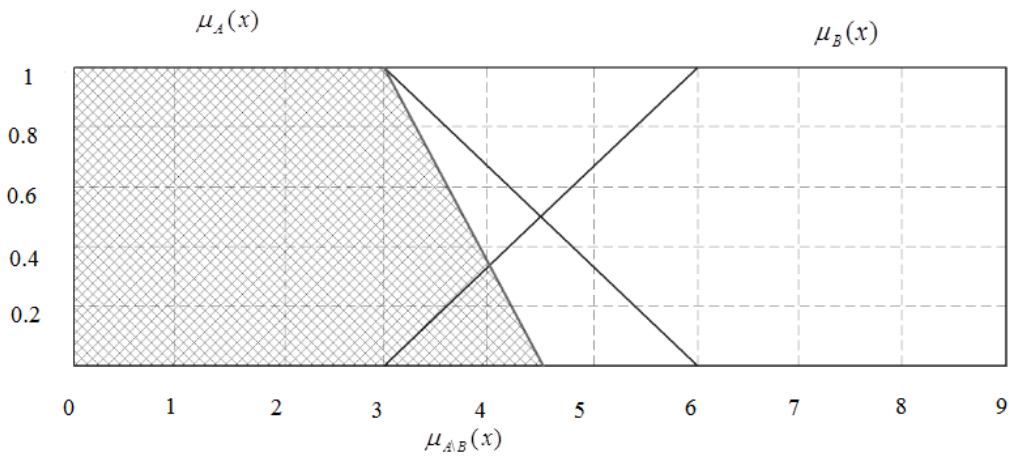
Результат операции разности двух нечетких множеств  $A$  и  $B$ , заданных на одном и том же универсуме  $X$  различными функциями принадлежности, изображен на рис. 9.5 (а, б).



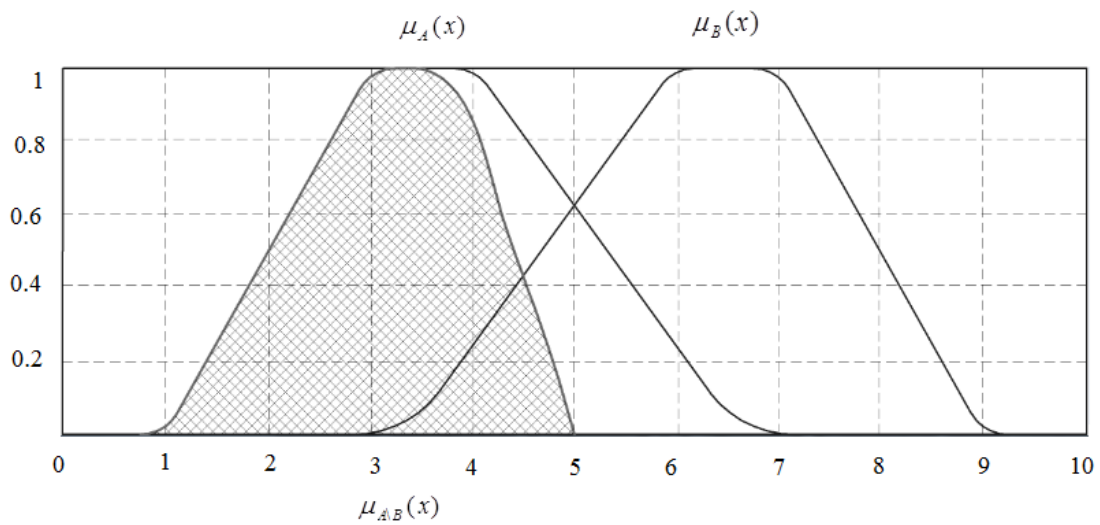


б

Рис. 9.4. Графическое представление операции объединения двух нечетких множеств А и В, заданных линейными Z-образной и S-образной функциями принадлежности (а) и П-образными (б) функциями принадлежности



а



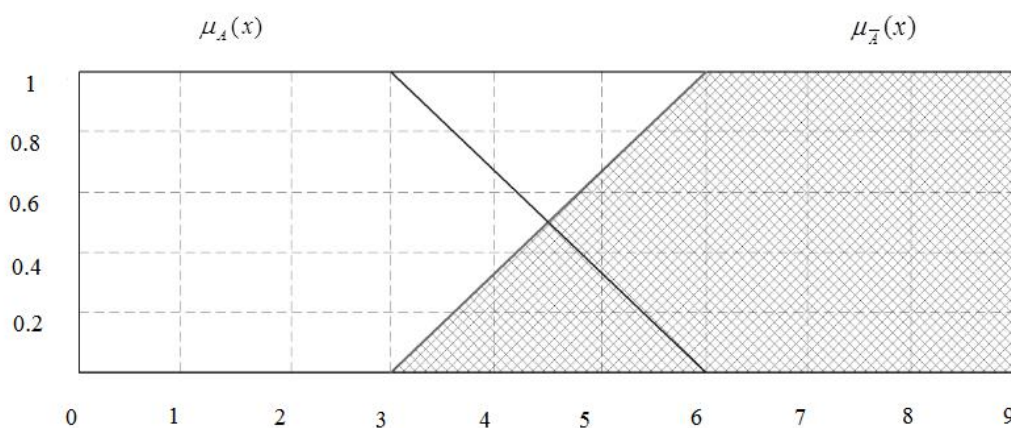
б

Рис. 9.5. Графическое, представление операции разности двух нечетких множеств  $A \setminus B$ , заданных линейными Z-образной и S-образной функциями принадлежности (а) и П-образными (б) функциями принадлежности

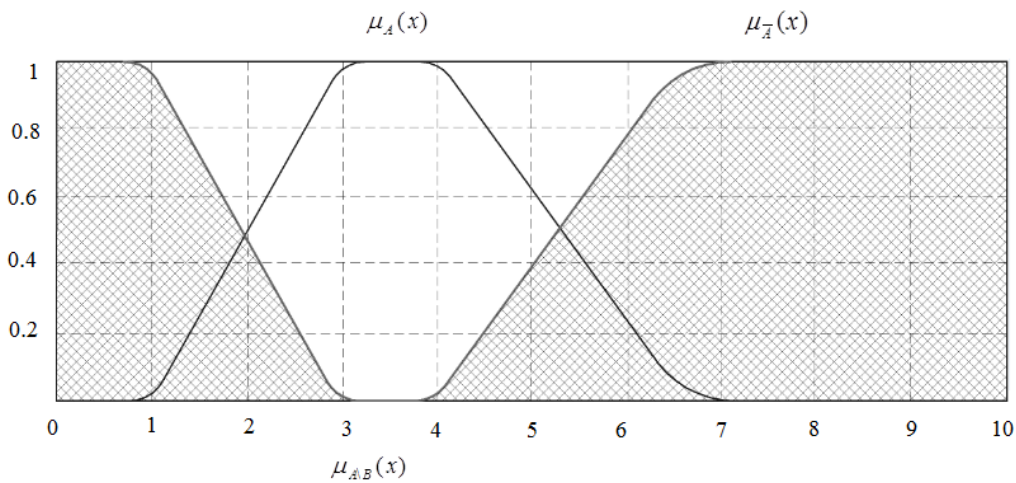
*Дополнение.* Дополнение нечеткого множества  $A$  обозначается через  $\bar{A}$  и определяется как нечеткое множество  $\bar{A} = \{x | \mu_{\bar{A}}(x)\}$ , функция принадлежности которого,  $\mu_{\bar{A}}(x)$  определяется по следующей формуле:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (\forall x \in X). \quad (9.10)$$

Операция дополнение нечёткого множества  $A$  может быть проиллюстрировано графически (рис. 9.6). При этом результату операции дополнения  $\bar{A}$  также соответствует более темная область на графике. Как нетрудно видеть, график функции принадлежности дополнения нечеткого множества симметричен графику функции принадлежности исходного нечёткого множества относительно линии:  $y=0.5$ .



а



б

Рис. 9.6. Графическое представление операции дополнения нечеткого множества  $A$ , которое задано линейной  $Z$ -образной (а) функцией принадлежности и  $\Pi$ -образной (б) функцией принадлежности

Для рассмотренных операции над нечеткими множествами имеют место следующие фундаментальные свойства, аналогичные свойствам, обычных теоретико-множественных операций.

Пусть  $A$ ,  $B$  и  $C$  произвольное (конечные или бесконечное) нечеткое множества, заданные на одном и том же универсуме  $X$ . Справедливы следующие утверждения:

Коммутативность операций объединения и пересечения нечетких множеств:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

Ассоциативность операций объединения и пересечения нечетких множеств:

$$A \cup (B \cap C) = (A \cup B) \cap C, \quad A \cap (B \cup C) = (A \cap B) \cup C.$$

Дистрибутивность операций объединения и пересечения нечетких множеств относительно друг друга:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Идемпотентность операций объединения и пересечения нечетких множеств:

$$A \cup A = A, \quad A \cap A = A.$$

Поглощение одного из нечетких множеств при операциях объединения и пересечения:

$$A \cup (A \cap B) = A \cap (A \cup B) = A.$$

Универсальные верхняя и нижняя границы (единичные элементы) операций пересечения и объединения нечетких множеств:

$$A \cup \emptyset = A \quad A \cup X = X$$

$$A \cap X = A \quad A \cap \emptyset = \emptyset .$$

Инволюция (двойное дополнение) нечёткого множества:

$$\overline{\overline{A}} = A .$$

Законы де Моргана:

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B} \quad \overline{(A \cap B)} = \overline{A} \cup \overline{B} .$$

Особенность рассматриваемых операций над нечеткими множествами состоит том, что для них не выполняется закон исключённого третьего и закон тождества (свойства дополняемости операций пересечения и объединения). А именно в общем случае оказываются справедливыми неравенства:

$$A \cap \overline{A} \neq \emptyset \quad A \cup \overline{A} \neq X .$$

Доказательство отмеченных свойств непосредственно следует из свойств операций минимума и максимума, используемых в определениях соответствующих нечетких операций.

Введенные в рассмотрение операции над нечеткими множествами, основанные на использовании операций  $\max(\bullet)$  и  $\min(\bullet)$ , получили наибольшее распространение при решении практических задач нечеткого моделирования. Эти операции обладают двумя основными достоинствами. Во-первых, они наиболее естественны для интуитивного представления неопределенности, связанной с использованием соответствующих логических связок "И", "ИЛИ", "НЕ". Во-вторых, они удовлетворяют перечисленным выше свойствам, что в максимальной степени приближает структуру нечетких множеств к булевой алгебре.

Тем не менее, операции  $\min$ -пересечения и  $\max$ -объединения нечетких множеств было бы неверно считать единственными, поскольку в общем случае возможны и другие альтернативные способы их определения. При этом большинство из подобных альтернативных операций также оказыва-

ются корректными в смысле соответствия обычным теоретико-множественным операциям.

Среди дополнительных операций, которые находят применение при построении нечетких моделей сложных систем, следует отметить унарные операции умножения нечеткого множества на число и возведение нечеткого множества в степень.

*Умножение нечеткого множества на число.* Пусть  $A = \{x, \mu_A(x)\}$  – произвольное нечеткое множество, заданное на универсуме  $X$ ;  $a$  – положительное действительное число ( $a \in \mathbb{R}^+$ ), такое, что,  $a \cdot h_A \leq 1$  (напомним,  $h_A$  – высота нечеткого множества  $A$ ). Результат операции умножения нечеткого множества  $A$  на число  $a$  определяется как нечеткое множество  $B = \{x, \mu_B(x)\}$ , заданное на этом же универсуме  $X$ , функция принадлежности которого определяется по формуле:

$$\mu_B(x) = a * \mu_A(x) (\forall x \in X). \quad (9.11)$$

Эту операцию в дальнейшем будем обозначать через  $a \cdot A$ .

*Возведение в степень.* Пусть  $A = \{x, \mu_A(x)\}$  произвольное нечеткое множество, заданное на универсуме  $X$ ;  $k$  – положительное действительное число ( $k \in \mathbb{R}^+$ ). В этом случае чисто формально можно определить операцию возведения нечеткого множества  $A$  в степень  $k$  как нечеткое множество  $B = \{x, \mu_B(x)\}$ , заданное на этом же универсуме  $X$ , функция принадлежности которого определяется по формуле:

$$\mu_B(x) = \mu_A(x)^k (\forall x \in X). \quad (9.12)$$

Эту операцию иногда обозначают через  $A^k$ .

Операции умножения бесконечного нечеткого множества на число и возведения бесконечного нечеткого множества в степень могут быть проиллюстрированы графически следующим образом (рис. 9.7).

На основе операции возведения в степень определяются две специальные операции над нечеткими множествами: операция концентрирования и операция растяжения нечеткого множества.

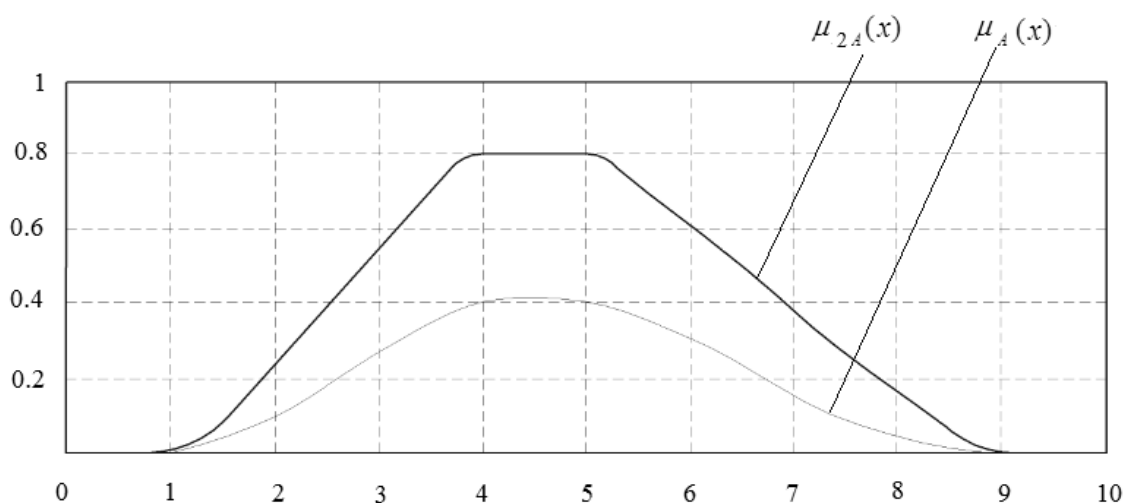
*Концентрирование.* Пусть на универсуме  $X$  задано произвольное нечеткое множество  $A = \{x, \mu_A(x)\}$ . Операция концентрирования, обозначаемая через  $CON(A)$ , дает в результате нечеткое множество  $C = \{x, \mu_C(x)\}$ , функция принадлежности которого равна значениям функции принадлежности исходного нечеткого множества, возведенным в квадрат, т. е.

$$\mu_C(x) = \mu_A(x)^2 \quad (\forall x \in X). \quad (9.13)$$

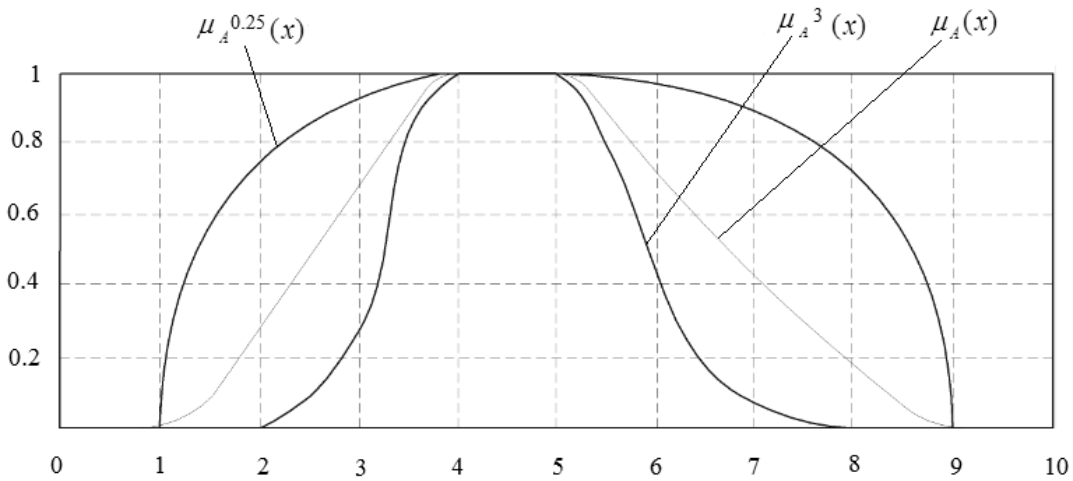
Очевидно, в этом случае  $CON(A) = A^2$ .

*Растяжение.* Операция растяжения, обозначаемая через  $DIL(A)$ , дает в результате нечеткое множество  $D = \{x, \mu_D(x)\}$ , функция принадлежности которого равна значениям функции принадлежности исходного нечеткого множества, возведенным в степень 0.5, т. е.

$$\mu_D(x) = \mu_A(x)^{0.5} \quad (\forall x \in X). \quad (9.14)$$



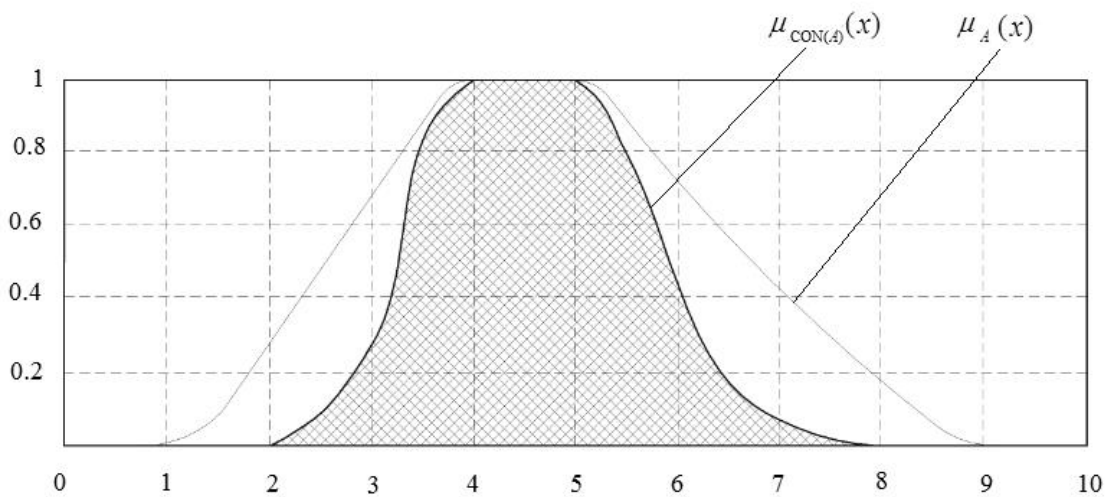
а



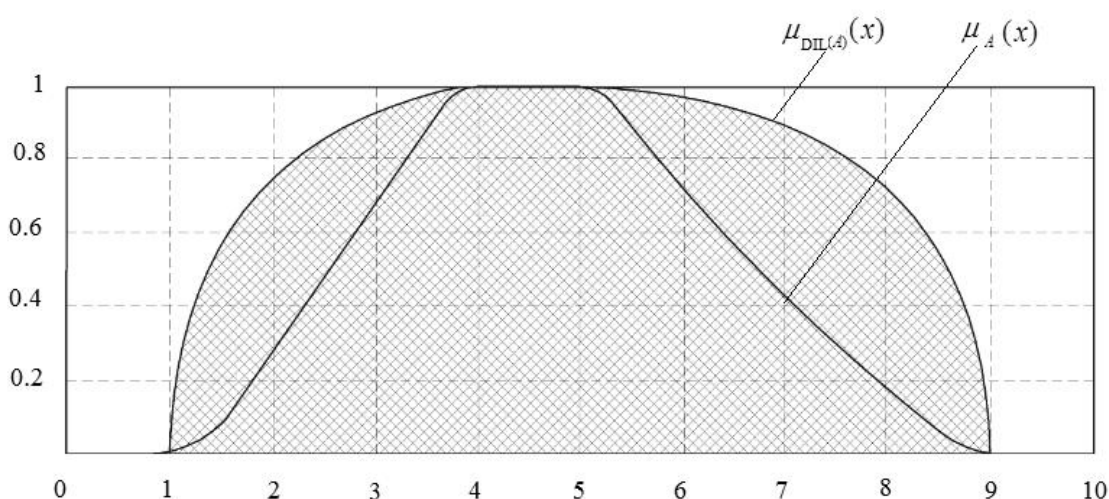
б

Рис. 9.7. Графическое представление операций умножения нечеткого множества  $A$  на число 2 (а) и возведения нечеткого множества  $A$  в степени 0.25 и 3 (б) для П-образных функций принадлежности, первая из которых не является нормальной

Очевидно, в этом случае  $DIL(A)=A^{0.5}$ . Операции концентрирования и растяжения нечеткого множества могут быть проиллюстрированы графически (см. рис. 9.8). Результатам этих операций соответствует более темная область на графике.



а



6

Рис. 9.8. Графическое представление операций концентрирования нечеткого множества (а) и растяжения нечеткого множества (б) для П-образных функций принадлежности

Применение операции концентрирования к нечёткому множеству означает уменьшение нечеткости или неопределённости в задании этого множества. Зачастую это может быть следствием поступления дополнительной информации. Напротив, применение операции растяжения означает усиление неопределенности в задании нечеткого множества, что может быть следствием либо потери части информации, либо поступления информации о дополнительных факторах, не учитываемых в исходной нечеткой модели. Эти операции находят применение при оперировании лингвистическими переменными, которые рассматриваются далее.

### 9.3. Нечеткие отношения

Понятие нечеткого отношения наряду с понятием самого нечеткого множества следует отнести к фундаментальным основам всей теории нечетких множеств. На основе нечетких отношений определяется целый ряд дополнительных понятий, используемых для построения нечетких моделей сложных систем. Нечеткое отношение обобщает понятие обычного отно-

шения и часто заменяется терминами нечеткая связь, ассоциация, взаимосвязь или соотношение.

Содержательно нечеткое отношение определяется как любое нечеткое подмножество упорядоченных кортежей, построенных из элементов тех или иных базисных множеств, в качестве которых в данном случае используются универсумы. При этом под кортежем, так же как и в случае обычных множеств, понимается произвольный набор или список упорядоченных элементов.

*Нечеткое отношение.* В общем случае нечетким отношением или, более точно, нечетким  $k$ -арным отношением, заданным на множествах (универсумах)  $X_1, X_2, \dots, X_k$ , называется некоторое фиксированное нечеткое подмножество декартова произведения этих универсумов. Другими словами, если обозначить произвольное нечеткое отношение через  $Q$ , то по определению  $Q = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_Q(\langle x_1, x_2, \dots, x_k \rangle) \}$ , где  $\mu_Q(\langle x_1, x_2, \dots, x_k \rangle)$  - функция принадлежности данного нечеткого отношения, которая определяется как отображение  $\mu_Q: X_1 \times X_2 \times \dots \times X_k \rightarrow [0, 1]$ . Здесь через  $\langle x_1, x_2, \dots, x_k \rangle$  обозначен кортеж из  $k$  элементов, каждый из которых выбирается из своего универсума:  $x_1 \in X_1, x_2 \in X_2, \dots, x_k \in X_k$ .

Так же как и в случае обычных множеств с целью характеризовать количество универсальных множеств, на основе которых строится то или иное нечеткое отношение, принято называть нечеткое отношение между элементами из двух универсальных множеств - бинарным, между элементами трех множеств - тернарным, а в общем случае -  $k$ -арным отношением. При этом на форму и вид функции принадлежности нечеткого отношения предварительно не накладываается никаких ограничений.

Особое значение в нечетком моделировании имеют бинарные нечеткие отношения, для задания которых используется одно или разные базисные множества (универсумы). В связи с этим приведем их формальное определение.

*Бинарное нечеткое отношение.* В общем случае бинарное нечеткое отношение задается на базисных множествах  $X_1, X_2$  и определяется как нечеткое отношение  $Q = \{ \langle x_i, x_j \rangle, \mu_Q(\langle x_i, x_j \rangle) \}$ . Здесь  $\mu_Q(\langle x_i, x_j \rangle)$  - функция принадлежности бинарного нечеткого отношения, которая определяется как отображение  $\mu_Q: X_1 \times X_2 \rightarrow [0, 1]$ , а через  $\langle x_i, x_j \rangle$  обозначен кортеж из двух элементов, при этом  $x_i \in X_1, x_j \in X_2$ .

Существуют различные способы, которыми в общем случае могут быть формально заданы те или иные нечеткие отношения. Наибольшее распространение из них получили следующие.

В форме *списка* с явным перечислением всех кортежей нечеткого отношения и соответствующих им значений функции принадлежности:  $Q = \{ (w_1, \mu_Q(w_1)), (w_2, \mu_Q(w_2)), \dots, (w_q, \mu_Q(w_q)) \}$ , где  $w_i$  -  $i$ -й кортеж  $\langle x_1, x_2, \dots, x_k \rangle$  элементов этого отношения, а  $q$  - рассматриваемое число кортежей нечеткого отношения  $Q$ . При этом для сокращения подобной записи кортежи с нулевыми значениями функции принадлежности не указываются в данном списке. Как нетрудно заметить, этот способ подходит только для задания нечетких отношений с конечным и небольшим числом кортежей  $q$ .

*Аналитически* в форме некоторого математического выражения для соответствующей функции принадлежности этого нечеткого отношения. Этот способ может быть использован для задания произвольных нечетких отношений как с конечным, так и с бесконечным числом кортежей. В этом случае нечеткое отношение записывается в виде:  $Q = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_Q(\langle x_1, x_2, \dots, x_k \rangle) \}$  или сокращенно:  $Q = \{ w, \mu_Q(w) \}$ , понимая под  $w$  общее обозначение кортежа длины  $k$ . Если в качестве универсумов используются числовые множества, то в этом случае удобно представить функцию принадлежности  $\mu_Q$  аналитически в форме некоторой функции  $f(x_1, x_2, \dots, x_k)$  от  $k$  переменных, которая конкретизирует отображение  $f(x): X_1 \times X_2 \times \dots \times X_k \rightarrow [0, 1]$ .

В дополнение к этим способам бинарные нечеткие отношения также могут быть заданы следующим образом.

*Графически* в форме некоторой поверхности или совокупности отдельных точек в трехмерном пространстве. При этом две координаты (независимые переменные) будут соответствовать значениям универсумов  $X_1$  и  $X_2$ , а третья координата – интервалу  $[0, 1]$ . Например, график математической функции:  $z=x^2+y^2$  для  $x, y \in [-0.5, 0.5]$  может служить примером графического способа формального задания некоторого нечеткого отношения. Здесь функция  $z$  является представлением функции принадлежности соответствующего нечеткого отношения. Этот способ зачастую используется в дополнение к аналитическому для визуализации бинарных нечетких отношений, с бесконечным числом кортежей,

В форму *матрицы* нечеткого отношения. Этот способ основан на представлении нечеткого бинарного отношения с конечным числом кортежей в форме матриц  $M_Q$ , строки которой соответствуют первым элементам кортежей, а столбцы - вторым элементам кортежей рассматриваемого нечеткого отношения. При этом элементами матрицы являются, соответствующие значения функции принадлежности  $\mu_Q$  данного отношения. Если бинарное нечеткое отношение задается на одном универсуме, то матрица такого отношения  $M_Q$  является квадратной. Определенную таким образом матрицу называют матрицей бинарного нечеткого отношения и обозначают  $M_Q$ . В этом контексте табличный способ может рассматриваться как разновидность матричного, поскольку конечная матрица всегда может быть представлена в форме таблицы.

В форме так называемого *нечеткого графа*, который формально может быть задан в виде двух обычных конечных множеств и некоторой функции принадлежности. А именно, нечеткий граф, а точнее, ориентированный нечеткий граф, есть  $G=(V,E,\mu_G)$ , где  $V=\{V_1, V_2, \dots, V_n\}$ - множество вершин

нечеткого графа,  $E = \{e_1, e_2, \dots, e_m\}$  - множество дуг нечеткого графа,  $\mu_G$  - функция принадлежности дуг данному нечеткому графу, т. е.  $\mu_G: E \rightarrow [0, 1]$ . При этом вершины нечеткого графа, как и в случае обычных графов, изображаются точками, дуги – отрезками прямых линий со стрелкой на одном из концов. Рядом с вершинами записываются условные обозначения соответствующих вершин, а рядом с каждой дугой - значение функции принадлежности для соответствующей дуги. Натуральное число  $n$  определяет общее количество вершин конкретного нечеткого графа, а натуральное число  $m$  - общее количество дуг нечеткого графа. При этом дуги с нулевой функцией принадлежности в нечетком графе обычно не изображаются.

Как нетрудно заметить, каждому ориентированному нечеткому графу  $G$  соответствует некоторое бинарное нечеткое отношение  $Q_G$ , состоящее из всех пар вида  $\langle v_i, v_j \rangle$ , где  $v_i, v_j \in V$ . При этом для каждой пары  $\langle v_i, v_j \rangle$  определено некоторое действительное число из интервала  $[0, 1]$ , которое равно значению функции принадлежности  $\mu_G(e_k)$  для дуги  $e_k \in E$ , которая соответствует этой паре вершин. И обратно, если вершины  $v_i$  и  $v_j$  соединяются в нечетком графе  $G$  некоторой дугой  $e_k \in E$  со значением функции принадлежности  $\mu_G(e_k)$  и направленной из вершины  $v_i$  в вершину  $v_j$ , то тем самым задается некоторое бинарное нечеткое отношение  $Q_G$ . При задании нечеткого отношения  $Q$  с помощью ориентированного нечеткого графа  $G$  каждому элементу универсума  $x \in X$  будет соответствовать отдельная вершина  $v_i \in G$  этого нечеткого графа, а каждому кортежу нечеткого отношения  $\langle x_i, x_j \rangle \in Q$  будет соответствовать дуга графа  $e_k = \langle v_i, v_j \rangle$  с началом в вершине  $v_i$ , концом в вершине  $v_j$ ; и значением функции принадлежности  $\mu_Q(\langle x_i, x_j \rangle)$ .

В зависимости от количества кортежей нечеткое отношение может быть конечным или бесконечным. Нечеткое отношение называется конечным, если его носитель является конечным отношением. При этом вполне уместно говорить, что такое нечёткое отношение имеет конечную мощность, которая численно равна количеству кортежей его носителя, рассматриваемого как обычное множество. В этом случае для обозначе-

ния мощности произвольного нечеткого отношения  $Q$  можно использовать общепринятое обозначение  $\text{card}(Q)$ . Аналогично счетным нечетким отношением будем называть нечеткое отношение со счетным носителем, т. е. носитель которого имеет счетную мощность  $\xi_0$  в обычном смысле. Несчетным нечетким отношением называется нечеткое отношение с несчетным носителем, т.е. носитель которого имеет несчетную мощность или мощность континуума  $c$  (или  $\xi_0$ ) в обычном смысле.

Для иллюстрации описанных способов задания отношений рассмотрим два примера конкретных нечетких отношений.

Пример 1. В качестве первого примера рассмотрим конечное бинарное нечеткое отношение  $Q_1$  заданное на одном универсуме  $X$ , в качестве которого возьмем подмножество первых 10 натуральных чисел:  $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ .

Пусть отношение  $Q_1$  описывает свойство: “натуральное число  $x_i$  приближенно равно натуральному числу  $x_j$ ”

Конкретное бинарное нечеткое отношение  $Q_1$  может быть задано в форме списка следующим образом:

$Q_1 = \{(\langle 1, 1 \rangle, 1.0), (\langle 1, 2 \rangle, 0.8), (\langle 1, 3 \rangle, 0.5), (\langle 1, 4 \rangle, 0.2), (\langle 2, 1 \rangle, 0.8), (\langle 2, 2 \rangle, 1), (\langle 2, 3 \rangle, 0.8), (\langle 2, 4 \rangle, 0.5), (\langle 2, 5 \rangle, 0.2), (\langle 3, 1 \rangle, 0.5), (\langle 3, 2 \rangle, 0.8), (\langle 3, 3 \rangle, 1), (\langle 3, 4 \rangle, 0.8), (\langle 3, 5 \rangle, 0.5); (\langle 3, 6 \rangle, 0.2), (\langle 4, 1 \rangle, 0.2), (\langle 4, 2 \rangle, 0.5), (\langle 4, 3 \rangle, 0.8), (\langle 4, 4 \rangle, 1), (\langle 4, 5 \rangle, 0.8), (\langle 4, 6 \rangle, 0.5), (\langle 4, 7 \rangle, 0.2), (\langle 5, 2 \rangle, 0.2), (\langle 5, 3 \rangle, 0.5), (\langle 5, 4 \rangle, 0.8), (\langle 5, 5 \rangle, 1), (\langle 5, 6 \rangle, 0.8), (\langle 5, 7 \rangle, 0.5), (\langle 5, 8 \rangle, 0.2), (\langle 6, 3 \rangle, 0.2), (\langle 6, 4 \rangle, 0.5), (\langle 6, 5 \rangle, 0.8), (\langle 6, 6 \rangle, 1), (\langle 6, 7 \rangle, 0.8), (\langle 6, 8 \rangle, 0.5), (\langle 6, 9 \rangle, 0.2), (\langle 7, 4 \rangle, 0.2), (\langle 7, 5 \rangle, 0.5), (\langle 7, 6 \rangle, 0.8), (\langle 7, 7 \rangle, 1), (\langle 7, 8 \rangle, 0.8), (\langle 7, 9 \rangle, 0.5), (\langle 7, 10 \rangle, 0.2), (\langle 8, 5 \rangle, 0.2), (\langle 8, 6 \rangle, 0.5), (\langle 8, 7 \rangle, 0.8), (\langle 8, 8 \rangle, 1); (\langle 8, 9 \rangle, 0.8), (\langle 8, 10 \rangle, 0.5), (\langle 9, 6 \rangle, 0.2), (\langle 9, 7 \rangle, 0.5), (\langle 9, 8 \rangle, 0.8), (\langle 9, 9 \rangle, 1), (\langle 9, 10 \rangle, 0.8), (\langle 10, 7 \rangle, 0.2), (\langle 10, 8 \rangle, 0.5), (\langle 10, 9 \rangle, 0.8), (\langle 10, 10 \rangle, 1)\}$ .

В этом списке отсутствуют кортежи с нулевым значением функции принадлежности.

Это же бинарное нечеткое отношение может быть задано матрицей  $M_{Q_1}$ :

$$M_{\varrho_1} = \begin{bmatrix} 1 & 0.8 & 0.5 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 1 & 0.8 & 0.5 & 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.8 & 1 & 0.8 & 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ 0.2 & 0.5 & 0.8 & 1 & 0.8 & 0.5 & 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0.5 & 0.8 & 1 & 0.8 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0.5 & 0.8 & 1 & 0.8 & 0.5 & 0.2 & 0 \\ 0 & 0 & 0 & 0.2 & 0.5 & 0.8 & 1 & 0.8 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0 & 0.2 & 0.5 & 0.8 & 1 & 0.8 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.5 & 0.8 & 1 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.5 & 0.8 & 1 \end{bmatrix}$$

Данное нечеткое отношение можно также представить графически в форме совокупности точек в трёхмерном представлении и в форме нечёткого графа. Однако эти представления не совсем удобны для визуализации рассмотренного нечёткого отношения и поэтому здесь не приводятся.

Пример 2. Предположим, необходимо построить нечеткое отношение, которое содержательно описывает упрощенную ситуацию поиска неисправности в автомобиле. С этой целью в качестве первого универсума рассмотрим множество предпосылок или причин неисправности  $X = \{x_1, x_2, x_3, x_4\}$ , в котором  $x_1$  - "неисправность аккумулятора",  $x_2$  - "неисправность карбюратора",  $x_3$  - "низкое качество бензина",  $x_4$  - "неисправность системы зажигания". В качестве второго универсума рассмотрим множество заключений или проявлений неисправности  $Y = \{y_1, y_2, y_3\}$ , где  $y_1$  - "двигатель не запускается",  $y_2$  - "двигатель работает неустойчиво",  $y_3$  - "двигатель не развивает полной мощности". При этом между каждым элементом множества предпосылок и каждым элементом множества следствий существует некоторая причинная взаимосвязь.

Особенность построения нечеткой модели для описываемой ситуации заключается в том, что рассматриваемая причинная взаимосвязь не является однозначной. Более того, исходя из субъективного опыта конкретного механика, марки автомобиля, условий его эксплуатации и учета других факторов эта причинная взаимосвязь наиболее адекватно может быть представлена в виде бинарного нечеткого отношения  $P = \{\langle x_i, y_j \rangle, \mu_P(\langle x_i, y_j \rangle)\}$ , заданного на базисных множествах  $X$  и  $Y$ . В этом случае функция принадлежности  $\mu_P(\langle x_i, y_j \rangle)$  этого бинарного нечеткого отношения количественно описывает степень уверенности в том, что та или иная причина неисправности может привести к тому или иному следствию.

Применительно к нашему примеру конкретное нечеткое отношение  $P$  может быть записано в форме списка следующим образом:

$$P = \{(\langle x_1, y_1 \rangle, 1), (\langle x_1, y_2 \rangle, 0.1), (\langle x_1, y_3 \rangle, 0.2), (\langle x_2, y_1 \rangle, 0.8), (\langle x_2, y_2 \rangle, 0.9), (\langle x_2, y_3 \rangle, 1), (\langle x_3, y_1 \rangle, 0.7), (\langle x_3, y_2 \rangle, 0.8), (\langle x_3, y_3 \rangle, 0.5), (\langle x_4, y_1 \rangle, 1), (\langle x_4, y_2 \rangle, 0.5), (\langle x_4, y_3 \rangle, 0.2)\}.$$

Поскольку нечеткое отношение  $P$  бинарное и конечное, оно может быть представлено в форме табл. 9.1, представленной ниже.

Таблица 9.1. Нечеткое отношение диагностики неисправности в автомобиле

	$y_1$	$y_2$	$y_3$
$x_1$	1	0,1	0,2
$x_2$	0,8	0,9	1
$x_3$	0,7	0,8	0,5
$x_4$	1	0,5	0,2

Эта таблица может быть легко преобразована в матрицу  $M_p$  нечеткого отношения, которая в данном конкретном случае имеет следующий вид:

$$M_p = \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.8 & 0.9 & 1 \\ 0.7 & 0.8 & 0.5 \\ 1 & 0.5 & 0.2 \end{bmatrix}$$

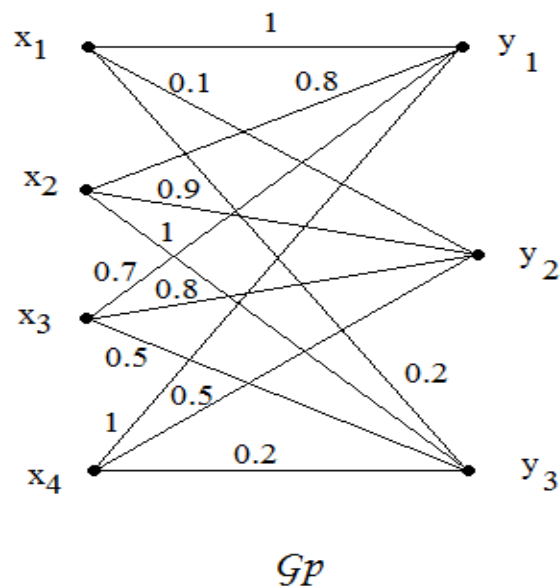


Рис. 9.9. Нечеткий граф отношения  $P$  (стрелки дуг, направленных от вершин  $x_i$ ; к вершинам  $y_j$  для удобства не указаны)

Для того чтобы представить это нечеткое отношение в форме нечеткого графа, изобразим на плоскости его вершины, в качестве которых выступают элементы множеств  $X$  и  $Y$ . Соединим эти вершины дугами, направленными от вершин, соответ-

вующих элементам множества  $X$ , к вершинам, соответствующим элементам множества  $Y$ . Рядом с каждой из дуг запишем значение её функции принадлежности. Тем самым получим нечеткий граф  $G_P$  рассматриваемого отношения  $P$  (рис.9.9).

Что касается аналитического способа представления данного нечеткого отношения, то поскольку отсутствует математическое выражение для записи соответствующей функции принадлежности, использовать этот способ в данном случае не представляется возможным.

#### 9.4. Понятие нечеткой и лингвистической переменной

Нечеткая переменная определяется как кортеж:  $\langle \alpha, X, A \rangle$ , где  $\alpha$  – наименование или название нечеткой переменной;  $X$  – область ее определения (универсум);  $A = \{x, \mu_A(x)\}$  – нечеткое множество на  $X$ , описывающее возможные значения, которые может принимать нечеткая переменная  $\alpha$ . Таким образом, говоря о нечеткой переменной  $\alpha$ , мы всегда будем иметь в виду некоторое нечеткое множество  $A$ , которое определяет ее возможные значения.

В качестве примера нечеткой переменной можно привести рассмотренное выше нечеткое множество  $B$ , которое характеризует «горячий кофе». В этом случае соответствующая нечеткая переменная может быть представлена следующим образом:  $\langle \text{Горячий кофе}, \{x \mid 0^\circ\text{C} < x < 100^\circ\text{C}\}, B \rangle$ , где  $B = \{x, \mu_B(x)\}$  – нечеткое множество с функцией принадлежности  $\mu_B(x)$ , которая может быть задана, в частности, графически (рис. 9.1, а или рис. 9.1, б).

Обобщением нечеткой переменной является так называемая лингвистическая переменная.

**Л и н г в и с т и ч е с к а я п е р е м е н н а я.** Лингвистическая переменная также определяется как кортеж:  $\langle \beta, T, X, G, M \rangle$ , где:

$\beta$  – наименование или название лингвистической переменной;

$T$  – базовое терм-множество лингвистической переменной или множество ее значений (термов), каждое из которых представляет собой наименование отдельной нечеткой переменной  $\alpha$ ;

$X$  – область определения (универсум) нечетких переменных, которые входят в определение лингвистической переменной  $\beta$ ;

$G$  – некоторая синтаксическая процедура, которая описывает процесс образования или генерирования из множества  $T$  новых, осмысленных в рассматриваемом контексте значений для данной лингвистической переменной;

$M$  – семантическая процедура, которая позволяет поставить в соответствие каждому новому значению лингвистической переменной, получаемому с помощью процедуры  $G$ , некоторое осмысленное содержание посредством формирования соответствующего нечеткого множества.

В качестве примера рассмотрим ситуацию со скоростью движения автомобильного транспорта в пределах городской черты. Хотя правила дорожного движения регламентируют величину этой скорости, однако многие автолюбители предпочитают давать собственную субъективную оценку своей скорости движения. При этом используются такие определения, как «малая скорость», «средняя скорость» и «высокая скорость» движения. Очевидно, что подобная практическая оценка скорости может относиться к диапазону скоростей в пределах интервала от 0 км/ч до некоторой величины, определяемой личными предпочтениями того или иного водителя. Пусть в нашем примере из соображений удобства это будет величина 100 км/ч. Формализация субъективной оценки скорости движения может быть выполнена с помощью следующей лингвистической переменной  $\langle \beta_1, T, X, G, M \rangle$ , где  $\beta_1$  – скорость движения автомобиля;  $T = \{ \text{«малая скорость»}, \text{«средняя скорость»}, \text{«высокая скорость»} \}$ ;  $X = [0, 100]$ ;

$G$  – процедура образования новых термов с помощью логических связок «И», «ИЛИ» и модификаторов типа «очень», «НЕ», «слегка» и др. Например: «малая или средняя скорость», «очень высокая скорость» и др.;

$M$  – процедура задания на  $X = [0, 100]$  нечетких переменных  $\alpha_1 = \text{«малая скорость»}$ ,  $\alpha_2 = \text{«средняя скорость»}$ ,  $\alpha_3 = \text{«высокая скорость»}$ , а также соответствующих нечетких

множеств для термов из  $G(T)$  в соответствии с правилами трансляции нечетких связей и модификаторов «И», «ИЛИ», «НЕ», «очень», «слегка».

Для рассматриваемого примера нечеткие множества  $A_1, A_2, A_3$ , соответствующие нечетким переменным:  $\alpha_1$  = «малая скорость»,  $\alpha_2$  = «средняя скорость»,  $\alpha_3$  = «высокая скорость», удобно задать графически с помощью кусочно-линейных функций принадлежности. Один из возможных конкретных вариантов этих нечетких множеств изображен на рис. 9.10.

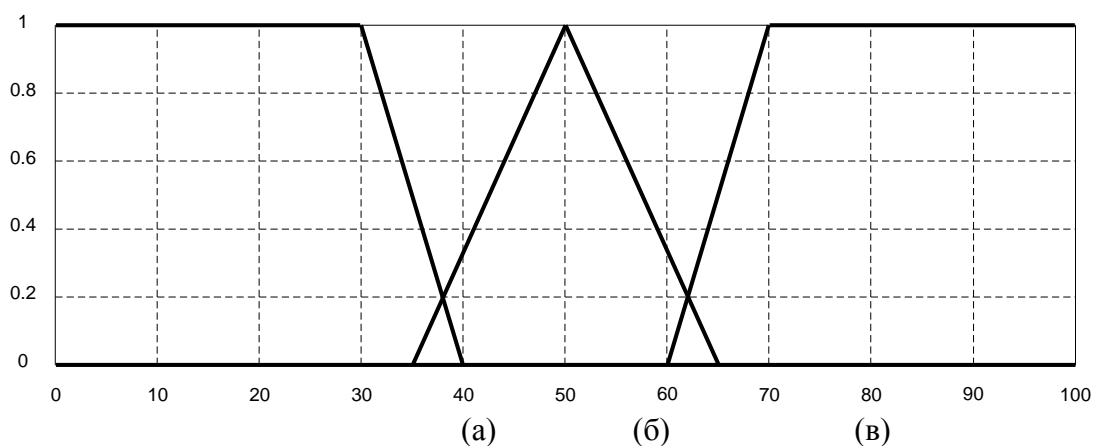


Рис. 9.10. Графики функций принадлежности нечетких множеств  $A_1, A_2, A_3$ , соответствующих нечетким переменным  $\alpha_1$  = «малая скорость» (а),  $\alpha_2$  = «средняя скорость» (б),  $\alpha_3$  = «высокая скорость» (в) для лингвистической переменной (скорость движения автомобиля)

Наряду с рассмотренными выше базовыми значениями лингвистической переменной «скорость движения автомобиля» ( $T = \{ \text{«малая скорость»}, \text{«средняя скорость»}, \text{«высокая скорость»} \}$ ) возможны и другие значения этой же лингвистической переменной, зависящие от конкретной величины скорости движения. Например, могут быть определены такие дополнительные значения лингвистической переменной «скорость движения автомобиля» как «около 30 км/ч», «около 50 км/ч», «около 70 км/ч». Как будет видно из дальнейшего изложения, эти значения лингвистической переменной удобно моделировать с помощью нечетких чисел.

*Нечеткие величины, числа и интервалы.* Процесс нечеткого моделирования основывается на количественном представлении входных и выходных переменных системы в форме нечетких множеств. Такое представление связано с рассмотрением специальных нечетких множеств, которые задаются на множестве действительных чисел и обладают некоторыми

дополнительными свойствами. Наиболее общим понятием в этом контексте является понятие нечеткой величины.

Нечеткой величиной называется произвольное нечеткое множество  $V = \{x, \mu_V(x)\}$ , заданное на множестве действительных чисел  $R$ , т. е. для которого универсумом  $X$  служит все множество  $R$ . Другими словами, функция принадлежности нечеткой величины есть отображение:  $\mu_V(x): R \rightarrow [0,1]$ . Если в качестве универсума взять подмножество неотрицательных действительных чисел  $R^+$ , то получим определение неотрицательной нечеткой величины  $V^+$ .

Наибольший интерес для нечеткого моделирования представляет конкретизация нечеткой величины в форме нечетких чисел и интервалов.

В общем случае нечетким интервалом называется нечеткая величина с выпуклой функцией принадлежности.

В общем случае нечетким числом называется такая нечеткая величина, функция принадлежности которой является выпуклой и унимодальной.

Нечеткое число в общем случае является частным случаем нечеткого интервала, что полностью согласуется с обычными числами и интервалами на множестве действительных чисел.

Поскольку нечеткие числа и интервалы представляют собой нечеткие множества, то для них оказываются справедливыми все свойства и операции, определенные ранее для нечетких множеств. Это в полной мере относится к определению нормального нечеткого числа и нормального нечеткого интервала, носителя и ядра, а также свойств выпуклости и унимодальности нечетких чисел и нечетких интервалов, которые были использованы при их определении.

Дополнительно нечеткие числа могут характеризоваться следующими свойствами.

Нечеткое число называется нечетким нулем, если его модальное значение (мода) равно 0.

Нечеткое число называется положительным (или отрицательным) нечетким числом, если оно имеет строго положительный (соответственно, строго отрицательный) носитель.

К основным вопросам, возникающим при разработке и реализации моделей при нечеткой исходной информации, можно отнести следующие:

- построение функций принадлежности нечетких множеств;
- выполнение операций над нечеткими числами;
- сравнение и упорядочение нечетких множеств и чисел;
- разработка нечетких моделей.

### **9.3. Построение функций принадлежности.**

#### **Прямые и косвенные методы**

Во всех случаях функция принадлежности строится методами экспертного опроса. В прямых методах эксперты непосредственно указывают степень принадлежности для каждого элемента. Методы простые, но имеют высокую степень субъективизма. Для уменьшения степени субъективизма используют различные приемы, например: усреднение по числу экспертов; параметрическое представление функции принадлежности.

Рассмотрим более подробно второй прием.

Формальное определение нечеткого множества не накладывает никаких ограничений на выбор конкретной функции принадлежности для его представления. Однако на практике удобно использовать те из них, которые допускают аналитическое представление в виде некоторой простой математической функции. Это упрощает не только соответствующие численные расчеты, но и сокращает вычислительные ресурсы, необходимые для хранения отдельных значений этих функций принадлежности. Необходимость типизации отдельных функций принадлежности также обусловлена наличием реализаций соответствующих функций в имеющихся инструментальных средствах.

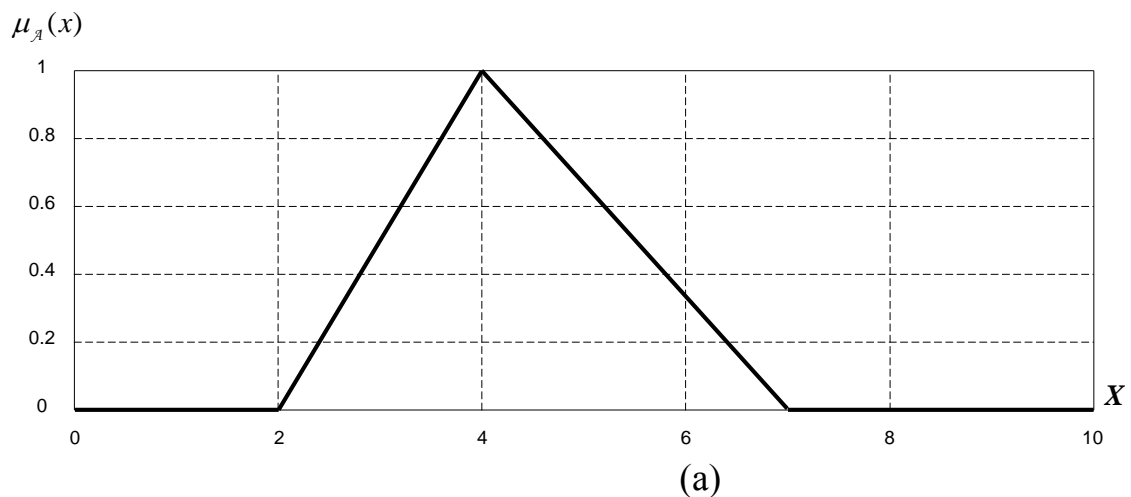
*Кусочно-линейные функции принадлежности.* В качестве первого типа функций принадлежности рассмотрим функции, которые, как следует из

их названия, состоят из отрезков прямых линий, образуя непрерывную или кусочно-непрерывную функцию. Наиболее характерным примером таких функций являются «треугольная» (рис. 9.11, а) и «трапециевидная» (рис. 9.11, б) функции принадлежности. В нашем случае каждая из этих функций задана на универсуме  $X=[0, 10]$ , в качестве которого выбран замкнутый интервал действительных чисел. В общем случае выбор универсума может быть произвольным и не ограничен никакими правилами.

Первая из этих функций принадлежности в общем случае может быть задана аналитически следующим выражением:

$$f_{\Delta}(x; a, b, c) = \left\{ \begin{array}{ll} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{array} \right\}, \quad (9.15)$$

где  $a, b, c$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a \leq b \leq c$ .



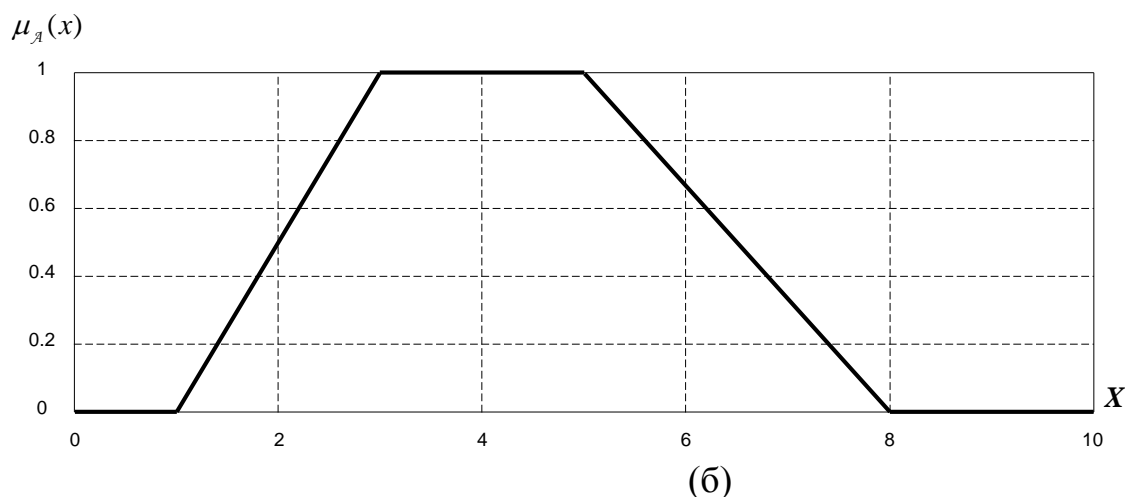


Рис. 9.11. Графики функций принадлежности треугольной (а) и трапециевидной (б) формы

Применительно к конкретной функции, изображенной на рис. 9.11,а, значения параметров равны:  $a=2$ ,  $b=4$ ,  $c=7$ . Как нетрудно заметить, параметры  $a$  и  $c$  характеризуют основание треугольника, а параметр  $b$  – его вершину. Как можно заметить, эта функция принадлежности порождает нормальное выпуклое унимодальное нечеткое множество с носителем – интервалом  $(a, c)$ , границами  $(a, c) \setminus \{b\}$ , ядром  $\{b\}$  и модой  $b$ .

*Трапециевидная* функция принадлежности в общем случае может быть задана аналитически следующим выражением:

$$f_T(x; a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}, \quad (9.16)$$

где  $a, b, c, d$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a \leq b \leq c \leq d$ . Эти функции используются для задания таких свойств множеств, которые характеризуют неопределенность типа: «приблизительно равно»,

«среднее значение», «расположен в интервале», «подобен объекту», «похож на предмет» и др.

*Z-образные и S-образные функции принадлежности.* Эти функции принадлежности также получили свое название по виду кривых, которые представляют их графики. Первая из функций этой группы называется *Z-образной кривой* или *сплайн-функцией* и в общем случае может быть задана аналитически следующим выражением:

$$f_{Z_1}(x; a, b) = \begin{cases} 1, & x < a \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-a}{b-a} \pi\right), & a \leq x \leq b \\ 0, & x > b \end{cases}, \quad (9.17)$$

где  $a, b$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b$ . График этой функции для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис. 9.12, а, при этом значения параметров соответственно равны  $a=3, b=6$ . Сплайн-функция может быть также задана другим выражением:

$$f_{Z_2}(x; a, b) = \begin{cases} 1, & x \leq a \\ 1 - 2\left(\frac{x-a}{b-a}\right)^2, & a < x \leq \frac{a+b}{2} \\ 2\left(\frac{b-x}{b-a}\right)^2, & \frac{a+b}{2} < x < b \\ 0, & b \leq x \end{cases}, \quad (9.18)$$

где  $a, b$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b$ . График этой функции для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис. 9.12, б, при этом значения параметров соответственно равны:  $a=3, b=6$ .

Данные функции принадлежности порождают нормальные выпуклые нечеткие множества с ядром  $(-\infty, a]$  и носителем  $(-\infty, b)$ .

Эти функции используются для представления таких свойств нечетких множеств, которые характеризуются неопределенностью типа: «малое количество», «небольшое значение», «незначительная величина», «низкая себестоимость продукции», «низкий уровень цен или доходов», «низкая процентная ставка» и многих других. Общим для всех таких ситуаций является слабая степень проявления того или иного качественного или количественного признака. Особенность нечеткого моделирования при этом заключается в представлении соответствующих нечетких множеств с помощью невозрастающих (монотонно убывающих) функций принадлежности.

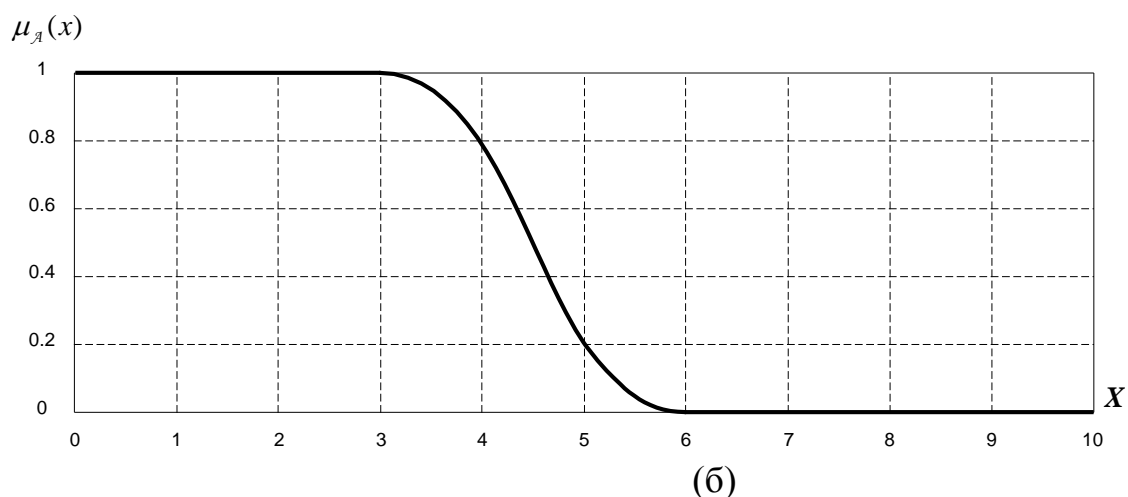
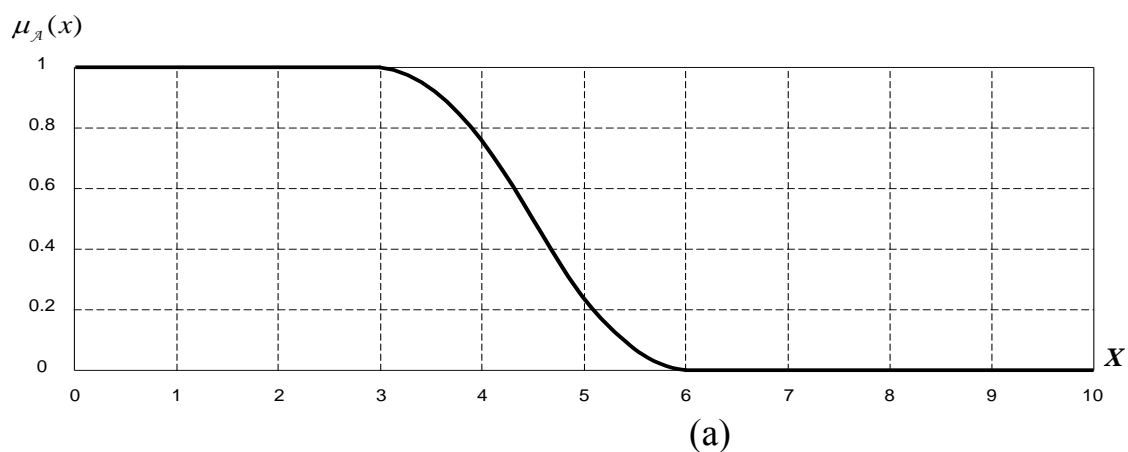


Рис. 9.12. Графики Z-образных функций принадлежности  $f_{Z1}$  и  $f_{Z2}$  для значений параметров  $a=3$ ,  $b=6$

Вторая из функций рассматриваемой группы называется S-образной кривой или сплайн-функцией и в общем случае может быть задана аналитически следующим выражением:

$$f_{S_1}(x; a, b) = \left\{ \begin{array}{ll} 0, & x < a \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-b}{b-a} \pi\right), & a \leq x \leq b \\ 1, & x > b \end{array} \right\}, \quad (9.19)$$

где  $a, b$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b$ . График этой функции для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис. 9.13, а, при этом значения параметров соответственно равны:  $a=3, b=6$ . Сплайн-функция может быть также задана другим выражением:

$$f_{S_2}(x; a, b) = \left\{ \begin{array}{ll} 0, & x \leq a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a < x \leq \frac{a+b}{2} \\ 1-2\left(\frac{b-x}{b-a}\right)^2, & \frac{a+b}{2} < x < b \\ 1, & b \leq x \end{array} \right\}, \quad (9.20)$$

где  $a, b$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b$ . График этой функции для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис. 9.13, б, при этом значения параметров соответственно равны:  $a=3, b=6$ .

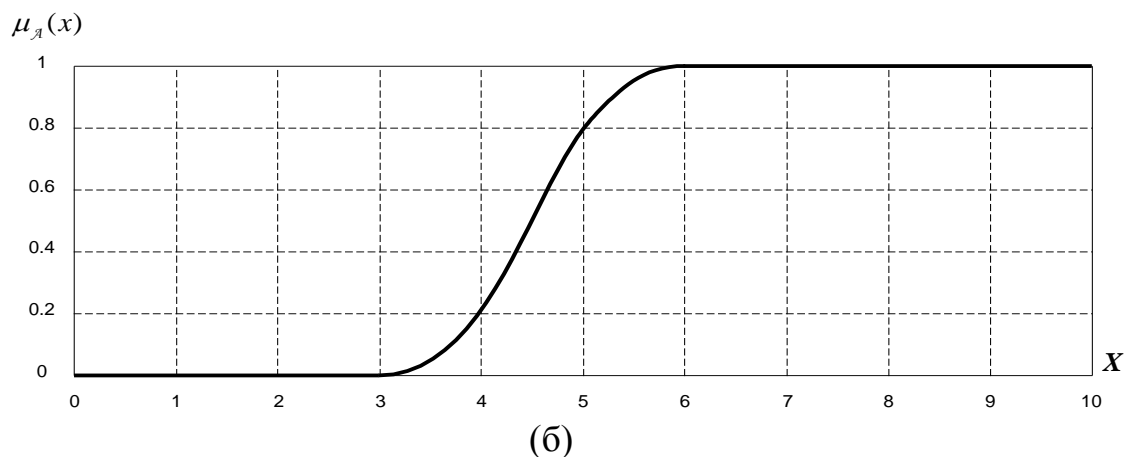
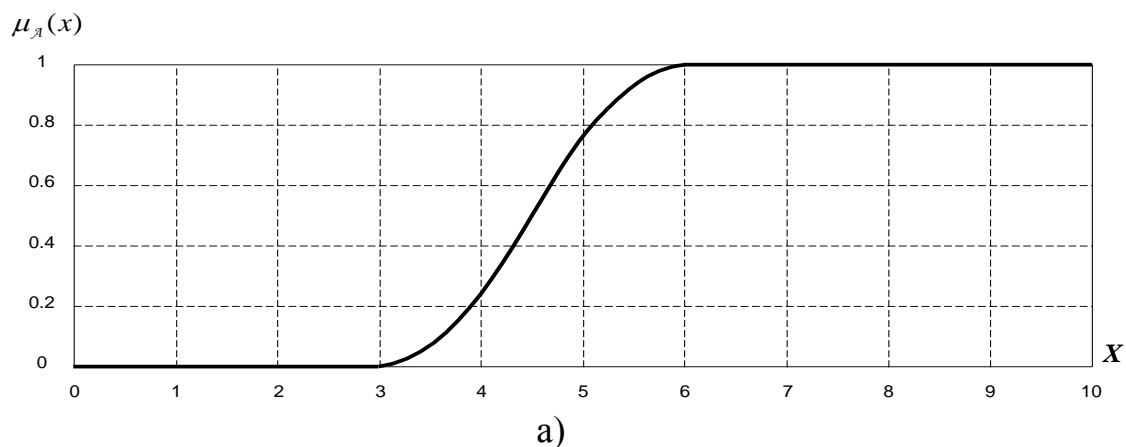


Рис. 9.13. Графики S-образных функций принадлежности  $f_{S_1}$  и  $f_{S_2}$  для значений параметров  $a=3$ ,  $b=6$

Данные функции принадлежности порождают нормальные выпуклые нечеткие множества с ядром  $[b, +\infty)$  и носителем  $(a, +\infty)$ .

К типу S-образных и одновременно Z-образных функций принадлежности может быть отнесена так называемая *сигмоидальная* функция (сигмоид), которая в общем случае задается аналитически следующим выражением:

$$f_{S_3}(x; a, b) = \frac{1}{1 + e^{-a(x-b)}}, \quad (9.21)$$

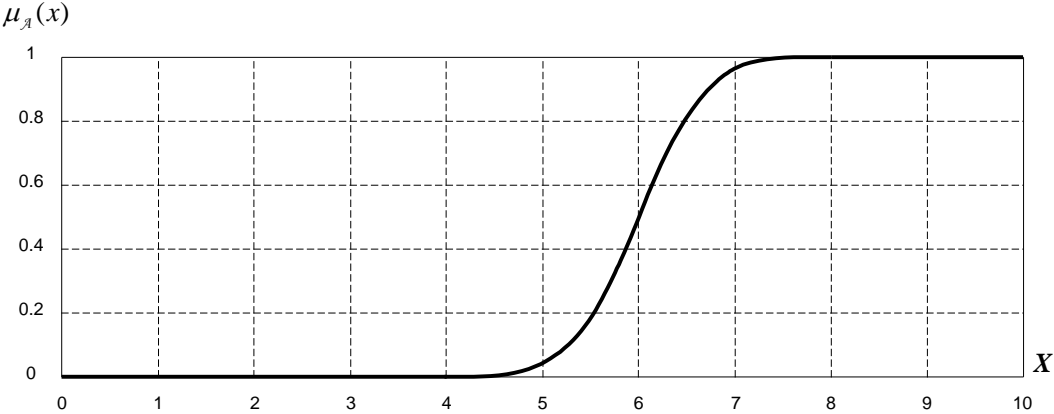
здесь  $a$ ,  $b$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b$ ,  $e$  – основание натуральных логарифмов, которое инициирует задание соответствующей экспоненциальной функции. При этом в случае  $a > 0$  может быть

получена S-образная функция принадлежности, а в случае  $a < 0$  – Z-образная функция принадлежности.

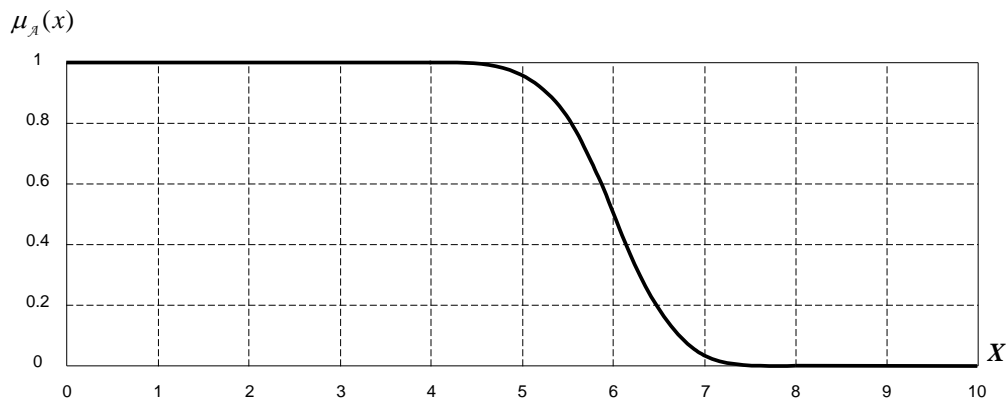
Графики этой функции для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображены на рис. 9.14. При этом S-образной функции принадлежности соответствуют значения параметров  $a=3, b=6$  (рис. 9.14, а), а Z-образной функции принадлежности соответствуют значения параметров  $a=-3, b=6$  (рис. 9.14, б).

Данные функции принадлежности порождают субнормальные выпуклые нечеткие множества с носителем и границей  $R$  и точкой перехода  $b$ .

Рассмотренные S-образные функции используются для представления таких нечетких множеств, которые характеризуются неопределенностью типа: «большое количество», «большое значение», «значительная величина», «высокий уровень доходов и цен», «высокая норма прибыли», «высокое качество услуг», «высокий сервис обслуживания» и многих других. Общим для всех таких ситуаций является высокая степень проявления того или иного качественного или количественного признака.



(a)



(б)

Рис. 9.14. Графики сигмоидальной функции принадлежности  $f_{S3}$  для значений параметров  $a=3, b=6$  (а) и  $a=-3, b=6$  (б)

Особенность нечеткого моделирования при этом заключается в представлении соответствующих нечетких множеств с помощью неубывающих (монотонно возрастающих) функций принадлежности.

*П-образные* функции принадлежности. К данному типу функций можно отнести целый класс кривых, которые по своей форме напоминают колокол, сглаженную трапецию или букву «П».

Первая из подобных функций так и называется – П-образная функция, и в общем случае задается аналитически следующим выражением:

$$f_{\Pi}(x; a, b, c, d) = f_S(x; a, b) \cdot f_Z(x; c, d), \quad (9.22)$$

где  $a, b, c, d$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a \leq b < c \leq d$ , а знак « $\cdot$ » обозначает обычное арифметическое произведение значений соответствующих функций.

При этом могут быть использованы любые из рассмотренных выше Z- и S-образных функций. В частности, если использовать функции  $f_{S1}$  и  $f_{Z1}$ , то получим П-функцию  $f_{\Pi1}$ , график которой для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис. 9.15, а. При этом

значения параметров для функции  $f_{S1}$  равны  $a=1$ ,  $b=4$ , а для функции  $f_{Z1} - c=5$ ,  $d=9$ . Если же использовать функции  $f_{S1}$  и  $f_{Z2}$ , то получим П-функцию  $f_{П2}$ , график которой для некоторого нечеткого множества  $A$  и универсума  $X=[0, 10]$  изображен на рис.9.15, б для тех же значений параметров.

Очевидно, этот тип функций принадлежности порождает нормальные выпуклые нечеткие множества с носителем  $(a, b)$  и ядром  $[b, c]$ .

Следующая функция этого типа П-образных функций определяется как произведение двух сигмоидальных функций и в общем случае может быть задана аналитически следующим выражением:

$$f_{П3}(x; a, b, c, d) = f_{S3}(x; a, b) \cdot f_{S3}(x; c, d), \quad (9.23)$$

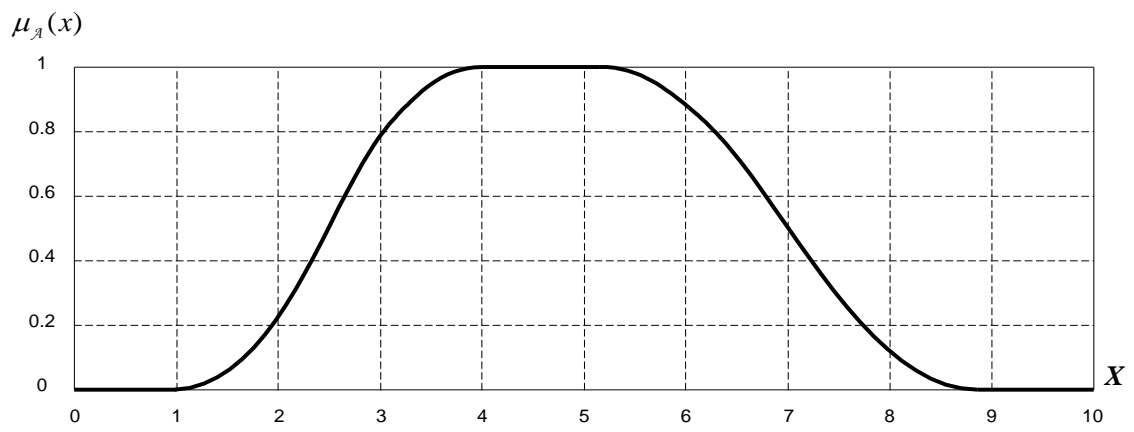
где  $a, b, c, d$  – некоторые числовые параметры, принимающие произвольные действительные значения, причем  $a > 0$ ,  $c < 0$ , и упорядоченные отношением:  $a \leq b < |c| \leq d$ .

Знак « $\cdot$ » обозначает арифметическое произведение значений соответствующих функций, а функция  $|x|$  – модуль действительного числа.

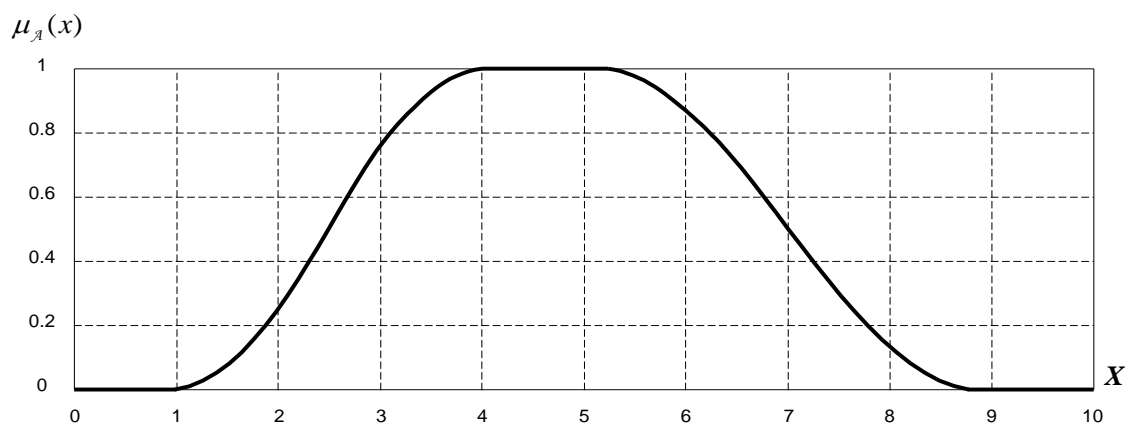
К П-образным функциям относится также так называемая колоколообразная (bell-shaped) функция, которая в общем случае задается аналитически следующим выражением:

$$f_{П4}(x; a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}}, \quad (9.24)$$

где  $a, b, c$  – некоторые числовые параметры, принимающие произвольные действительные значения и упорядоченные отношением:  $a < b < c$ , причем параметр  $b > 0$ . Здесь функция  $|x|$  обозначает модуль действительного числа.

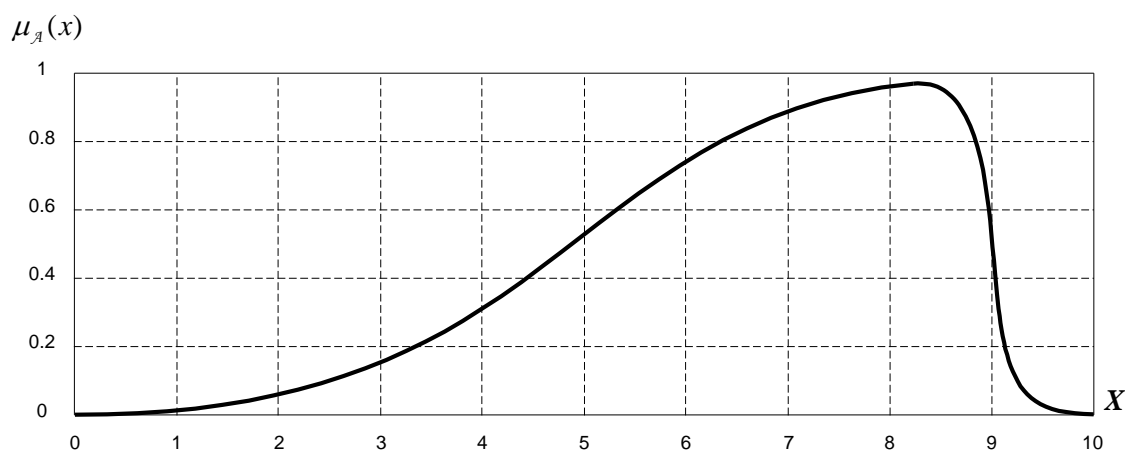


(a)

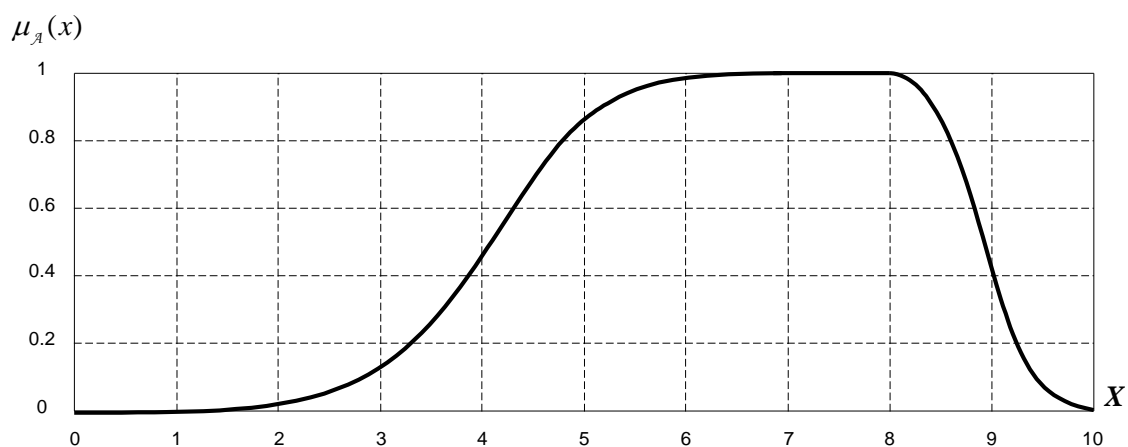


(б)

Рис. 9.15. Графики П-образных функций принадлежности  $f_{П1}$  (а) и  $f_{П2}$  (б) для значений параметров  $a=1, b=4, c=5, d=9$



(а)



(б)

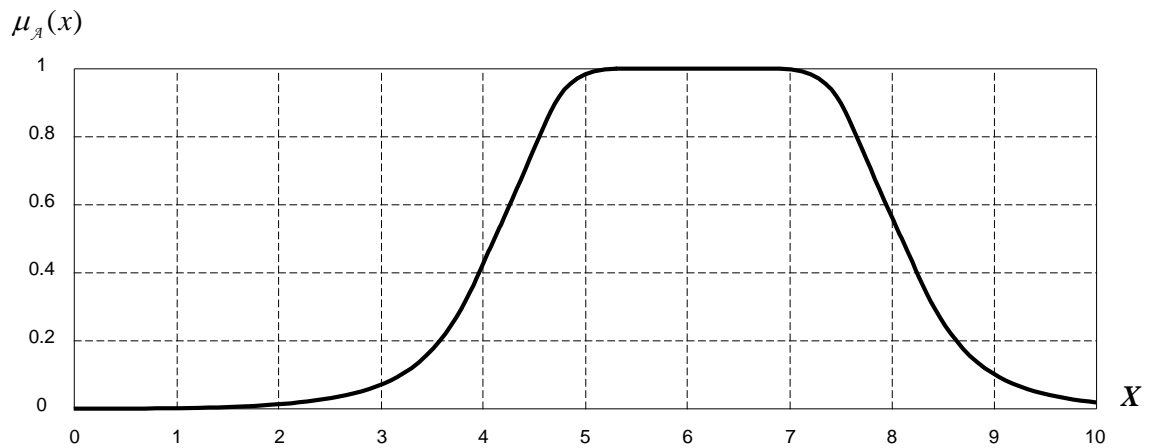
Рис. 9.16. Графики П-образных функций принадлежности  $f_{ПЗ}$  для значений параметров  $a=1, b=5, c=-7, d=9$  (а) и для значений параметров  $a=2, b=4, c=-5, d=9$  (б)

Наконец, последней из рассматриваемых функций данного типа является хорошо известная в теории вероятностей функция плотности нормального распределения в предположении, что  $\sqrt{2\pi}\sigma = 1$ , и которая в нашем случае задается аналитически следующим выражением:

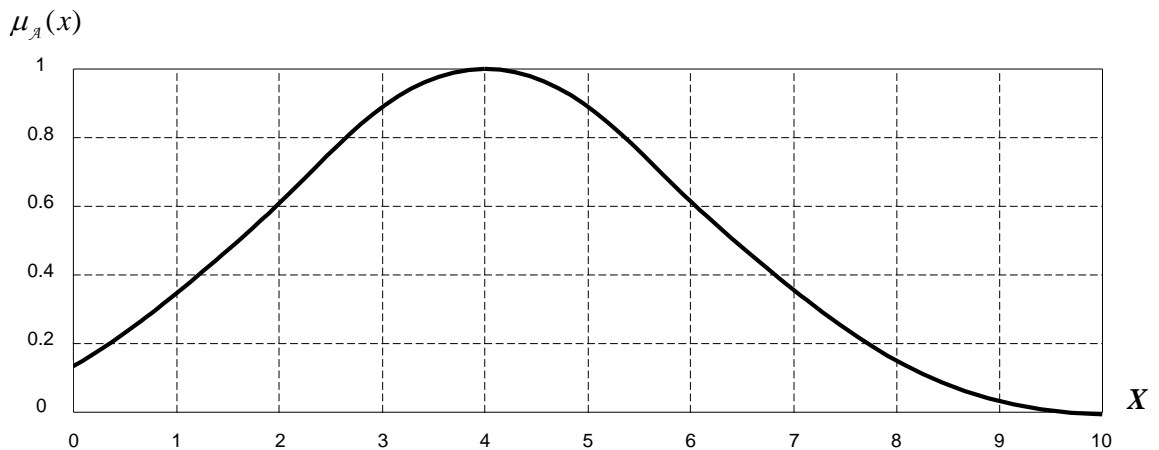
$$f_{П5}(x; \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}}. \quad (9.25)$$

Здесь  $\sigma$  и  $c$  – числовые параметры, при этом квадрат первого из них  $\sigma^2$  в теории вероятностей называется дисперсией распределения, а второй параметр  $c$  – математическим ожиданием.

Очевидно, эти последние типы функций принадлежности порождают нормальные выпуклые нечеткие множества, при этом плотность нормального распределения обеспечивает унимодальность соответствующего нечеткого множества.



(а)



(б)

Рис. 9.17. Графики П-образных функций принадлежности  $f_{П4}$  для значений параметров  $a=2$ ,  $b=3$ ,  $c=6$  (а) и  $f_{П5}$  для значений параметров  $\sigma=2$ ,  $c=4$  (б)

В косвенных методах значения функции принадлежности выбираются таким образом, чтобы удовлетворить заранее сформированным условиям (вычисляются по определенным правилам из информации экспертного опроса).

*Метод равноделения (метод психологического шкалирования).*

Пусть  $U$  – универсальное множество, а  $u_m$  и  $u_M$  элементы с  $\min$  и  $\max$  степенями принадлежностями. Для определенности примем  $\mu(u_m) = 0$ ,  $\mu(u_M) = 1$ . Алгоритм метода:

- найти среднюю по принадлежности точку интервала  $[u_m, u_M]$ :  $u_{0,5}$   
 $\mu(u_{0,5}) = 0,5$ ;
- найти среднюю по принадлежности точку интервала  $[u_{0,5}, u_M]$ :  $u_{0,75}$ ;
- найти среднюю по принадлежности точку интервала  $[u_m, u_{0,5}]$ :  $u_{0,25}$ ;
- найти среднюю по принадлежности точку интервала  $[u_{0,25}, u_{0,75}]$  для проверки согласованности, если это не  $u_{0,5}$ , то найденные точки изменяются до получения согласованности;
- шаги 1-4 необходимо повторить для получения большого числа точек;
- по набору точек  $\{u_\mu, \mu\}$  определяется искомая функция принадлежности подходящим методом интерполяции.

*Метод попарного сравнения.*

Обработка матрицы оценок, отражающих мнение эксперта об относительной принадлежности элементов множеству или степени выраженности у них свойств, формализуемого множеством. Потребуем, чтобы для всех элементов множества  $S$  выполнялось равенство

$$\sum_{i=1}^n \mu_s(x) = 1.$$

Степень принадлежности будем определять посредством парных сравнений. Оценка элемента  $x_i$  по сравнению с  $x_j$  с точки зрения свойства  $S$  -  $a_{ij} = \frac{\mu_S(x_i)}{\mu_S(x_j)}$ . Для обеспечения согласованности примем  $a_{ij} = \frac{1}{a_{ji}}$ . Итак, мы имеем матрицу оценок  $A = \|a_{ij}\|$ . Найдем  $w = (\omega_1, \dots, \omega_n)$  собственный вектор матрицы  $A$ , решая уравнение  $A \cdot \omega = \lambda \omega$ , где  $\lambda$  собственное значение матрицы  $A$ . Вычисленные значения  $\omega_i$  принимаются за  $\mu_S(x_i) = \omega_i$   $i = 1, \bar{n}$ . Т.к. всегда выполняется равенство  $A\omega = n\omega$ , то найденные значения тем точнее, чем ближе  $\lambda_{\max}$  к  $n$ . Отклонения  $\lambda_{\max}$  от  $n$  может служить мерой согласованности суждений экспертов при составлении матрицы оценок.

Нахождение собственного вектора и собственных чисел матрицы.

Иногда приходится рассматривать уравнения типа:

$$\|A\| \cdot \vec{\omega} = \lambda \cdot \vec{\omega}, \quad (9.26)$$

где  $\omega$  - неизвестный числовой вектор, высота которого равна порядку  $\|A\|$ , а  $\lambda$  неизвестное число. При любом  $\lambda$  уравнение (9.26) обладает тривиальным решением  $\vec{\omega}$ , однако нас будут интересовать только такие  $\lambda$ , при которых эта система имеет нетривиальные решения, эти значения  $\lambda$  называются собственными значениями матрицы  $A$ , а решения  $\vec{\omega}$  при таких значениях - её собственными векторами. Преобразуем (9.26), используя очевидное равенство:  $\vec{\omega} = \vec{E} \cdot \vec{\omega}$ , тогда получаем

$$(\vec{A} - \lambda \vec{E}) \vec{\omega} = 0. \quad (9.27)$$

(9.27) представляет собой систему из  $n$  однородных алгебраических уравнений с  $n$  неизвестными. Для наличия нетривиального решения необходимо и достаточно, чтобы  $\det(\vec{A} - \lambda \vec{E}) = 0$ . Это уравнение называется характеристическим уравнением матрицы  $\vec{A}$ , оно служит для разыскания собственных значений  $\lambda$ .

$$\vec{A} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \quad \begin{vmatrix} a_1 - \lambda & b_1 & c_1 \\ a_2 & b_2 - \lambda & c_2 \\ a_3 & b_3 & c_3 - \lambda \end{vmatrix} = 0$$

Раскрыв определитель, получим алгебраическое уравнение степени  $n$  (порядок матрицы  $A$ ). Матрица порядка « $n$ » имеет « $n$ » собственных значений. Обычно выбирают максимальное по модулю значение.

Найдя, какое либо собственное значение  $\lambda$ , мы можем соответственно собственные вектора найти из (9.27). Из этого уравнения вытекает, что при зафиксированной  $\lambda$  сумма решений  $\vec{y} = \vec{\omega}^1 + \vec{\omega}^2$  будет снова решением, произведение  $\vec{y} = k\vec{\omega}$  решения на число будет также решением уравнения. Значит, совокупность всех собственных векторов, отвечающих заданному собственному значению, образует линейное пространство (бесконечное число ненулевых решений, система обязательно зависима, т.е. одно уравнение является следствием остальных, одно из уравнений можно отбросить и задавать произвольные значения для одной из переменных). Для нахождения единственного собственного вектора используется замена одного из уравнений условием нормировки.

#### 9.4. Арифметические операции над нечеткими числами

Под нечетким числом будем понимать нечеткое множество с областью определения в виде интервала действительной оси  $R_1$ . Пусть  $A$  и  $B$  – два нечетких числа с носителями  $S_A = (a_1, a_2)$  и  $S_B = (b_1, b_2)$  соответственно:  $a_2 > a_1$  и  $b_2 > b_1$ .

Согласно *принципу обобщения* нечеткое число  $D = A \otimes B$  определяется функцией принадлежности:

$$\mu_D(x) = \sup \min \{ \mu_A(a), \mu_B(b) \} \quad \otimes = +, -, \cdot, /$$

$$x = a \otimes b$$

$$a \in S_A, B \in S_B.$$

Первоначально принцип обобщения был введен как некоторый эвристический прием, затем он был получен дедуктивно, а его физический смысл был объяснен в рамках вероятностной интерпретации функции принадлежности.

Носитель  $S_D$  нечеткого числа  $D$  можно найти по правилам интервальной арифметики

$$S_D = \{x : x = a \otimes b, a \in S_A, b \in S_B\}.$$

Очевидно, что в множестве  $S_A \times S_B$  для любого  $x \in S_A \otimes S_B$ , где операция  $\otimes$  над носителями нечетких чисел  $A$  и  $B$  выполняется по правилам интервальной арифметики, существует бесконечно много пар  $(a, b)$  таких, что  $x = a \otimes b$ .

Пусть  $A = \left\{ \frac{0.1}{5} + \frac{0.8}{6} + \frac{0.4}{7} \right\}$  и  $B = \left\{ \frac{0.2}{4} + \frac{0.9}{5} + \frac{0.3}{6} \right\}$ . Тогда получаем

$$A + B = \left\{ \frac{0.1}{9} + \frac{0.2}{10} + \frac{0.8}{11} + \frac{0.4}{12} + \frac{0.3}{13} \right\}$$

$$A \cdot B = \left\{ \frac{0.1}{20} + \frac{0.2}{24} + \frac{0.1}{25} + \frac{0.2}{28} + \frac{0.8}{30} + \frac{0.4}{35} + \frac{0.3}{36} + \frac{0.3}{42} \right\}.$$

В случае непрерывных нечетких чисел для вычисления результата выполнения арифметической операции разработано несколько алгоритмов (алгоритм перебора, алгоритм для кусочно-непрерывной функции принадлежности (требуется числовое решение нелинейных уравнений или ряда задач безусловной оптимизации)).

Другой способ вычисления может быть использован в случае треугольных функций принадлежности. Пусть  $X$  - нечеткое число, определим левую  $x'$ , правую  $x''$  границы нечеткого числа и его вершину:

$$\forall \delta : \mu(x') = 0 \quad \mu(x' - \delta) = 0 \quad \mu(x' + \delta) \neq 0$$

$$\forall \delta : \mu(x'') = 0 \quad \mu(x'' - \delta) \neq 0 \quad \mu(x'' + \delta) = 0$$

$$\mu(x^*) = 1.$$

Можно доказать, что нечеткое число  $Z = X \otimes Y$  также определяется функцией принадлежности треугольного типа, границы и вершина которой находятся следующим образом

$$z' = x' \otimes y', \quad z'' = x'' \otimes y'', \quad z^* = x^* \otimes y^*.$$

*Связь четких и нечетких значений лингвистической переменной.*

Система управления запасами сформировала решение: закупить небольшое количество деталей. Нечеткое понятие «небольшое количество» задано дискретным нечетким числом

$$\langle \text{"небольшое количество"}, \overline{1,50}, C^+ \rangle$$

$$C^+ = \left\{ \frac{0}{1} + \dots + \frac{0}{10} + \frac{0,1}{11} + \frac{0,15}{12} + \frac{0,2}{13} + \dots + \frac{0,9}{21} + \frac{1,0}{22} + \dots + \frac{1,0}{30} + \frac{0,9}{31} + \frac{0,8}{32} + \frac{0,4}{33} + \frac{0,2}{34} + \frac{0}{35} + \dots \right\}$$

.Но закупить можно лишь четкое количество деталей, необходимо произвести выбор четкого значения лингвистической переменной КОЛИЧЕСТВО ДЕТАЛЕЙ при наличии её нечеткого значения «НЕБОЛЬШОЕ». Аналогичных ситуаций в реальных задачах принятия решений достаточно много. Для решения данной задачи в литературе предложен ряд процедур, например:

1. каждый раз выбирать  $x^* = \operatorname{argsup}_{x \in X} \mu_S(x)$ .
2. всегда выбирать такой элемент  $x_0 \in X$ , который делит площадь под графиком функции принадлежности пополам

$$\int_{-\infty}^{x_0} \mu_E(x) dx = \frac{1}{2} \int_{-\infty}^{+\infty} \mu_E(x) dx.$$

## 9.5. Сравнение нечетких чисел

Результат расчетов в модели принятия решений в нечёткой среде (интегральная оценка альтернативы), как правило, является нечётким числом. Поэтому необходимо определить процедуру сравнения нечётких чи-

сел. Рассмотрим два нечетких числа:  $\langle A, R^1, \mu_A \rangle$  и  $\langle B, R^1, \mu_B \rangle$ , у которых  $S_A \cap S_B \neq \emptyset$ .

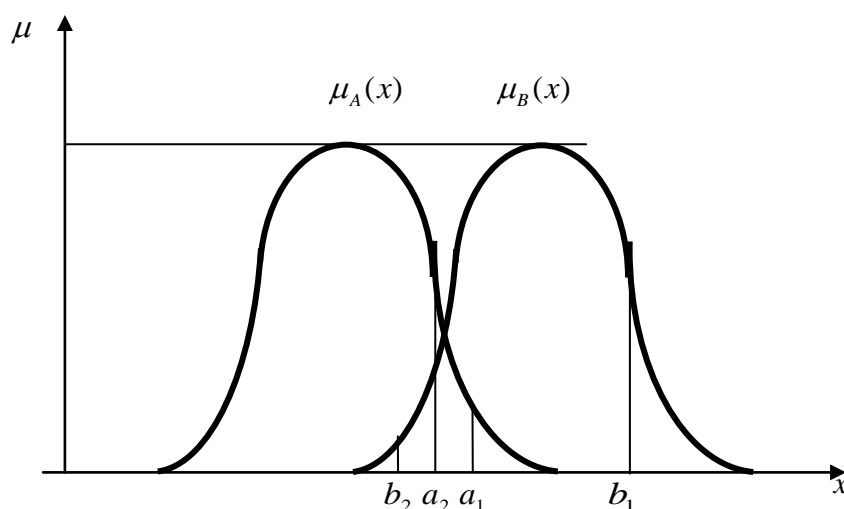


Рис. 9.18. Соотношения между четкими значениями нечетких чисел

В разных реализациях соотношения между четкими значениями нечетких чисел (а значит, и между именами нечетких чисел) может быть различным

- 1: ситуация  $b_1 > a_1$  ( $A < B$ );
- 2: ситуация  $a_2 > b_2$  ( $A > B$ ).

Таким образом, в общем случае отношения порядка типа «больше, меньше» и т.п. на множестве нечетких чисел являются нечеткими. Лишь в том случае, когда пересечение носителей нечетких чисел  $A$  и  $B$  пусто, отношение между числами будет четким.

Все предложенные процедуры основаны на вычислении некоторой четкой функции  $H(A, B)$  от нечетких аргументов  $A$  и  $B$ , которая называется индексом ранжирования. Значения индекса для конкретной пары чисел дает основание решить вопрос о том, какое из двух чисел больше (или с какой степенью больше). Будем использовать в дальнейшем следующий индекс ранжирования:

$$H_5^1(A, B) = \sup_{a \geq b} \min \{ \mu_a, \mu_b \}.$$

Индекс  $H_5^1$  выделяет в качестве наибольшего нечеткое число  $A$ , у которого величина  $\sup_{x \in S_A} \arg \sup \mu_A(x)$  является наибольшей (т.е. число, у которого конец «пика» функции принадлежности расположен правее). При этом форма функции принадлежности не учитывается.

Возможность того, что  $A \geq B$  обозначается  $V(A \geq B)$ :

$$V(A \geq B) = \sup_{a \geq b} \min(\mu_A(a), \mu_B(b)).$$

Ответ на вопрос «Может ли  $A$  быть больше  $B$ ?» нечеткий и представляется как нечеткое подмножество множества {да, нет}:

$$A = \{V(A \geq B) / да; V(A \leq B) / нет\}.$$

Для нашего примера:  $V(A \geq B) = 0,48$ ;  $V(B \geq A) = 1$ .

Больше ли  $A$ , чем  $B$ ? Ответ:  $\left\{ \frac{0,48}{да}, \frac{1}{нет} \right\}$ .

Возможность для нечеткого числа  $\tilde{M}$  быть больше  $k$  нечетких чисел  $\tilde{M}_i$ :

$$V(\tilde{M} \geq \tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_k) = \sup_{\substack{t \geq t_1 \\ t \geq t_2 \\ t \geq t_k}} \min(\mu_{\tilde{M}}(t), \mu_{\tilde{M}_1}(t_1), \dots, \mu_{\tilde{M}_k}(t_k)).$$

Замечание. Используя введенную выше процедуру, можно определить так называемое «множество оптимальности альтернатив» -  $O = \{ \mu_o(i) / i \}, i \in N$ ,  $\mu_o(i)$  – интерпретируется как степень соответствия альтернативы  $a_i$  понятию «наилучшая альтернатива».  $R_i(r_i)$  – нечеткая оценка альтернативы  $a_i$  и  $\mu_o(i) = \sup_{r_i \geq r_j; \forall j \neq i} \min(\mu_{R_i}(r_i), \mu_{R_j}(r_j))$ .

## 9.6. Примеры использования теории нечетких множеств

### 9.6.1. Оценка и упорядочение альтернатив при аддитивности критериев

Метод использует аддитивную свертку, обобщенную на случай нечеткой исходной информации.

Пусть необходимо упорядочить  $m$  альтернатив  $a_1, a_2, \dots, a_m$ , оцениваемых по  $n$  критериям  $C_1, C_2, \dots, C_n$ . Пусть оценка альтернативы  $a_i$  по критерию  $C_j$  задается нечетким числом  $R_{ij}$ , относительная важность каждого критерия задается нечетким числом  $W_j$ ,  $j = \overline{1, n}$ . При такой постановке задачи для нахождения решения необходимо произвести следующие действия:

1. находится взвешенная оценка  $i$ -той альтернативы

$$R_i = \sum_{j=1}^n W_j R_{ij} \quad (\text{если оценки нормированы}). \quad (9.28).$$

Оценки альтернатив по критериям и коэффициенты относительной важности задаются функциями принадлежности соответственно  $\mu_{R_{ij}}(r_{ij})$  и  $\mu_{W_j}(\omega_j)$ , где  $r_{ij}, \omega_j \in R^1$ . Т.к. в данном случае  $R_{ij}$  и  $W_j$  являются нечеткими числами,  $R_i$  определяется в соответствии с (9.28) на основе принципа обобщения (см. выше 9.4).

2. находится множество оптимальности альтернатив. После того, как взвешенные оценки  $R_i$  получены, необходимо сравнить альтернативы на их основе. Для этого вводится нечеткое множество  $O$ , заданное на множестве индексов альтернатив  $\{1, 2, \dots, m\}$  и значение соответствующей функции принадлежности интерпретируется как характеристика степени того, насколько альтернатива  $a_i$  является лучшей.

Значение  $\mu_o(i)$  вычисляется как  $\mu_o(i) = \sup_{r_i \geq r_j; \forall j \neq i} \min(\mu_{R_i}(r_i), \mu_{R_j}(r_j))$  и равно, как нетрудно видеть, ординате точки пересечения взвешенной оценки  $i$ -той альтернативы и оценки наилучшей альтернативы (см. 9.5).

### 9.6.2. Нечеткие сети Петри

Один из подклассов нечетких сетей Петри (НСП), называемый в дальнейшем НСП типа  $V_f$  или кратко – НСП  $V_f$ , получается в результате введения описания неопределенности нечеткого характера в начальную маркировку и правила изменения маркировок базового формализма классических СП.

*Нечеткая сеть Петри типа  $V_f$*  определяется как  $V_f = (N, M_0)$ , где:

$N = (P, T, I, O)$  – структура НСП  $V_f$ , в которой:  $P = \{p_1, p_2, \dots, p_n\}$  – конечное множество позиций;  $T = \{t_1, t_2, \dots, t_u\}$  – конечное множество переходов;  $I: P \times T \rightarrow N_0$  – входная функция переходов;  $O: T \times P \rightarrow N_0$  – выходная функция переходов;

$M_0$  – матрица начальной маркировки, размерность которой равна  $(n \times (d+1))$ . Каждый элемент  $m_{ij}^0$  этой матрицы равен значению функции принадлежности наличия  $j-1$  числа маркеров в позиции  $p_i$  НСП на момент начала ее запуска. По определению функции принадлежности элементы матрицы начальной маркировки должны удовлетворять следующему условию:

$$m_{ij}^0 \in [0,1], (\forall i \in \{1,2,\dots,n\}, \forall j \in J). \quad (9.29)$$

Здесь, как и ранее,  $N_0 = \{0, 1, 2, 3, \dots\}$  обозначает множество натуральных чисел и ноль. Множество  $J$  определяется как:  $J = \{1, \dots, d, d+1\} \subset N_0$ , т.е. как некоторое конечное подмножество  $N_0$ , состоящее из  $d+1$  первых натуральных чисел. При этом общее количество столбцов матрицы на-

чальной маркировки определяется максимальным количеством вводимых в рассмотрение маркеров в позициях НСП  $V_f$ , которое в общем случае принимается равным  $d$ .

В определении НСП  $V_f$  каждая строка матрицы  $M_0$  может рассматриваться как значения функции принадлежности нечеткого множества маркеров для соответствующей позиции НСП  $V_f$ . Характерным свойством указанного нечеткого множества является его дискретный характер, что позволяет задать его конечным множеством положительных значений функции принадлежности для носителя данного множества. В этом случае базовый формализм классических СП может быть получен в качестве частного случая НСП  $V_f$ .

Из определения введенного в рассмотрение подкласса НСП  $V_f$  можно видеть, что структура  $N$  НСП  $V_f$  полностью идентична структуре базового формализма классических СП. Поэтому НСП  $V_f$  графически также изображается ориентированным двудольным мультиграфом аналогично обычным СП.

**П р и м е р 1.** Структура изображенной на рис. 9.19 НСП типа  $V_f$ , имеет множество позиций  $P = \{p_1, p_2, \dots, p_n\}$  и множество переходов  $T = \{t_1, t_2, \dots, t_n\}$ . Начальная маркировка для этой НСП  $V_f$  может быть задана, например, следующей матрицей:

$$M_0 = \begin{bmatrix} 0.1 & 0.2 & 0.6 & 0.8 \\ 0.8 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.6 & 0.5 & 0 & 0 \end{bmatrix}$$

Строки матрицы начальной маркировки образованы из векторов  $m_i^0 = (m_{i1}^0, m_{i2}^0, m_{i3}^0, m_{i4}^0)$ , ( $\forall i \in \{1, 2, \dots, n\}$ ), компоненты которых определяют степени принадлежности  $m_{ij}^0 = \mu_i(j-1)$  нечеткого наличия  $j-1$  количества маркеров в позиции  $p_i$  данной НСП. При этом общее количество вво-

димых в рассмотрение маркеров равно  $d=3$ . Так, для позиции  $p_1$  значение  $m_{11}^0=0.1$  означает, что степень принадлежности нечеткого отсутствия маркеров в данной позиции для начальной маркировки равна 0.1. Значение  $m_{12}^0=0.2$  означает, что степень принадлежности нечеткого наличия одного маркера в позиции  $p_1$  для начальной маркировки равно 0.2. Значение  $m_{13}^0=0.6$  означает степень принадлежности нечеткого наличия двух, а значение  $m_{14}^0=0.8$  – степень принадлежности нечеткого наличия трех маркеров в позиции  $p_1$  данной НСП для начальной маркировки. Значения  $m_{15}^0=m_{16}^0=\dots=0$  в матрице  $M_0$  не приводятся.

Другими словами, в матрице начальной маркировки  $M_0$  указывается лишь значения степеней принадлежности для носителя нечеткого множества количества маркеров в позициях НСП  $V_f$ .

Динамика изменения начальной и последующих маркировок НСП  $V_f$  после момента ее запуска подчиняется следующим *Правилам*  $P(V_f)$ :

- $(P_1)$  *Правило определения текущей маркировки.* Любое текущее состояние НСП  $V_f$  определяется некоторой матрицей  $M$  размерности  $(n \times (d+1))$ , элементы которой удовлетворяют условию (9.29) и интерпретируются как значения степеней принадлежности нечеткого наличия  $\{0, 1, 2, \dots, d\}$  маркеров в соответствующих позициях  $p_i \in P$  НСП  $V_f$ . Начальное состояние НСП определяется матрицей начальной маркировки  $M_0$ .

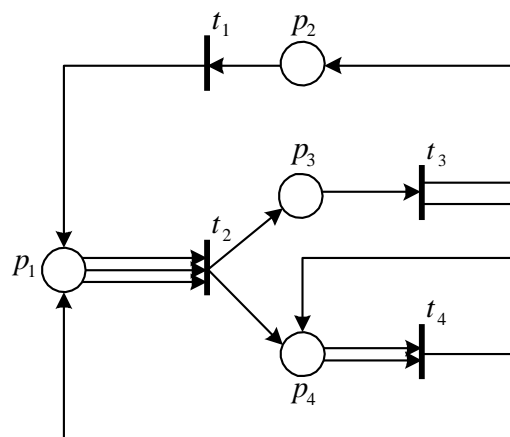


Рис. 9.19. Графическое изображение структуры НСП типа  $V_f$

- (P<sub>2</sub>) *Правило (условие) активности перехода.* Переход  $t_k \in T$  НСП  $V_f$  называется активным (разрешенным, возбужденным) при некоторой текущей маркировке  $M$ , если выполнено следующее условие:

$$\sigma \geq (I(p_1, t_k), I(p_2, t_k), \dots, I(p_n, t_k)), \quad (9.30)$$

где компоненты вектора  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$  определяются по формуле:

$$\sigma_i = \max_{(j \in J) \wedge (m_{ij} > 0)} \{j\} - 1 \quad (\forall p_i \in P), \quad (9.31)$$

т. е. значение равно максимальному индексу отличной от нуля степени принадлежности нечеткого наличия маркеров в позициях НСП  $V_f$ . Другими словами, некоторый переход  $t_k \in T$  НСП  $V_f$  является активным или разрешенным, если для текущей маркировки  $M$  во всех его входных позициях имеется нечеткое количество маркеров, которое больше или равно количеству дуг, соединяющих соответствующие входные позиции с рассматриваемым переходом.

- (P<sub>3</sub>) *Правило срабатывания перехода.* Если переход  $t_k \in T$  НСП  $V_f$  является активным при некоторой текущей маркировке  $M$  (т. е. выполнено условие (9.30)), то срабатывание данного перехода, осуществляемое мгновенным образом, приводит к новой маркировке  $M_v$ , элементы которой определяются следующим образом:

для каждой из входных позиций  $p_i \in P$ , для которых  $I(p_i, t_k) \neq 0$ , по формулам:

$$m_{i1}^v = \max_{(j \in \{1, 2, \dots, I(p_i, t_k) + 1\})} \{m_{ij}\}; \quad (9.32)$$

$$m_{ij}^v = m_{i, j+I(p_i, t_k)}, \quad (\forall j \in J \setminus \{1\}). \quad (9.33)$$

для каждой из выходных позиций  $p_i \in P$ , такой что:  $O(t_k, p_i) \neq 0$  по формулам:

$$m_{ij}^v = \min\{m_{ij}, 1 - q_k\}, \quad (\forall j \in \{1, 2, \dots, O(t_k, p_i)\}); \quad (9.34)$$

$$m_{ij}^v = \max\{\min\{m_{ij}, 1 - q_k\}, \min\{m_{i,j-O(t_k, p_i)}, q_k\}\}, \quad (9.35)$$

$$(\forall j \in J) \wedge (j > O(t_k, p_i)),$$

где  $q_k$  – степень принадлежности или мера возможности нечеткого срабатывания (запуска) перехода  $t_k \in T$ , рассчитываемая по формуле:

$$q_k = \min_{(i \in \{1,2,\dots,n\})} \{ \max_{(j \in J) \wedge (j > I(p_i, t_k))} \{m_{ij}\} \}, (\forall t_k \in T). \quad (9.36)$$

Если некоторые из позиций  $p_i \in P$  являются одновременно входными и выходными для разрешенного перехода  $t_k \in T$ , то для них элементы матрицы новой маркировки  $M_v$  рассчитываются последовательно, вначале по формулам (9.32) – (9.33), а затем – по формулам (9.34) – (9.35).

Проиллюстрируем применение описанных правил  $P(V_f)$  на примере изменения начальной маркировки  $M_0$  НСП  $V_f$ , изображенной на рис. 9.19.

Согласно правилу  $P_2$ , при данной маркировке разрешенным является единственный переход  $t_2$ , поскольку только для него выполнено условие (9.30):  $(3, 0, 0, 1) \geq (3, 0, 0, 0)$ . Срабатывание этого перехода приводит к новой маркировке  $M_1$ , элементы матрицы которой определяются по правилу  $P_3$  следующим образом:

- для позиции  $p_1$ , являющейся входной для перехода  $t_2$  со значением  $I(p_1, t_2)=3$ , с использованием соотношений (9.32) – (9.33):  
 $m_{11}^1 = \max\{0.1, 0.2, 0.6, 0.8\} = 0.8$ ,  $m_{12}^1 = m_{13}^1 = m_{14}^1 = 0$ ;

- для позиций  $p_3$  и  $p_4$ , являющимися выходными для перехода  $t_2$  со значениями  $O(t_2, p_3)=1$  и  $O(t_2, p_4)=1$ , с использованием соотношений (9.34) – (9.35), при этом степень принадлежности нечеткого срабатывания перехода  $t_2$  равна  $q_2=0.8$ :

$$m_{31}^1 = \min\{m_{31}^0, 1-0.8\} = \min\{0.7, 0.2\} = 0.2; \quad m_{32}^1 = \max\{\min\{m_{32}^0, 1-0.8\}, \min\{m_{31}^0, 0.8\}\} = \max\{\min\{0, 0.2\}, \min\{0.7, 0.8\}\} = 0.7; \quad m_{33}^1 = m_{34}^1 = 0, \\ m_{41}^1 = \min\{m_{41}^0, 1-0.8\} = \min\{0.6, 0.2\} = 0.2, \quad m_{42}^1 = \max\{\min\{m_{42}^0, 1-0.8\}, \min\{m_{41}^0, 0.8\}\} = \max\{\min\{0.6, 0.2\}, \min\{0.6, 0.8\}\} = 0.6;$$

$$\min\{m_{41}^0, 0.8\} = \max\{\min\{0.5, 0.2\}, \min\{0.6, 0.8\}\} = 0.6, m_{43}^1 = \max\{\min\{m_{43}^0, 1-0.8\}, \min\{m_{42}^0, 0.8\}\} = \max\{\min\{0, 0.2\}, \min\{0.5, 0.8\}\} = 0.5, m_{44}^1 = 0;$$

- для позиции  $p_2$ , не являющейся для перехода  $t_2$  ни входной, ни выходной, компоненты новой маркировки  $M_1$  будут равны компонентам исходной маркировки  $M_0$ , т.е.  $m_2 = (0.8, 0, 0, 0)$ .

Таким образом, после нечеткого срабатывания перехода  $t_2$  начальная маркировка  $M_0$  НСП  $V_f$  изменится на маркировку  $M_1$ , матрица которой равна:

$$M_1 = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 \\ 0.2 & 0.7 & 0 & 0 \\ 0.2 & 0.6 & 0.5 & 0 \end{bmatrix}.$$

Полученная маркировка  $M_1$  разрешает переходы  $t_3$  и  $t_4$ , в чем можно убедиться, используя правило  $P_2$ . Поэтому процесс изменения маркировок для данной НСП  $V_f$  может быть продолжен.

Поскольку процесс изменения маркировок НСП  $V_f$  во многом аналогичен соответствующему процессу для базового формализма классических СП с такой же структурой, хотелось бы представить динамику изменения маркировок НСП  $V_f$  также в виде дерева или диаграммы достижимых маркировок. Строгое определение диаграммы достижимых маркировок НСП  $V_f$  базируется на отношениях непосредственного следования и достижимости маркировок, определения которых приводятся ниже.

Маркировка  $M_v$  НСП  $V_f = (N, M_0)$  непосредственно достижима из маркировки  $M$  (непосредственно следует за маркировкой  $M$ ) этой же НСП, если маркировка  $M_v$  получается в результате срабатывания некоторого разрешенного при  $M$  перехода  $t_k \in T$ . Отношение непосредственного следования маркировок НСП  $V_f$  будем обозначать через:  $M - (t_k) \rightarrow M_v$ , где  $M$  и  $M_v$  – матрицы соответствующих маркировок.

Маркировка  $M_w$  НСП  $V_f = (N, M_0)$  называется *достижимой* из маркировки  $M$  этой же НСП, если существует конечная упорядоченная последовательность маркировок  $M_k = \langle M_1, M_2, \dots, M_q \rangle$ , где  $M_1=M$ ,  $M_q=M_w$ , и соответствующая ей конечная упорядоченная последовательность переходов  $\pi_1 = \langle t_1, t_2, \dots, t_{q-1} \rangle$  такие, что любая пара соседних маркировок в  $M_k$  удовлетворяет отношению непосредственного следования маркировок:  $M_j \xrightarrow{(t_j)} M_{j+1}$  ( $\forall j \in \{1, 2, \dots, q-1\}$ ). Отношение достижимости для НСП  $V_f$  будем обозначать через  $M \xrightarrow{(\pi_1)} M_w$ .

Обозначим через  $D(M)$  множество маркировок, достижимых из некоторой фиксированной маркировки  $M$  для заданной НСП  $V_f = (N, M_0)$ , т.е.  $D(M) = \{M_w \mid (\exists \pi_1) M \xrightarrow{(\pi_1)} M_w\}$ .

Множество  $D(M_0)$  будем называть *множеством достижимых маркировок* для НСП  $V_f = (N, M_0)$  или просто – *множеством достижимости* НСП  $V_f$ . *Диаграммой достижимых маркировок* (графом достижимости) НСП  $V_f$  будем называть ориентированный мультиграф  $D(V, E, \gamma)$ , где  $V = \{v_1, v_2, \dots, v_q\}$  – множество вершин мультиграфа, которое соответствует множеству достижимых маркировок  $D(M_0)$ ;  $E = \{e_{ij}^k\}$  ( $\forall v_i, v_j \in V$ ) – в общем случае мультимножество (комплект) дуг, которое соответствует отношению непосредственной достижимости маркировок НСП  $V_f$ ;  $\gamma: E \rightarrow T$  – отображение, помечающее каждую дугу мультиграфа переходом, срабатывание которого обеспечивает выполнение соответствующего отношения непосредственной достижимости для рассматриваемой пары маркировок.

Диаграмма достижимых маркировок, построенная для НСП из примера 1 в форме диаграммы деятельности нотации UML, представлена на рис.9.20.

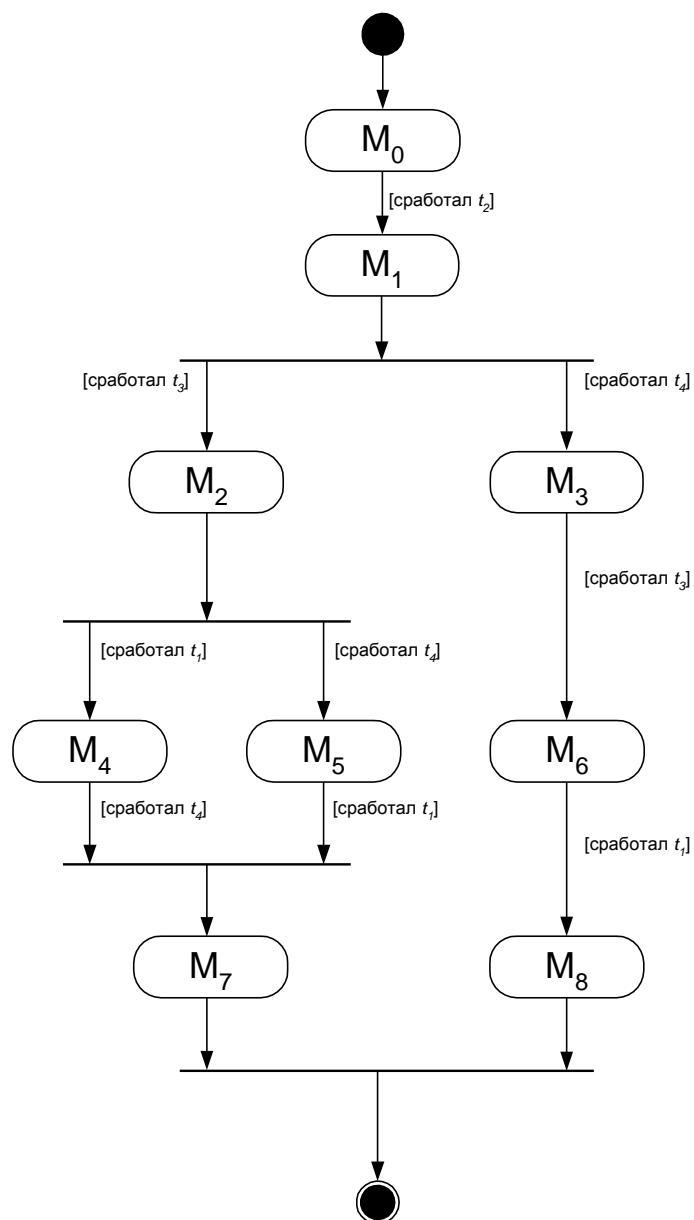


Рис. 9.20. Диаграмма достижимости маркировок для НСП  $V_f$  из примера 1  
 Для наглядности значения матриц всех достижимых маркировок записаны отдельно от диаграмм

$$M_1 = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 \\ 0.2 & 0.7 & 0 & 0 \\ 0.2 & 0.6 & 0.5 & 0 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.6 & 0 \end{bmatrix}$$

$$\begin{aligned}
M_3 &= \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.8 & 0 & 0 & 0 \\ 0.2 & 0.7 & 0 & 0 \\ 0.6 & 0 & 0 & 0 \end{bmatrix} & M_4 &= \begin{bmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.6 & 0.5 \end{bmatrix} \\
M_5 &= \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.6 & 0.5 & 0 & 0 \end{bmatrix} & M_6 &= \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.3 & 0.6 & 0 & 0 \end{bmatrix} \\
M_7 &= \begin{bmatrix} 0.3 & 0.4 & 0.6 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.6 & 0.5 & 0 & 0 \end{bmatrix} & M_8 &= \begin{bmatrix} 0.3 & 0.5 & 0.5 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.3 & 0.6 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Как показывает анализ данной диаграммы достижимости (см. рис. 9.20), для рассматриваемой НСП  $V_f$  справедливы следующие отношения:  $M_0 \rightarrow (t_2) \rightarrow M_1$ ,  $M_1 \rightarrow (t_3) \rightarrow M_2$ ,  $M_1 \rightarrow (t_4) \rightarrow M_3$  и  $M_0 \rightarrow (< t_2, t_3 >) \rightarrow M_2$ .

Для НСП  $V_f$  могут быть введены в рассмотрение свойства, соответствующие приведенным выше свойствам классических СП. В этом случае при решении задач системного моделирования с использованием рассматриваемых моделей НСП  $V_f$  основной интерес также представляют множества  $D(M_0)$  и соответствующие им диаграммы достижимости, которые содержат исчерпывающую информацию об соответствующих свойствах НСП  $V_f$ .

**О г р а н и ч е н н о с т ь п о з и ц и й.** Позиция  $p_i \in P$  НСП  $V_f = (N, M_0)$  называется  $k$  – ограниченной, если для любой достижимой маркировки  $M \in D(M_0)$  выполняется условие:  $\max\{m_{ij} \mid m_{ij} > 0\} \leq k-1$  для некоторого фиксированного значения  $k \in \{1, 2, 3, \dots\}$ . Здесь максимум берется по всем компонентам вектора  $m_i$ , соответствующего  $i$ -ой строке матрицы достижимой маркировки  $M$ .

НСП  $V_f = (N, M_0)$  называется  $k$  – ограниченной, если все позиции множества  $P$  являются  $k$ -ограниченными.

**Безопасность позиций.** 1-ограниченная позиция НСП  $V_f = (N, M_0)$  называется безопасной позицией. 1-ограниченная НСП  $V_f = (N, M_0)$  называется безопасной НСП.

**Устойчивость переходов.** Переход  $t_i$  НСП  $V_f$  называется устойчивым, если этот переход  $t_i$  остается активным при срабатывании любого другого активного при этой же маркировке перехода  $t_j$ . НСП называется устойчивой, если все ее переходы являются устойчивыми.

**Уровни активности переходов** НСП. Пусть задана НСП  $V_f$ .

Переход  $t_i \in T$  обладает активностью уровня 0 и называется пассивным или мертвым, если не существует ни одной достижимой в этой НСП маркировки, при которой этот переход был бы активным.

Переход  $t_i \in T$  обладает активностью уровня 1 и называется потенциально активным или живым, если существует достижимая в этой НСП маркировка, при которой этот переход является активным.

Переход  $t_i \in T$  обладает активностью уровня 2, если для любого натурального  $s \in \mathbb{N}$  в этой НСП существует последовательность срабатываемых переходов  $\pi_k$ , в которой  $t_i$  присутствует по крайней мере  $s$  раз.

Переход  $t_i \in T$  обладает активностью уровня 3, если в этой НСП существует последовательность срабатываемых переходов  $\pi_k$ , в которой переход  $t_i$  присутствует неограниченное число раз.

Переход  $t_i \in T$  обладает активностью уровня 4 и называется активным или живым, если для любой достижимой в этой НСП маркировки существует последовательность срабатываемых переходов  $\pi_k$ , так что  $t_i$  является активным при некоторой другой достижимой в этой НСП маркировке, получаемой в результате последовательного срабатывания переходов из  $\pi_k$ .

Переход  $t_i \in T$  называется потенциально мертвым, если существует некоторая достижимая в этой НСП маркировка, такая, что при любой достижимой из данной маркировке переход  $t_i$  не является активным.

Рассматриваемая НСП обладает активностью уровня  $k$ , где  $k \in \{0, 1, 2, 3, 4\}$ , если все ее переходы обладают активностью уровня  $k$ .

**Тупиковая маркировка.** Достижимая в НСП маркировка называется  $t_i$ -тупиковой ( $t_i \in T$ ), если переход  $t_i$  потенциально мертвый для данной маркировки. Если рассматриваемая достижимая в этой НСП маркировка  $t_i$  – тупиковая для всех переходов  $t_i \in T$ , то она называется тупиковой.

При моделировании процессов функционирования дискретных динамических систем с использованием НСП также могут быть сформулированы некоторые задачи, аналогичные задачам базового формализма СП. Так, например, задача достижимости маркировки в НСП состоит в установлении отношения достижимости некоторой заданной маркировки из начальной маркировки для рассматриваемой НСП.

### **9.7. Вопросы для самопроверки**

1. Приведите определение нечеткого множества и назовите способы его задания.
2. Перечислите основные характеристики (параметры) нечетких множеств.
3. Как вводятся операции над нечеткими множествами?
4. Какие существуют способы задания нечетких отношений?
5. Приведите определения понятий нечеткой и лингвистической переменных.

6. Назовите основные типы функций принадлежности нечетких множеств.

7. В чем заключаются особенности косвенных методов построения функций принадлежности?

8. Сформулируйте процедуры арифметических операций над нечеткими числами и сравнения нечетких чисел.

# ГЛАВА 10. КОГНИТИВНЫЕ ПОДХОДЫ К РЕШЕНИЮ СЛАБОСТРУКТУРИРОВАННЫХ И ПЛОХО ФОРМАЛИЗОВАННЫХ ЗАДАЧ

## 10.1 Когнитивные модели

Интенсивное проникновение методов искусственного интеллекта в различные сферы деятельности для решения сложных задач управления и поддержки принятия решений не может быть плодотворным без привлечения информации, которая не выражается количественно. Часто это смысловая, качественная (не количественная) информация. Существуют области ответственной и сложной деятельности, в которых информация и знания меняются очень быстро. Конечно, лучше всего использовать в этих случаях соответствующие математические модели. Однако их построение достаточно сложно и не всегда возможно. В этих случаях целесообразно использование более простых – качественных моделей.

Как соотносятся между собой последовательность разных моделей? На рис.10.1 показан процесс последовательно усложняющихся моделей.

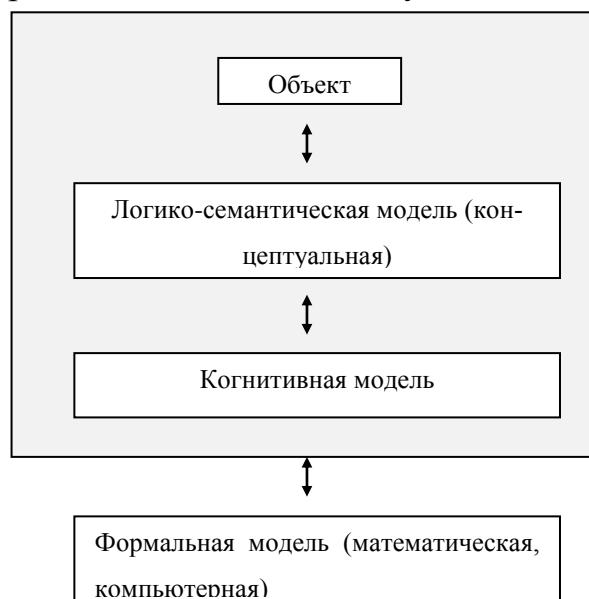


Рис 10.1. Процесс последовательно усложняющихся моделей

По мере получения информации об объекте модель его уточняется, переходя на новый качественный уровень. На первом этапе выясняется, какие переменные характеризуют модель, формируется так называемая логико-семантическая модель.

Например, в самом упрощенном случае, можно предположить, что прибыль предприятия  $P$  зависит от: рыночной стоимости  $R$  товара, выпускаемого предприятием, издержек производства  $J$  предприятия, объёма научно-исследовательских работ  $N$  предприятия. В этом случае логико-семантическая модель прибыли предприятия будет иметь вид

$$P = f(J, N, R) \quad (10.1)$$

Естественно, показатели, входящие в зависимость (10.1), могут в свою очередь зависеть от других показателей, так, что исходная зависимость имеет иерархический вид.

На втором этапе строят более информативную, когнитивную модель, в которой устанавливаются на основе экспертной информации взаимосвязи между переменными.

На третьем этапе создаётся более формализованная модель, в которой идентифицируются параметры взаимосвязей между переменными. Часто эта формализованная модель описывает динамический процесс изменения переменных во времени и пространстве.

*Когнитивное моделирование* является одним из классов имитационного моделирования, в основе которого лежит построение и исследование *когнитивной карты* ситуации. Для этих целей используется аппарат взвешенных знаковых и функциональных знаковых графов. Он позволяет работать с данными как качественного, так и количественного типа. Поэтому его достаточно удобно использовать при исследовании развития и функционирования сложных систем различного типа.

Когнитивное моделирование применяется при работе с слабоструктурированными (слабоформализованными) системами, может использоваться в тех областях, где никогда не делались попытки провести моделирование. В таких системах невозможен традиционный эконометрический подход к анализу процессов для выработки комплексных (т.е. затрагивающих различные аспекты исследуемой системы) решений.

Что такое когнитивная карта? Когнитивная карта есть то или иное представление связей, которые существуют между атрибутами в данной предметной области знания. Если отношения численные, то они могут быть положительными, отрицательными или нейтральными и характеризовать положительное или отрицательное влияние друг на друга. Применяя специальные алгоритмы, можно на когнитивной карте (заданной графом) «проследить» путь влияния положительных и отрицательных связей на один или несколько когнитивных элементов. Если эти когнитивные элементы будут по смыслу отражать понятие цели, то можно уже говорить о принятии решений по отдельным целям.

Как может использоваться когнитивная карта для принятия стратегических решений? При когнитивном моделировании создается модель нашего воображения, концепты и отношения, характеризующие предметную область, «связываются» друг с другом. Затем определяются и выделяются стратегические цели и синтезируются стратегии по их достижению. Таким образом, вводится понятие управляемого когнитивного процесса для использования его при поддержке принятия стратегических решений.

Термин когнитивная карта появился задолго до возникновения когнитологии в работе выдающегося американского психолога Э. Толмена "Когнитивные карты у крыс и человека" (1948). Анализируя поведение крыс в лабиринте, Толмен пришел к выводу, что в результате беганья по лабиринту у крысы формируется особая структура, которую можно на-

звать когнитивной картой окружающей обстановки. Толмен предположил, что когнитивные карты определяют также поведение людей, и призывал к "расширению" когнитивных карт как к средству борьбы с избытком агрессивности — только так, по его мнению, можно научить людей адекватно ориентироваться в решении жизненных проблем. Таким образом, когнитивную карту можно понимать как схематичное, упрощенное описание картины мира индивида, точнее, ее фрагмента, относящегося к данной проблемной ситуации.

Когнитивные карты были впервые предложены для работы с представлением ситуации Р. Аксельродом в 1976 году для моделирования причинных взаимосвязей, выявленных между концептами некоторой области. Позднее Ф.С. Робертс в 1984 предложил методы расчета динамики развития на этих картах. Особым видом когнитивных карт являются нечеткие когнитивные карты, предложенные Б. Коско в 1986 для представления нечетких причинно-следственных связей.

Когнитивная карта – ключевое понятие в когнитивном моделировании. Сама когнитивная карта представляет собой ориентированный взвешенный граф  $G(V,A)$ . Переменные в когнитивной модели, именуемые базисными факторами, на когнитивной карте представляют собой вершины (V). Дуги (A) моделируют причинно-следственные взаимосвязи между факторами (усиливающие – положительные; тормозящие – отрицательные).

На рис.10.2 представлена когнитивная карта упрощенной экономической модели предприятия, логико-семантическая модель которой представлена формулой (10.1). Стрелками обозначены взаимосвязи между факторами; если увеличение фактора 1 приводит к увеличению фактора 2, стрелка направлена от первого фактора ко второму.



Рис.10.2 Когнитивная карта экономики предприятия

## 10.2. Суть когнитивного подхода. Когнитивный анализ

Когнитивный анализ предусматривает последовательную причинно-следственную структуризацию информации о происходящих в исследуемой системе процессах в соответствии со следующей схемой:

- всякое событие, произошедшее в системе, вызывается определенными причинами (предпосылками), появление которых связано с движением материальных потоков (товары, деньги, ресурсы и т.п.) и нематериальных потоков (информационные взаимодействия). Движение каждого потока может быть описано в самом общем виде соответствующими цепочками причинно-следственных отношений, составляющих знания аналитика или его предположения о действующих в данной системе закономерностях. Вычленение таких потоков является первым шагом при когнитивном анализе исследуемой системы;

- каждый из выделенных потоков описывается соответствующей совокупностью факторов. Объединение всех этих совокупностей составляет

множество факторов, в терминах которых описываются процессы в системе;

- определяются взаимосвязи между факторами путем рассмотрения причинно-следственных цепочек, описывающих движение каждого потока. Считается, что факторы, входящие в первую часть «если...» цепочки «если то...», влияют на факторы её второй части «то...», причем это влияние может быть либо усиливающим (положительным) либо тормозящим (отрицательным), либо переменного знака в зависимости от возможных дополнительных условий;

- сила воздействия факторов друг на друга описывается с помощью лингвистических переменных типа "значительная", "умеренная", "слабая" и т.п. Сопоставим совокупности таких лингвистических переменных некоторую числовую шкалу так, что каждой переменной будет соответствовать некоторое число в этой шкале. В качестве такой шкалы выберем интервал  $[0,1]$ .

Взаимовлияния факторов отображаются с помощью когнитивной карты, являющейся, по существу, моделью исследуемой системы в виде взвешенного орграфа  $G(V,A)$ , в котором:

- множество вершин взаимно однозначно соответствует совокупности  $V$  рассматриваемых факторов слабоструктурированной системы;

- если влияние фактора  $v_i$  на фактор  $v_j$  положительно (отрицательно), то в графе  $G$  проводится дуга  $a_{ij}$  с весом  $+w_{ij}$  ( $-w_{ij}$ ) от вершины  $v_i$  к  $v_j$ . Число  $w_{ij} \in [0,1]$  отображает степень воздействия фактора  $v_i$  на  $v_j$  в выбранной шкале  $[0,1]$ . Таким образом, каждой дуге приписывается определенный вес, являющийся числом из интервала  $[-1,1]$ ;

- $A = \|\alpha_{ij}\|$  матрица смежности взвешенного орграфа  $G(V,A)$ , размера  $n \times n$ .

В зависимости от условий, в которых аналитик разрешает поставленную перед ним проблему, возможно использование какого-либо из следующих методов построения когнитивных карт.

1. Когнитивная карта строится аналитиком на основе своих знаний и представлений о разрешаемой проблеме без привлечения экспертов и справочных материалов. Эффективность этого метода определяется квалификацией аналитика, его знаниями и умением определять необходимый для решения проблемы набор базисных факторов и характер отношений между факторами. Построение когнитивной карты позволяет аналитику яснее представить себе проблему, лучше понять роль отдельных факторов и характер отношений между ними. Это, скорее всего, самый быстрый способ построения когнитивной карты, обеспечивающий, кроме того, конфиденциальность.

2. Построение когнитивных карт на основе изучения документов. Этот метод имеет два преимущества: он удобен и позволяет использовать данные, имеющиеся у аналитика. Однако изучение документов экспертами - процесс достаточно трудоемкий, а согласование мнений экспертов требует специальной технологии.

3. Построение когнитивных карт на основе опросов группы экспертов, имеющих возможность оценивать причинно-следственные связи в исследуемой системе. Преимущество этого метода - в возможности агрегировать индивидуальные мнения. При этом можно использовать разные механизмы агрегирования - от простого усреднения оценок разных экспертов, исключения «выпадающих» оценок и т.п. до специализированных процедур выработки коллективных решений.

4. Построение когнитивных карт, основанное на открытых выборочных опросах. Этот метод может быть использован для построений сравнимых когнитивных карт. Его достоинство заключается в предоставлении

исследователю возможности вести активный диалог с источниками информации.

Графовая модель слабоструктурированной системы строится с двойкой целью:

- провести некоторую (по возможности, более детальную в пределах имеющейся информации) структуризацию рассматриваемой слабоструктурированной системы;

- с использованием такой структуризации решить исходные проблемы, возникшие в рассматриваемой системе.

Под решением всякой проблемы, исследуемой в рамках графовой модели, будем понимать обеспечение требуемой динамики факторов, составляющих суть данной проблемы. Эти факторы (являющиеся «выходными» факторами графовой модели) назовем *целевыми* факторами.

*Управляющие* факторы («входные» факторы графовой модели) – это те факторы («рычаги»), с помощью которых аналитик может управлять процессами в модели, подавая на них те или иные управляющие воздействия.

*Внешние* факторы модели – это факторы, являющиеся висячими вершинами в графовой модели. Внешние факторы соответствуют тем факторам внешней среды, динамика которых не зависит от состояния слабоструктурированной системы. Таким образом, динамика внешних факторов модели задается аналитиком не с целью управления процессами в модели, а лишь для отражения соответствующих влияний внешней среды на моделируемую систему.

*Задача управления динамикой модели* состоит в выборе такой совокупности управляющих воздействий (управлений), которые компенсируют нежелательные для аналитика последствия начальных условий и обеспечивают заданную динамику изменения целевых факторов.

### 10.3. Моделирование динамики слабоструктурированной системы

Для моделирования временных процессов, происходящих в слабоструктурированной системе, необходимо перейти от графовой модели  $G(V,A)$  к временным уравнениям для факторов, отразив в структуре этих уравнений структурные свойства графа  $G(t)$ . Компьютерная реализация модели слабоструктурированной системы предусматривает исходное описание такой модели в дискретном времени. Моменты дискретного времени будут обозначаться символом  $t$  ( $t=0,1,2,\dots$ ).

Разность  $v_i(t) - v_i(t-1) = p_i(t)$  называется импульсом в вершине  $v_i$  в момент  $t$ . Взаимосвязь между импульсами описывается следующим соотношением.

$$p_i(t+1) = \sum_{j \in I_i} a_{ji} p_j(t) + p_j^0(t), \quad (10.2)$$

где  $I_i$  – совокупность индексов дуг, входящих в вершину  $v_i$ ,  $p_j^0(t)$  – внешний импульс, вносимый в вершину  $v_i$  в момент  $t$ .

Обозначим через  $u_i$  внешний вход в вершину  $v_i$ , так, что

$$p_i^0(t+1) = u_i(t+1) - u_i(t). \quad (10.3)$$

Будем считать, что

$$v_i(t)=0, t < 0, i=1 \dots n,$$

а внешние импульсы подаются, начиная с  $t=0$ . Тогда получим

$$v_i(t+1) = \sum a_{ij} v_j(t) + u_j^0(t), i=1, \dots, n \quad (10.4)$$

или в матричной форме

$$v(t+1) = (I + A)v(t) + Bu(t), \quad (10.5)$$

где:  $A = \|a_{ij}\|$ , матрица смежности размера  $n \times n$  графа  $G(V,A)$ ,  $I$  – единичная ( $n \times n$ ) – матрица,  $u = (u_1, \dots, u_m)^T$  – вектор управлений (внешних

управляющих воздействий), а  $B$  - матрица размера  $n \times m$ , ненулевые элементы которой указывают точки приложения управлений.

Если матрица  $A$  неустойчива (т.е. некоторые из её собственных значений по модулю больше 1), то временные процессы для некоторых факторов будут расходящимися, что делает невозможным использование такой модели. В таком случае необходимо преобразование исходной матрицы  $A$  в устойчивую матрицу, что можно сделать умножением каждого столбца (или строки) матрицы  $A$  на множитель  $1/(s_i + \varepsilon)$ , где  $s_i$  - число ненулевых элементов  $i$ -го столбца (строки), а  $\varepsilon$  - малое число. Поскольку исходная матрица  $A$  является слабо заполненной, она может содержать столбцы лишь с одним ненулевым элементом. Но тогда после умножения её строк (или столбцов) на множители вида  $1/s_i$  норма модифицированной матрицы  $A_m$  будет равна 1, что не гарантирует сходимости некоторых процессов, определяемых уравнениями вида (10.4). Именно для сходимости всех процессов и добавляется некоторое  $\varepsilon > 0$  в сомножители вида  $1/(s_i + \varepsilon)$ .

Для простоты записи будем считать, что в уравнении (10.5) матрица  $A$  изначально устойчивая или сделана устойчивой указанным выше образом.

К уравнению (10.5) можно добавить уравнение, выделяющее наблюдаемые  $k$  факторов из всей совокупности  $n$  базисных факторов:

$$y(t) = Cv(t), \quad (10.6)$$

где  $C$  - матрица размера  $k \times n$ , указывающая, какие из факторов  $v_1, \dots, v_n$  "наблюдаемы", т.е. интересуют исследователя в первую очередь.

Уравнения (10.5)–(10.6) являются динамической моделью слабоструктурированной системы. Поскольку факторы модели (10.5)–(10.6) в соответствии с уравнениями вида (10.4) могут принимать положительные и отрицательные значения, будем считать эти значения оценками приращений факторов. Это означает, что модель (10.5)–(10.6) определена на тен-

денциях факторов. В рамках такой интерпретации определим понятия начального состояния модели и цели управления.

Начальное состояние  $x(0)$  модели – совокупность направлений и скоростей изменения всех факторов модели в начальный момент времени. Цель (вектор цели) управления – это вектор  $u^*$  целевых факторов с заданными значениями направления и скорости изменения всех его координат.

*Моделирование саморазвития ситуации.* Прежде чем решать задачу поиска управлений, обеспечивающих целенаправленное развитие процессов в модели, необходимо проанализировать их свободное развитие (без подачи управлений) из заданного начального состояния  $x(0)$ . Хотелось бы, чтобы структура модели и начальные условия  $x(0)$  оказались такими, что её состояние в свободном движении (за счет лишь взаимодействия факторов) будет смещаться «в сторону цели», т.е. текущие тенденции изменения целевых факторов будут приближаться к желательным целевым тенденциям. Динамика свободного движения состояния  $x(t)$  модели (10.5) описывается уравнением

$$x(t) = (I + A + A^2 + \dots + A^t)x(0). \quad (10.7)$$

Если при этом текущие значения (т.е. оцениваемые тенденции) целевых факторов в векторе  $x(t)$  "уходят" от заданных тенденций целевых факторов, то модель движется "не в ту сторону" и необходимы либо коррекция цели, т.е. изменение количества и/или числа целевых факторов, либо коррекция её поведения с использованием внешних по отношению к модели управлений.

В обоих случаях необходим предварительный структурный анализ соотношений между интегральными (т.е. непосредственными и опосредованными) влияниями факторов друг на друга.

*Структурный анализ модели.* Из построения матрицы  $A$  взаимовлияний факторов следует, что каждый фактор помимо непосредственных

влияний от "соседних" факторов подвержен также опосредованным влияниям со стороны более "отдаленных" факторов. Полная совокупность влияний, которым подвержен каждый фактор, описывается рядом  $I + A + A_2 + \dots + A_t + \dots$ , сумма которого при устойчивой матрице  $A$  практически совпадает с матрицей  $Q$ , называемой транзитивным замыканием матрицы  $A$ . Интегральное воздействие фактора  $x_i$  на фактор  $x_j$ , найденное с учетом как непосредственных, так и всех опосредованных воздействий, есть  $(i, j)$ -й элемент матрицы  $Q$ . При этом важным моментом является то, что каждый  $(i, j)$ -й элемент матрицы  $Q$  связывает знак исходного приращения  $\Delta x_i$  фактора  $x_i$  со знаком интегрального приращения  $\Delta x_j$  фактора  $x_j$ . Другими словами, если в ситуации возмущается (получает приращение) только фактор  $x_i$ , то с учетом всех опосредованных влияний фактора  $x_i$  на  $x_j$  знак результирующего приращения фактора  $x_j$  определится как

$$\text{sign}(\Delta x_j) = \text{sign}(q_{ij}) \text{sign}(\Delta x_i). \quad (10.8)$$

При формировании вектора цели в рамках данной модели аналитик должен указать, какое направление изменения (возрастание, убывание) каждого целевого фактора он рассматривает как желательное. Желательное направление изменения фактора  $x_i$  будем характеризовать показателем  $r_i$ , принимающим значение  $+1$ , если требуется увеличение значения фактора  $x_i$ , и  $-1$ , если необходимо уменьшение фактора  $x_i$ . Показатель  $r_i$  назовем "оценкой пользователем изменения фактора (ОИФ)  $x_i$ ". Полагаем, что при постановке целей никакой из показателей ОИФ целевых факторов не может быть равен нулю.

Пусть  $Y$  - множество целевых факторов, а  $r$  - вектор показателей ОИФ этих факторов. Цель  $y^*$  называется непротиворечивой, если изменение любого целевого фактора в желательном направлении не приводит (через це-

почку «промежуточных» факторов) к изменению какого-либо другого целевого фактора в нежелательном направлении. Формально условие непротиворечивости цели выглядит как

$$r_i \text{sign}(q_{ij}) = r_j \quad , \quad (10.9)$$

где  $r_i$  и  $r_j$  - любые ненулевые координаты вектора  $r_i$ , а  $q_{ij}$  – соответствующие элементы матрицы  $Q$ . При невыполнении условия (10.9) хотя бы для одного целевого фактора, цель называется противоречивой.

Совокупность управляющих факторов  $U$  называется согласованной с целью, если для каждого управляющего фактора найдется такое направление его изменения, которое не приводит к изменению какого-либо из целевых факторов в нежелательном направлении. В рамках модели (10.5-10.7) условие согласованности управляющих факторов из множества  $U$  с целью  $u^*$  означает, что для любого управляющего фактора  $u_i$  из  $U$  возможно такое его изменение со знаком  $\text{sign}(u_i)$ , которое не может привести к нежелательным изменениям каких-либо целевых координат, что можно записать так:

$$\forall u_i \in U \exists \text{sign}(u_i) \forall x_i \in Y (\text{sign}(u_i) \text{sign}(q_{ij}) = r_i) \quad . \quad (10.10)$$

Вернемся к результатам моделирования саморазвития ситуации. Если оказалось, что под влиянием начальных условий состояние модели не движется «в сторону цели», то необходимо либо проанализировать исходную цель на непротиворечивость и если она оказалась противоречивой, то скорректировать исходную цель и для скорректированной цели ещё раз промоделировать режим саморазвития ситуации; либо перейти к поиску управляющих воздействий, с помощью которых возможно «переломить» негативное развитие ситуации и направить её «в сторону цели».

*Моделирование управляемого развития и решение обратной задачи управления. Шкалы и единицы измерения.* Нахождение искомой совокуп-

ности управляющих воздействий производится путем решения обратной задачи управления, суть которой состоит в следующем. Выберем каждое управляющее воздействие  $g(t)$  в виде «мгновенного» импульса величины  $g(0)$ , подаваемого в момент  $t=0$ . Тогда динамика модели (10.5) описывается уравнением

$$x(t) = (I + A + A^2 + \dots + A^t)x(0) + (I + A + A^k + \dots + A^{i-1})Bg(0) . \quad (10.11)$$

При обеспечении устойчивости матрицы  $A$  ряд  $(I + A + A^2 + \dots + A^t)$  сходится довольно быстро с ростом  $t$ , так что в нем достаточно рассмотреть лишь  $N$  членов, сумма которых практически совпадает с матрицей  $Q$ .

Задачу управления будем считать решенной, если найден вектор  $g^*(0)$  управляющих воздействий такой, что значения целевых координат в установившемся состоянии модели совпадают с заданными целевыми значениями (координатами вектора  $y^*$ ), что можно записать с учетом (10.11) так:

$$y_{уст} = Cx_{уст} = CQx(0) + CQBg(0) = y^* , \quad (10.12)$$

где элементы матрицы  $C$  указывают, какие из факторов модели являются целевыми,  $x_{уст}$  - установившееся состояние модели, а  $y^*$  - заданный вектор цели. Искомый вектор управляющих воздействий находится методом наименьших квадратов (МНК):

$$g^*(0) = (CQB)^{-1}(y^* - CQx(0)) , \quad (10.13)$$

где  $(CQB)^{-1}$  – обратная матрица для матрицы  $CQB$ . Вектор  $g^*(0)$  является наилучшим с точки зрения величины невязки правой и левой частей соотношения (10.8).

Возникает естественный вопрос - не приведет ли подача управлений из  $g^*(0)$ , полученного чисто формальным методом, к недопустимым (с точки зрения поставленной цели) изменениям целевых факторов? Дело в том, что в случае векторной цели каждое управляющее воздействие может из-

менять всю совокупность целевых факторов и не все целевые факторы при этом могут изменяться желательным образом. Что может произойти с целевыми факторами, если подать на модель всю совокупность управлений - координат вектора  $g^*(0)$ ?

Ответ на это вопрос содержится в следующем утверждении в понятиях структурного анализа: если цель непротиворечива и совокупность управляющих факторов согласована с целью, то совокупность управлений - координат вектора  $g^*(0)$  - изменяет каждый из целевых факторов в желательном направлении. Если же условие (10.12) не выполняется, то необходимо проанализировать «качество» найденного МНК-решения  $g^*(0)$  (в плане влияния его координат на целевые факторы) путем моделирования и при необходимости - скорректировать его.

Таким образом, структурный анализ помогает аналитику в решении следующих вопросов:

- если поставленная цель противоречива, то имеет смысл сразу внести в неё коррективы, поскольку противоречивая цель, скорее всего, не будет достигнута;
- если совокупность управляющих факторов не согласована с целью, то это приведет к дополнительной затрате ресурсов (дополнительному увеличению необходимых управляющих воздействий) для достижения цели.

В силу указанных причин структурный анализ следует проводить до решения задачи управления.

При построении когнитивных карт и проведении сценариев когнитивного моделирования все факторы, их начальные и конечные значения, время, веса взаимосвязей – принимаются безразмерными и меняются в диапазоне  $[-1;+1]$ . Для перевода количественных значений в качественные используется шкала желательности Харрингтона. В таблице 10.1

приведены диапазоны значений на шкале желательности в зависимости от вида желательности.

Таблица 10.1 Шкала желательности Харрингтона

Желательность	Отметки на шкале желательности
Очень хорошо	1.00 - 0.80
Хорошо	0.80 - 0.63
Удовлетворительно	0.63 - 0.37
Плохо	0.37 - 0.20
Очень плохо	0.20 - 0.00

#### 10.4. Аналитическая программная система «СТРАТЕГ»

АПС «СТРАТЕГ», разработанная в Волгоградском государственном техническом университете, основана на когнитивном подходе и обеспечивает большие функциональные возможности для построения и анализа когнитивных моделей. АПС «Стратег» может применяться для моделирования возможных путей развития ситуаций, формирования и анализа различных стратегий, ситуационного моделирования.

Система предназначена для поддержки принятия управленческих решений, выявления и наглядного представления знаний эксперта о предметной области, помощи в изучении и анализе функционирования сложных слабоструктурированных систем.

Область применения АПС «СТРАТЕГ» – любые слабоструктурированные системы в различных сферах: социология, политика, экономика, управление, техника и т.д.

Основные возможности АПС «СТРАТЕГ»:

- Вариативность построения моделей (таблицы, граф);

- Расчет начальных тенденций;
- Нахождение взаимовлияний по статистическим данным;
- Определение опосредованных влияний;
- Определение взаимосвязанности факторов,  
Парето-анализ взаимовлияний и тенденций;
- Широкие возможности работы с моделью на графе;
- Анализ целей на противоречивость;
- Анализ согласованности управлений;
- Моделирование свободного и управляемого развития;
- Решение обратной задачи управления;
- Генерация отчетов;

АПС «СТРАТЕГ» состоит из 7 подсистем:

- подсистема извлечения знаний эксперта;
- подсистема визуализации данных;
- подсистема анализа;
- подсистема моделирования;
- подсистема работы со сценариями;
- подсистема построения отчетов;
- подсистема импорта/экспорта.

Подсистема извлечения знаний эксперта реализует методы извлечения знаний эксперта о проблемной ситуации, а также необходимый функционал для построения когнитивной модели. Построение модели реализовано двумя способами – визуально с помощью прямого редактирования когнитивной карты и посредством редактирования списка факторов и попарного оценивания силы влияния факторов.

Подсистема визуализации данных предназначена для представления различной информации в удобном виде (графики, диаграммы, схемы и т.д.). Она реализует следующие возможности:

- работа с когнитивной картой (отображение факторов, связей между ними, а также редактирование факторов и связей);
- графическое представление результатов свободного и управляемого развития;
- визуализация Парето анализа;
- визуализация анализа графа влияющих факторов.

Подсистема анализа реализует методы статического и динамического анализа когнитивных карт. К статическим методам относится метод анализа влияющих факторов. К динамическим методам можно отнести следующие методы:

- анализ цепочек влияния;
- анализ Парето.

Анализ влияющих факторов позволяет получить список всех факторов влияющих на выбранный фактор, а также всех факторов, испытывающих влияние от выбранного фактора. Данная информация посредством системы визуализации раскрашивает граф по уровням влияния. Данное представление является удобным инструментом для анализа глубины связи между факторами.

Анализ цепочек влияния позволяет находить цепочки и динамически рассчитывать результирующие воздействие всей цепочки между двумя заданными факторами.

Подсистема моделирования реализует все необходимые методы для расчета в рамках когнитивного подхода. Предусмотрены следующие виды расчетов:

- расчет транзитивного замыкания матрицы и графа;

- нахождение и расчет цепочек влияния;
- расчет импульсных процессов для моделирования свободного и управляемого развития ситуации.

Подсистема работы со сценариями реализует возможности по работе со сценариями в модели. Эта подсистема позволяет выбрать целевые и управляющие факторы, а также создать группы наблюдения для каждого сценария. После настройки параметров сценария с помощью системы моделирования можно провести моделирование свободного и управляемого развития ситуации, а полученные результаты для каждого фактора сохранить для текущего сценария. Также данная подсистема представляет возможности по сравнению сценариев и выбор наиболее оптимальных по разным критериям – например, по наибольшей близости к целевым значениям.

Подсистема построение отчетов позволяет строить разнообразные отчеты по модели и сохранять их в различных форматах.

В отчет можно выводить следующую информацию:

- общая информация о модели (название, описание, сведения об авторстве);
- список факторов;
- детальная информация о факторах;
- информация о связях;
- когнитивная карта;
- информация о сценариях;
- результаты моделирования по каждому сценарию.

Подсистема импорта/экспорта реализует возможность импорта и экспорта матрицы взаимовлияния факторов в текстовой файл с разделителями. Получаемый файл имеет табличную структуру.

## 10.5. Пример. Модель потребления электроэнергии в регионе (модель Робертса)

Рассмотрим пример когнитивной карты для анализа проблемы потребления электроэнергии в регионе (рис.10.3).

Предположим, что в первом приближении исследуемую проблему достаточно полно можно описать семью факторами F, J, P, Q, R, C, U (наименования факторов указаны на рисунке). Дугами на рис.10.3 отмечены существенные причинно-следственные отношения, влиянием остальных можно пренебречь.

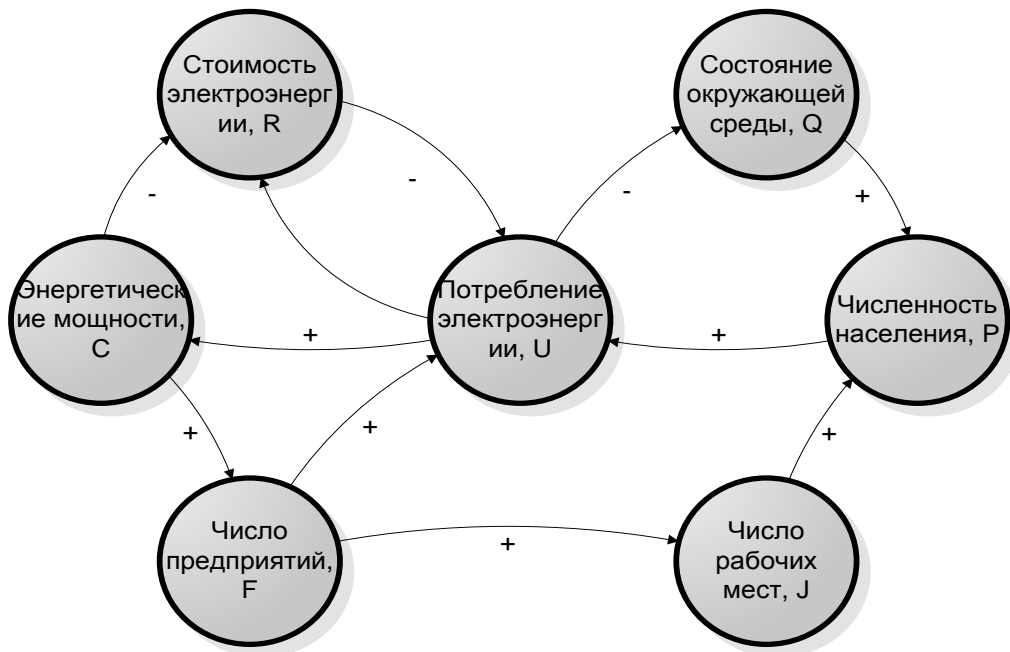


Рис.10.3. Модель потребления электроэнергии в регионе

Дуга (Q, P) имеет знак «+», так как улучшение окружающей среды ведет к увеличению числа жителей, а ухудшение состояния окружающей среды вызывает отток населения. Дуга (U, Q) имеет знак «-», так как увеличение потребления энергии ухудшает состояние окружающей среды, а уменьшение потребления энергии благоприятно сказывается на ее состоянии. Дуга (P, U) имеет знак «+» в виду того, что рост числа жите-

лей вызывает увеличение потребления энергии и, наоборот, уменьшение населения приводит к падению потребления энергии.

В таблице 10.2 представлена матрица взаимовлияний, заполненная экспертом.

Зададим начальные значения всех факторов и произведем моделирование саморазвития ситуации. Результаты саморазвития ситуации представлены в таблице 10.3 и графически отображены на рис.10.4.

Таблица 10.2. Матрица взаимодействия

	Число предприятий	Число рабочих мест	Численность населения	Состояние окружающей среды	Стоимость электроэнергии	Энергетические мощности	Потребление электроэнергии
Число предприятий	0	0,5	0	0	0	0	0,5
Число рабочих мест	0	0	0,3	0	0	0	0
Численность населения	0	0	0	0	0	0	0,2
Состояние окружающей среды	0	0	0,2	0	0	0	0
Стоимость электроэнергии	0	0	0	0	0	0	-0,4
Энергетические мощности	0,3	0	0	0	-0,5	0	0
Потребление электроэнергии	0	0	0	-0,5	0,2	0,2	0

Таблица 10.3. Саморазвитие ситуации

	Начальные значения	Саморазвитие
Число предприятий	0,25	0,30
Число рабочих мест	0,25	0,40
Численность населения	0,25	0,36
Состояние окружающей среды	0,10	-0,03
Стоимость электроэнергии	0,05	0,03
Энергетические мощности	0,10	0,15
Потребление электроэнергии	0,05	0,26

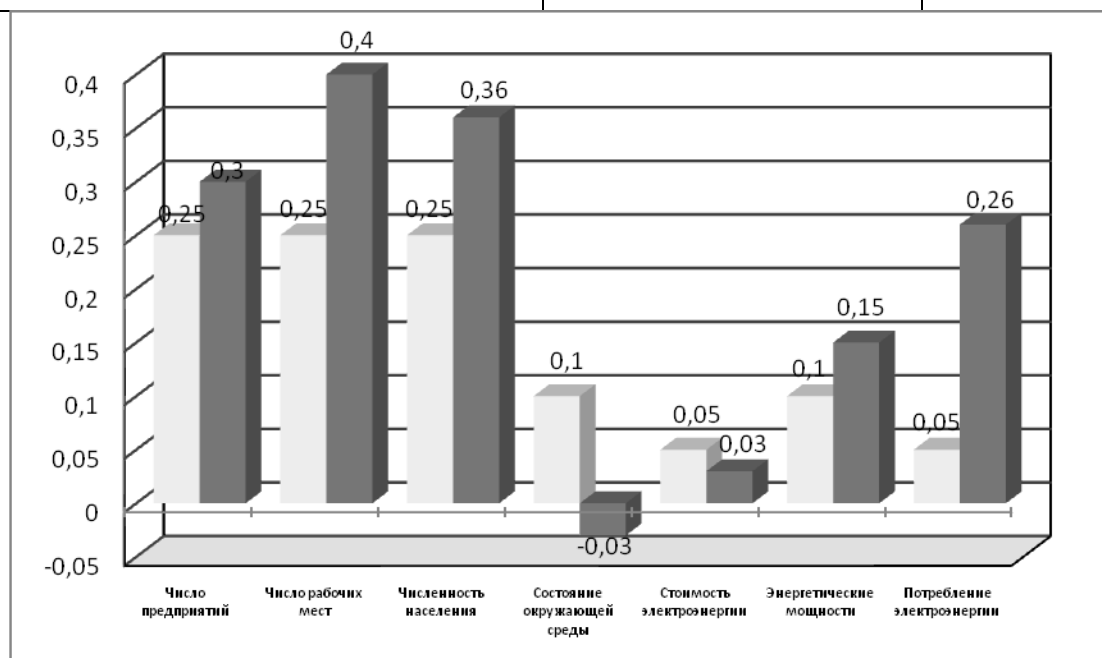


Рис.10.4. Саморазвитие ситуации

Анализ показывает, что при заданных начальных параметрах, несмотря на улучшение всех наблюдаемых факторов, достаточно сильно ухудшается состояние окружающей среды.

Для улучшения ситуации (состояние окружающей среды), повысим стоимость электроэнергии до величины 0,2 (сделаем этот фактор управ-

ляющим) и наблюдаем за управляемым развитием ситуации. Результаты сведены в таблицу 10.4 и представлены на рис.10.5

Таблица 10.4 Сравнение саморазвития и управляемого развития

	Начальные значения	Саморазвитие	Управляемое развитие
Число предприятий	0,25	0,30	0,29
Число рабочих мест	0,25	0,40	0,39
Численность населения	0,25	0,36	0,37
Состояние окружающей среды	0,10	-0,03	0,01
Стоимость электроэнергии	0,05	0,03	0,22
Энергетические мощности	0,10	0,15	0,13
Потребление электроэнергии	0,05	0,26	0,17

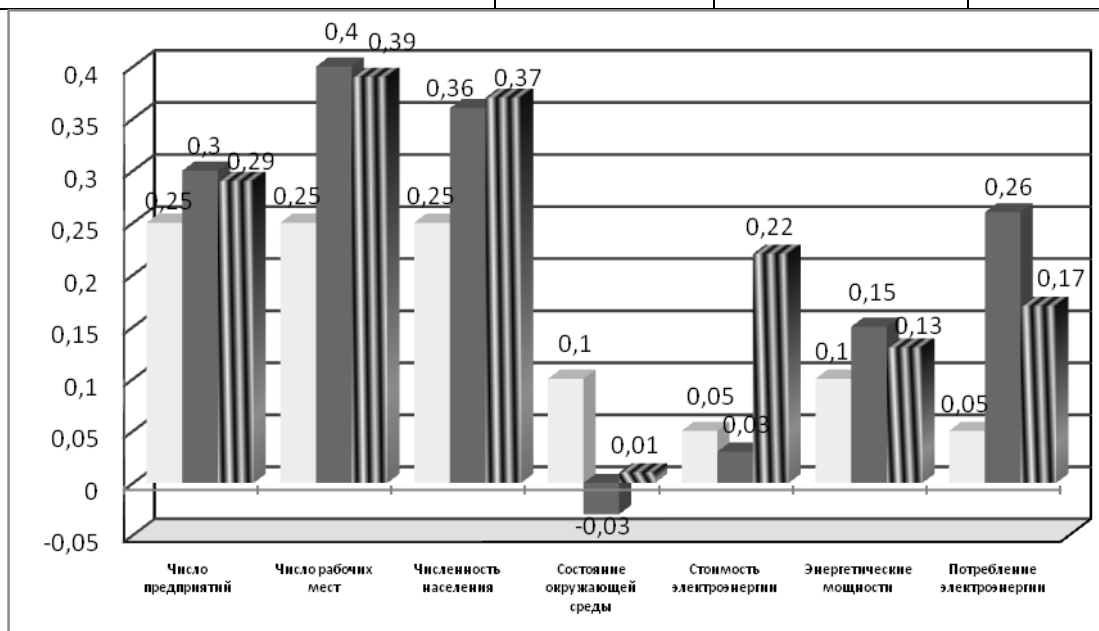


Рис.10.4. Управляемое развитие ситуации

Видно, что при увеличении стоимости электроэнергии, нам удалось улучшить состояние окружающей среды, незначительно увеличить численность населения, при этом снизив потребление электроэнергии в регионе, но потеряв некоторый процент энергетических мощностей.

Конечно, говорить о какой-то адекватности модели Робертса невозможно, но ее удобно применять для объяснения аппарата когнитивного моделирования.

### **10.6. Вопросы для самопроверки**

- 1) Как необходимо выбирать факторы, характеризующие модель?
- 2) Каким образом можно влиять на развитие ситуации в развитии модели?
- 3) Откуда берутся данные для заполнений начальных условий?
- 4) Как численно показать в матрице взаимовлияний «сильное усиливающее влияние» факторов друг на друга?
- 5) Каковы основные этапы построения когнитивной модели?
- 6) Каковы основные этапы когнитивного моделирования?

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. - СПб.: БХВ-Петербург, 2007.
2. Бусленко Н.П. Моделирование сложных систем. - М.: Наука, 1978.
3. Вентцель Е.С. Исследование операций. Задачи, принципы, методология. - М.: Высшая школа, 2001.
4. Вентцель Е.С. Исследование операций. - М.: Советское радио, 1972.
5. Дворецкий С.И., Муромцев Ю.А. и др. Моделирование систем. - М.: Издательский центр «Академия», 2009.
6. Емельянов В.В., Ясиновский С.И. Имитационное моделирование систем. - М.: МГТУ им. Н.Э. Баумана, 2009.
7. Карпов Ю. Имитационное моделирование систем. Введение в моделирование с Any Logic 5. - СПб.: БХВ-Петербург, 2005.
8. Костерин В.В. Оптимизация технических систем и устройств. Учебное пособие. - Волгоград: ВолгГТУ, 1996.
9. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. - СПб.: БХВ-Петербург, 2003.
10. Львовский Е.Н. Статистические методы построения эмпирических формул. - М.: Высшая школа, 1988.
11. Мышкис А.Д. Элементы теории математических моделей. - М.: КомКнига, 2007.
12. Неуймин Я.Г. Модели в науке и технике. - Ленинград: Наука, 1984.
13. Орлова, Ю.А. Основы системного анализа: учебное пособие (гриф). Доп. УМО вузов по университетскому политехническому образованию / Орлова Ю.А., Фоменков С.А., Заболеева-Зотова А.В.; ВолгГТУ. - 2-е изд., перераб. и доп. - Волгоград, 2013. - 232 с.
14. Пегат А. Нечеткое моделирование и управление. - М.: Бинوم. Лаборатория знаний, 2009.
15. Питерсон Дж. Теория сетей Петри и моделирование систем. - М.: Мир, 1984.

16. Плотинский Ю.М. Модели социальных процессов. Учебное пособие для вузов. - М.: Логос, 2001.
17. Рыжиков Ю.И. Имитационное моделирование. Теория и технология. - СПб.: КОРОНА принт; М. - Альтекс-А, 2004.
18. Советов Б.Я., Яковлев С.А. Моделирование систем. - М.: Высшая школа, 2007.
19. Таха Х. Введение в исследование операций. – М.: Издательский дом «Вильямс», 2005.

Учебное издание

Сергей Алексеевич **Фоменков**,

Валерий Анатольевич **Камаев**,

Юлия Александровна **Орлова**

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
СИСТЕМНЫХ ОБЪЕКТОВ**

Редактор

Компьютерная верстка

Темплан заказных изданий 2013 г. Поз. №

Подписано в печать .... Формат 60x84 1/16. Бумага офсетная.

Гарнитура Times. Печать офсетная Усл. Печ. Л. \_\_. Уч.изд. л. \_\_\_\_.

Тираж 500 экз. Заказ

Волгоградский государственный технический университет

400005, г. Волгоград, пр. им. В. И. Ленина, 28, корп.1

Отпечатано в типографии ИУНЛ ВолгГТУ.

400005, г. Волгоград, пр. им. В. И. Ленина, 28, корп.7