

# **УЧЕБНОЕ ПОСОБИЕ**

**по дисциплине**

## **Моделирование**

**Для студентов третьего курса**

**специальности I 40 02 01**

**«Вычислительные машины, системы и сети»**

**Минск 2007**

Автор – Мельник Николай Иосифович, старший преподаватель кафедры «Программное обеспечение информационных технологий» Белорусского государственного университета информатики и радиоэлектроники.

## СОДЕРЖАНИЕ

1. Моделирование. Основные понятия и принципы .....	4
2. Аналитическое моделирование .....	11
2.1. Математические модели. ....	11
2.2. Типовые схемы моделирования.....	12
2.3. Непрерывно-детерминированные модели ( <i>D</i> -схемы) .....	13
2.4. Дискретно-детерминированные модели ( <i>F</i> -схемы).....	14
2.5. Дискретно-стохастические модели ( <i>P</i> -схемы).....	18
2.6. Марковский случайный процесс .....	19
2.7 Непрерывно – стохастические модели ( <i>Q</i> – схемы) .....	26
2.7.1. Системы массового обслуживания. Потоки событий.....	26
2.7.2. Простейший поток.....	30
2.7.3. Непрерывные марковские цепи. Уравнения Колмогорова .....	33
2.7.4. Диаграмма интенсивностей переходов.....	37
2.7.5 Формула Литтла.....	39
2.7.6. Исследование СМО с помощью диаграмм интенсивностей переходов .....	41
2.7.7. Замкнутые системы массового обслуживания (СМО с ожиданием ответа).....	56
2.7.8. Распределение Эрланга. Метод этапов.....	60
2.7.8. Немарковские СМО.....	69
3. Имитационное моделирование.....	73
3.1. Условия применения имитационного моделирования .....	73
3.2. Этапы имитационного моделирования .....	76
3.3. Способы моделирования случайных величин .....	78
3.4. Равномерно-распределённые случайные числа (РРСЧ).....	80
3.4.1. Методы формирования РРСЧ. ....	83
3.4.2. Проверка качества последовательностей РРСЧ.....	85
3.5. Формирование случайных величин с заданным законом распределения.....	91
3.5.1. Метод обратной функции.....	91
3.5.2. Универсальный метод .....	93
3.5.3. Метод исключения (отбраковки, режекции, Дж. Неймана) .....	96
3.5.4. Метод композиции (суперпозиции).....	100
3.6. Формирование случайных векторов с заданными вероятностными характеристиками.....	102
3.7. Моделирование случайных событий.....	103
3. 8. Сетевые модели .....	106
3.8.1. Сети Петри .....	106
3.8.2. Е-сети.....	109
3.8.3. Сетевая модель взаимодействующих параллельных процессов в операционной системе. ....	118
3.9. Управление модельным временем .....	122
3.10. Планирование машинных экспериментов.....	128

3.11. Обработка экспериментальных данных .....	136
3.11.1. Экспериментальные оценки .....	137
3.11.2. Оценки для математического ожидания и дисперсии .....	138
3.11.2. Доверительные интервал и вероятность.....	139
3.11.3. Точность. Определение числа реализаций.....	142
Литература.....	145

## 1. МОДЕЛИРОВАНИЕ. ОСНОВНЫЕ ПОНЯТИЯ И ПРИНЦИПЫ

В настоящее время трудно назвать область человеческой деятельности, в которой в той или иной степени не использовались бы методы моделирования.

Формально мы будем понимать под моделированием замещение одного объекта (оригинала) другим (моделью) и фиксацию или изучение свойств оригинала путем исследования свойств модели.

Модель – представление объекта, системы или понятия (идеи) в некоторой форме, отличной от формы их реального существования.

Польза от моделирования может быть достигнута только при соблюдении следующих достаточно очевидных условий:

- модель адекватно отображает свойства оригинала, существенные с точки зрения цели исследования;
- модель позволяет устранять проблемы, присущие проведению измерений на реальных объектах.

При экспериментировании с моделью сложной системы можно получить больше информации о внутренних взаимодействующих факторах системы, чем при манипулировании с реальной системой благодаря изменяемости структурных элементов, легкости изменения параметров модели и т.д.

Исторически сложились два основных подхода при моделировании процессов и систем.

**Классический** (индуктивный) рассматривает систему путем перехода от частного к общему, т.е. модель системы синтезируется путем слияния моделей ее компонент, разрабатываемых отдельно.

При **системном** подходе предполагается последовательный переход от общего к частному, когда в основе построения модели лежит цель исследования. Именно из нее исходят, создавая модель. Подобие процесса, протекающего в модели, реальному процессу является не целью, а лишь условием правильного функционирования модели, поэтому в качестве цели должна быть поставлена задача изучения какой-либо стороны функционирования объекта. Для правильно построенной модели характерно то, что она выявляет только те закономерности, которые нужны исследователю и не рассматривает те свойства системы, которые не существенны для данного исследования.

Качество моделирования определяется тем, в какой степени решаются задачи, поставленные исследователем.

Рассмотрим основные принципы моделирования.

1) *Принцип информационной достаточности.* При полном отсутствии информации об исследуемом объекте построить его модель невозможно. Если информация полная, то моделирование лишено смысла. Должен существовать некоторый критический уровень априорных сведений об объекте (уровень информационной достаточности), при достижении которого может быть построена его адекватная модель.

2) *Принцип осуществимости.* Модель должна обеспечивать достижения поставленной цели с вероятностью отличной от нуля и за конечное время. Обычно задают некоторое пороговое значение и приемлемую границу времени  $t_0$  достижения цели. Модель осуществима, если

$$P(t) \geq P_0 \text{ и } t \leq t_0.$$

3) *Принцип множественности моделей.* Создаваемая модель должна отражать в первую очередь те свойства моделируемой системы или процесса, которые влияют на выбранный показатель эффективности. Соответственно, с помощью конкретной модели можно изучить лишь некоторые стороны реальности. Для более полного ее исследования необходим ряд моделей, позволяющих более разносторонне и с разной степенью детальности отражать рассматриваемый объект или процесс.

4) *Принцип агрегирования.* Сложную систему обычно можно представить состоящей из подсистем (агрегатов), для математического описания которых используются стандартные математические схемы. Кроме того, этот принцип позволяет гибко перестраивать модель в зависимости от целей исследования.

5) *Принцип параметризации.* В ряде случаев моделируемая система может иметь относительно изолированные подсистемы, которые характеризуются определенным параметром (в том числе векторным). Такие подсистемы можно заметить в модели соответствующими числами, а не описывать процесс их функционирования. При необходимости зависимость этих величин от ситуации может быть задана в виде таблицы, графика или аналитического выражения (формулы). Это позволяет сократить объем и продолжительность моделирования. Однако надо помнить, что параметризация снижает адекватность модели.

При проведении исследования различных объектов с использованием моделирования экспериментатор должен последовательно решить ряд возникающих перед ним проблем.

Если цель поставлена, то возникает проблема построения модели. Это возможно, если имеется информация или выдвинуты гипотезы относительно структуры, алгоритмов функционирования и параметров исследуемого объекта.

Если модель построена, то возникает проблема организации работы с ней. Основные задачи здесь - минимизация времени получения результатов и обеспечения их достоверности.

Наконец, если в результате эксперимента с моделью получены результаты, возникает задача их обработки и интерпретации. Средства вычислительной техники, используемые при моделировании могут помочь с точки зрения эффективности реализации сложной модели, но не являются гарантами правильности той или иной модели. Только на основе обработанных данных и опыта исследователя можно достоверно оценить адекватность модели по отношению к реальному процессу.

В основе моделирования лежит *теория подобия*, которая утверждает, что абсолютное подобие может иметь место лишь при замене одного объекта другим точно таким же. При моделировании абсолютное подобие не имеет места и создатели моделей стремятся к тому, чтобы модель достаточно хорошо отображала исследуемую сторону функционирования объекта. Поэтому в качестве одного из первых признаков классификации видов моделирования можно выбрать степень полноты модели и разделить модели в соответствии с этим признаком на полные, неполные и приближенные. В основе полного моделирования лежит полное подобие, которое проявляется как во времени, так и в пространстве. Для неполного моделирования характерно неполное подобие модели изучаемому объекту. В основе приближенного моделирования лежит приближенное подобие, при котором некоторые стороны функционирования реального объекта не моделируются совсем.

Методы моделирования и, соответственно, модели могут быть классифицированы с использованием различных критериев.

В зависимости от характера изучаемых процессов в системе все виды моделирования могут быть разделены на детерминированные и стохастические. *Детерминированное моделирование* отображает детерминированные процессы, т. е. процессы, в которых предполагается отсутствие всяких случайных воздействий; *стохастическое моделирование* отображает вероятностные процессы и события. В этом случае анализируется ряд реализаций случайного процесса и оцениваются средние характеристики, т. е. набор однородных реализаций.

По признаку развития процессов во времени различают статические и динамические. *Статическое моделирование* служит для описания поведения объекта в какой-либо момент времени, а *динамическое моделирование* отражает поведение объекта во времени.

По представлению информации в модели различают дискретные, непрерывные и дискретно-непрерывные модели. *Дискретное моделирование* служит для описания процессов, которые

предполагаются дискретными, соответственно непрерывное моделирование позволяет отразить непрерывные процессы в системах, а *дискретно-непрерывное моделирование* используется для случаев, когда хотят выделить наличие как дискретных, так и непрерывных процессов.

В зависимости от формы представления объекта можно выделить мысленное и реальное моделирование. *Мысленное моделирование* часто является единственным способом моделирования объектов, которые либо практически не реализуемы в заданном интервале времени, либо существуют вне условий, возможных для их физического создания. Например, на базе мысленного моделирования могут быть проанализированы многие ситуации микромира, которые не поддаются физическому эксперименту. Мысленное моделирование может быть реализовано в виде наглядного, символического и математического.

При *наглядном моделировании* на базе представлений человека о реальных объектах создаются различные наглядные модели, отображающие явления и процессы, протекающие в объекте. В основу *гипотетического моделирования* исследователем закладывается некоторая гипотеза о закономерностях протекания процесса в реальном объекте, которая отражает уровень знаний исследователя об объекте и базируется на причинно-следственных связях между входом и выходом изучаемого объекта. Гипотетическое моделирование используется, когда знаний об объекте недостаточно для построения формальных моделей.

*Аналоговое моделирование* основывается на применении аналогий различных уровней. Наивысшим уровнем является полная аналогия, имеющая место только для достаточно простых объектов. С усложнением объекта используют аналогии последующих уровней, когда аналоговая модель отображает несколько либо только одну сторону функционирования объекта.

Существенное место при мысленном наглядном моделировании занимает *макетирование*. Мысленный макет может применяться в случаях, когда протекающие в реальном объекте процессы не поддаются физическому моделированию, либо может предшествовать проведению других видов моделирования. В основе построения мысленных макетов также лежат аналогии, однако обычно базирующиеся на причинно-следственных связях между явлениями и процессами в объекте. Если ввести условное обозначение отдельных понятий, т. е. знаки, а также определенные операции между этими знаками, то можно реализовать *знаковое моделирование* и с помощью знаков отображать набор понятий – составлять отдельные цепочки из слов и предложений. Используя операции объединения, пересечения и дополнения теории множеств, можно дать описание какого-то реального объекта.

В основе *языкового моделирования* лежит некоторый тезаурус. Последний образуется из набора входящих понятий, причем этот набор должен быть фиксированным. Следует отметить, что между тезаурусом и обычным словарем имеются принципиальные различия. Тезаурус – словарь, который очищен от неоднозначности, т. е. в нем каждому слову может соответствовать лишь единственное понятие, хотя в обычном словаре одному слову могут соответствовать несколько понятий.

*Символическое моделирование* представляет собой искусственный процесс создания логического объекта, который замещает реальный и выражает основные свойства его отношений с помощью определенной системы знаков или символов.

Для исследования характеристик процесса функционирования любой системы математическими методами, включая в машинные, должна быть проведена формализация этого процесса, т. е. построена математическая модель.

Под *математическим моделированием* будем понимать процесс установления соответствия данному реальному объекту некоторого математического объекта, называемого математической моделью, и исследование этой модели, позволяющее получать характеристики рассматриваемого реального объекта. Вид математической модели зависит как от природы реального объекта, так и от задач исследования объекта и требуемой достоверности и точности решения этой задачи. Любая математическая модель, как и всякая другая, описывает реальный объект лишь с некоторой степенью приближения к действительности. Математическое моделирование для исследования характеристик процесса функционирования систем можно разделить на аналитическое, имитационное и комбинированное.

Для *аналитического моделирования* характерно то, что процессы функционирования элементов системы записываются в виде некоторых функциональных соотношений (алгебраических, интегро-дифференциальных, конечно-разностных и т. п.) или логических условий. Аналитическая модель может быть исследована следующими методами:

а) аналитическим, когда стремятся получить в общем виде явные зависимости для искомых характеристик;

б) численным, когда, не умея решать уравнений в общем виде, стремятся получить числовые результаты при конкретных начальных данных;

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

Наиболее полное исследование процесса функционирования системы можно провести, если известны явные зависимости, связывающие искомые характеристики с начальными условиями, па-



раметрами и переменными системы S. Однако такие зависимости удается получить только для сравнительно простых систем. При усложнении систем исследование их аналитическим методом наталкивается на значительные трудности, которые часто бывают непреодолимыми. Поэтому, желая использовать аналитический метод, в этом случае идут на существенное упрощение первоначальной модели, чтобы иметь возможность изучить хотя бы общие свойства системы. Такое исследование на упрощенной модели аналитическим методом помогает получить ориентировочные результаты для определения более точных оценок другими методами. Численный метод позволяет исследовать по сравнению с аналитическим методом более широкий класс систем, но при этом полученные решения носят частный характер. Численный метод особенно эффективен при использовании ЭВМ.

В отдельных случаях исследователя системы могут удовлетворить и те выводы, которые можно сделать при использовании качественного метода анализа математической модели. Такие качественные методы широко используются, например, в теории автоматического управления для оценки эффективности различных вариантов систем управления.

В настоящее время распространены методы машинной реализации исследования характеристик процесса функционирования больших систем. Для реализации математической модели на ЭВМ необходимо построить соответствующий моделирующий алгоритм.

При *имитационном моделировании* реализующий модель алгоритм воспроизводит процесс функционирования системы во времени. Причем имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры и последовательности протекания во времени, что позволяет по исходным данным получить сведения о состояниях процесса в определенные моменты времени, дающие возможность оценить характеристики моделируемой системы.

Основным преимуществом имитационного моделирования по сравнению с аналитическим является возможность решения более сложных задач. Имитационные модели позволяют достаточно просто учитывать такие факторы, как наличие дискретных и непрерывных элементов, нелинейные характеристики элементов системы, многочисленные случайные воздействия и др., которые часто создают трудности при аналитических исследованиях. В настоящее время имитационное моделирование – наиболее эффективный метод исследования больших систем, а часто и единственный практически доступный метод получения информации о поведении системы, особенно на этапе ее проектирования.

Когда результаты, полученные при воспроизведении на имитационной модели процесса функционирования моделируемой

системы, являются реализациями случайных величин и функций, тогда для нахождения характеристик процесса требуется его многократное воспроизведение с последующей статистической обработкой информации и целесообразно в качестве метода машинной реализации имитационной модели использовать метод статистического моделирования. Первоначально был разработан метод статистических испытаний, представляющий собой численный метод, который применялся для моделирования случайных величин и функций, вероятностные характеристики которых совпадали с решениями аналитических задач (такая процедура получила название метода Монте-Карло). Затем этот прием стали применять и для машинной имитации с целью исследования характеристик процессов функционирования систем, подверженных случайным воздействиям, т. е. появился метод статистического моделирования.

Метод имитационного моделирования позволяет решать задачи анализа больших систем, включая задачи оценки: вариантов структуры системы, эффективности различных алгоритмов управления системой, влияния изменения различных параметров системы. Имитационное моделирование может быть положено также в основу структурного, алгоритмического и параметрического синтеза больших систем, когда требуется создать систему с заданными характеристиками при определенных ограничениях, которая является оптимальной по некоторым критериям оценки эффективности.

*Комбинированное* (аналитико-имитационное) *моделирование* при анализе и синтезе систем позволяет объединить достоинства аналитического и имитационного моделирования. При построении комбинированных моделей проводится предварительная декомпозиция процесса функционирования объекта на составляющие подпроцессы, и для тех из них, где это возможно, используются аналитические модели, а для остальных подпроцессов строятся имитационные модели. Такой комбинированный подход позволяет охватить качественно новые классы систем, которые не могут быть исследованы с использованием только аналитического и имитационного моделирования в отдельности.

При *реальном моделировании* используется возможность исследования различных характеристик либо на реальном объекте целиком, либо на его части. Такие исследования могут проводиться как на объектах, работающих в нормальных режимах, так и при организации специальных режимов для оценки интересующих исследователя характеристик (при других значениях переменных и параметров, в другом масштабе времени и т. д.). Реальное моделирование является наиболее адекватным, но при этом его возможности с учетом особенностей реальных объектов ограничены. Например, проведение реального моделирования АСУ предприятием потребует, во-первых, создания такой АСУ, а во-вторых, проведения

экспериментов с управляемым объектом, т. е. предприятием, что в большинстве случаев невозможно. Рассмотрим разновидности реального моделирования.

*Натурным моделированием* называют проведение исследования на реальном объекте с последующей обработкой результатов эксперимента на основе теории подобия. При функционировании объекта в соответствии с поставленной целью удастся выявить закономерности протекания реального процесса. Надо отметить, что такие разновидности натурального эксперимента, как *производственный эксперимент и комплексные испытания*, обладают высокой степенью достоверности. При *физическом моделировании* эксперимент проводится на специальных установках, процессы в которых имеют физическое подобие процессам в моделируемых объектах.

## 2. АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

### 2.1. Математические модели.

Исходной информацией для математических моделей служат данные о назначении и условиях работы моделируемой системы.

При построении математических моделей исследователь должен обеспечить с одной стороны адекватность отображения в модели реальных процессов, протекающих в исследуемой системе, а с другой стороны – возможность реализации моделирования, достигаемую, как правило, за счет выделения основных свойств системы и игнорирования несущественных (с точки зрения конкретной исследовательской задачи).

В общем случае при построении математических моделей, описание объекта должно учитывать следующие множества факторов:

- входные воздействия  $x_i \in X, i = \overline{1, n_x}$ ;
- воздействия внешней среды  $v_e \in V, e = \overline{1, n_v}$ ;
- внутренние параметры  $h_k \in H, k = \overline{1, n_h}$ ;
- выходные характеристики  $y_j \in Y, j = \overline{1, n_y}$

При моделировании множества  $X, V, H$  являются независимыми (экзогенными) переменными, а  $Y$  зависимым (эндогенным).

Функционирование системы во времени в общем случае описывается оператором:

$$\vec{y}(t) = F_s(\vec{x}, \vec{v}, \vec{h}, t). \quad [2.1.1]$$

Для статистических моделей математическая модель может быть записана как

$$\vec{y} = F_s(\vec{x}, \vec{v}, \vec{h}). \quad [2.1.2]$$

Соотношения [2.1.1] и [2.1.2] могут быть заданы разными способами (аналитически, таблично, графически и т.д.). В ряде случаев они могут быть получены через свойства системы в конкретные моменты времени, так называемые *состояния*

$$z_i \in Z.$$

Процесс функционирования моделируемой системы во времени  $t_0 < t \leq T$  в этом случае можно представить как последовательную смену состояний, которые полностью определяются начальным состоянием  $\vec{z}_0$ , входными воздействиями, внутренними параметрами и воздействиями внешней среды

$$\vec{z}(t) = \Phi(\vec{z}_0, \vec{x}, \vec{v}, \vec{h}, t); \quad [2.1.3]$$

$$\vec{y}(t) = F_s(\vec{z}, t). \quad [2.1.4]$$

Если элементы случайности отсутствуют или не учитываются, то имеем детерминированную модель

$$\vec{y}(t) = f(\vec{x}, t).$$

Это самые общие математические соотношения, а на практике на первоначальных этапах используют т.н. типовые математические схемы (типовые схемы моделирования). Они более просты и наглядны.

## 2.2. Типовые схемы моделирования

Не обладая такой степенью общности, как рассмотренные модели, типовые математические схемы имеют преимущества простоты и наглядности, но при существенном сужении возможностей применения. Различают следующие виды типовых схем моделирования:

D-схемы, или непрерывно-детерминированные модели;

F-схемы, или дискретно-детерминированные модели (конечные автоматы);

P-схемы, или дискретно-стохастические модели (вероятностные автоматы);

Q-схемы, или непрерывно-стохастические модели (системы массового обслуживания);

A-схемы - обобщенные модели (агрегативные системы).

Перечисленные типовые математические схемы, естественно, не могут претендовать на возможность описания на их базе всех возможных

процессов и систем. Каждая из них имеет свою область использования. В качестве детерминированных моделей, когда при исследовании случайные факторы не учитываются, для представления систем, функционирующих в непрерывном времени, используются дифференциальные, интегральные, интегродифференциальные и другие уравнения, а для представления систем, функционирующих в дискретном времени, – конечные автоматы и конечноразностные схемы. В качестве стохастических моделей (при учете случайных факторов) для представления систем с дискретным временем используются вероятностные автоматы, а для представления системы с непрерывным временем – системы массового обслуживания и т. д.

### 2.3. Непрерывно-детерминированные модели (D-схемы)

В качестве математических моделей в D-схемах (от английского *dynamic*) используются дифференциальные уравнения. Дифференциальными уравнениями называются такие уравнения, в которых неизвестными будут функции одной или нескольких переменных, причем в уравнение входят не только функции, но и их производные различных порядков. Если неизвестные – функции многих переменных, то уравнения называются уравнениями в частных производных, в противном случае при рассмотрении функций только одной независимой переменной уравнения называются обыкновенными дифференциальными уравнениями.

Обычно в таких математических моделях в качестве независимой переменной, от которой зависят неизвестные искомые функции, служит время  $t$ . Тогда математическое соотношение для детерминированных систем в общем виде будет

$$\begin{cases} \vec{y}' = \vec{f}(\vec{y}, t), \\ \vec{y}(t_0) = \vec{y}_0, \end{cases}$$

где  $\vec{y} = (y_1, y_2, \dots, y_n)$   $n$ -мерный вектор,

$\vec{f} = (f_1, f_2, \dots, f_n)$   $n$ -мерная функция.

Наиболее широко этот подход используется в теории систем автоматического управления (САУ) при анализе точности и устойчивости САУ в переходных процессах.

Задачей системы является изменение выходной переменной (выходного сигнала)  $y(t)$  согласно заданному закону с определенной точ-

ностью (с допустимой ошибкой). При проектировании и эксплуатации систем автоматического управления необходимо выбрать такие параметры системы, которые обеспечили бы требуемую точность управления, а также устойчивость системы в переходном процессе.

Если система устойчива, то представляет практический интерес поведение системы во времени, максимальное отклонение регулируемой переменной  $y(t)$  в переходном процессе, время переходного процесса и т. п. Выводы о свойствах систем автоматического управления различных классов можно сделать по виду дифференциальных уравнений, приближенно описывающих процессы в системах. Порядок дифференциального уравнения и значения его коэффициентов полностью определяются статическими и динамическими параметрами системы

## 2.4. Дискретно-детерминированные модели ( $F$ -схемы)

При использовании такого рода моделей система представляется в виде автомата, перерабатывающего дискретную информацию и меняющего свои внутренние состояния лишь в допустимые моменты времени. Автомат можно представить как некоторое устройство (черный ящик), на которое подаются входные сигналы и снимаются выходные и которое может иметь некоторые внутренние состояния.

Конечным автоматом называется автомат, у которого множество внутренних состояний и входных сигналов (а следовательно, и множество выходных сигналов) являются конечными множествами.

Абстрактно конечный автомат (англ. *finite automata*) можно представить как математическую схему ( $F$ -схему), характеризующуюся шестью элементами: конечным множеством  $X$  входных сигналов (входным алфавитом); конечным множеством  $Y$  выходных сигналов (выходным алфавитом); конечным множеством  $Z$  внутренних состояний (внутренним алфавитом или алфавитом состояний); начальным состоянием  $z_0$ , функцией переходов  $\Phi(z, x)$ , функцией выходов  $\Psi(z, x)$ . Автомат, задаваемый  $F$ -схемой, функционирует в дискретном автоматном времени, моментами которого являются такты, т. е. примыкающие друг к другу равные интервалы времени. Изменение состояния автомата и выходного сигнала может произойти только в тактовые моменты. Таким образом, работа конечного автомата происходит по следующей схеме: в каждом  $t$ -м такте на вход автомата, находящегося в состоянии  $z(t)$ , подается некоторый сигнал  $x(t)$ , на который он реагирует переходом в  $(t+1)$ -м такте в новое состояние  $z(t+1)$  и выдачей некоторого выходного сигнала. Это можно описать следующими уравнениями:

$$z(t+1) = \Phi[z(t), x(t)], t=0, 1, 2 \dots;$$

$$y(t) = \Psi[z(t), x(t)], t=0, 1, 2 \dots$$

Различают два вида конечных автоматов. Приведенные выше уравнения описывают работу автомата первого рода, называемого также автоматом Мили. Для F-автомата второго рода

$$z(t+1) = \Phi[z(t), x(t)], t=0, 1, 2 \dots;$$

$$y(t) = \Psi[z(t), x(t-1)], t=1, 2, 3 \dots$$

Автомат второго рода, для которого функция выходов не зависит от входной переменной  $x(t)$  называется автоматом Мура. Для него

$$z(t+1) = \Phi[z(t), x(t)], t=0, 1, 2 \dots;$$

$$y(t) = \Psi[z(t)], t=1, 2, 3 \dots$$

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 схемы) обладают лишь одним состоянием. При этом работа комбинационной схемы заключается в том, что она ставит в соответствие каждому входному сигналу  $x(t)$  определенный выходной сигнал  $y(t)$ .

Чтобы задать конечный F-автомат, необходимо описать все элементы множества  $F = (Z, X, Y, \Phi, \Psi)$ , т. е. входной, внутренний и выходной алфавиты, а также функции переходов и выходов. Причем среди множества состояний необходимо выделить состояние  $z_0$ , в котором автомат находился в момент времени  $t = 0$ . Существуют несколько способов задания работы F-автоматов, 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. При этом обычно первый слева столбец соответствует начальному состоянию  $z_0$ . На пересечении  $i$ -й строки и  $k$ -го столбца таблицы переходов помещается соответствующее значение  $\Phi(z_k, x_i)$  функции переходов, а в таблице выходов — соответствующее значение  $\Psi(z_k, x_i)$  функции выходов. Для F-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 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  $z_j$ , отмечается сигналом  $x_k$ . Для того чтобы задать функцию выходов, дуги графа необходимо отметить соответствующими выходными сигналами  $\Psi(z_i, x_k)$ . Для автомата Мура выходные сигналы связаны только с состояниями и поэтому значениями  $\Psi(z_k)$  отмечают соответствующие вершины графа.

На рис. 2.1 и 2.2 приведены примеры задания автоматов Мили и Мура.

а)                                      Функция переходов  $z_j=\Phi(z_i, x_k)$

Входной сигнал	Состояния		
	$z_0$	$z_1$	$z_2$
$x_1$	$z_0$	$z_0$	$z_1$
$x_2$	$z_2$	$z_2$	$z_2$

Функция выходов  $y_j=\Psi(z_i, x_k)$

Входной сигнал	Состояния		
	$z_0$	$z_1$	$z_2$
$x_1$	$y_0$	$y_0$	$y_1$
$x_2$	$y_1$	$y_0$	$y_1$

б)

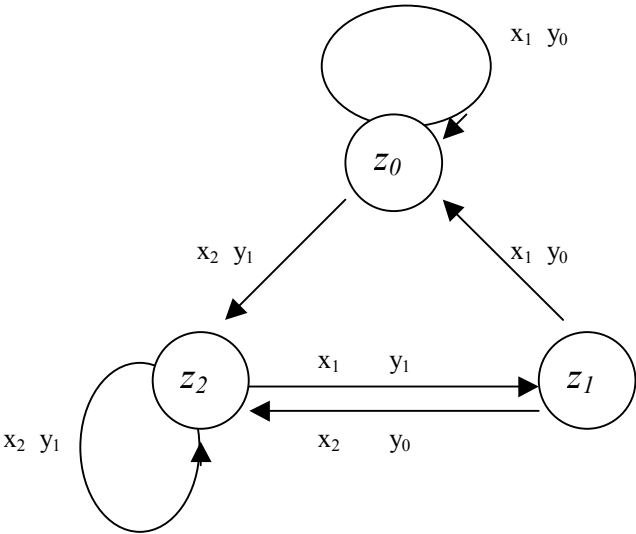


Рис. 2.1. Задание конечного автомата Мили

- а) табличным способом;
- б) с помощью графа.



а)                      Функция переходов  $z_j = \Phi(z_i, x_k)$

Входной сигнал	Состояния		
	$z_0$	$z_1$	$z_2$
$x_1$	$z_0$	$z_0$	$z_1$
$x_2$	$z_2$	$z_2$	$z_2$

Функция выходов  $y_j = \Psi(z_i)$

Состояния		
$z_0$	$z_1$	$z_2$
$y_0$	$y_0$	$y_1$

б)

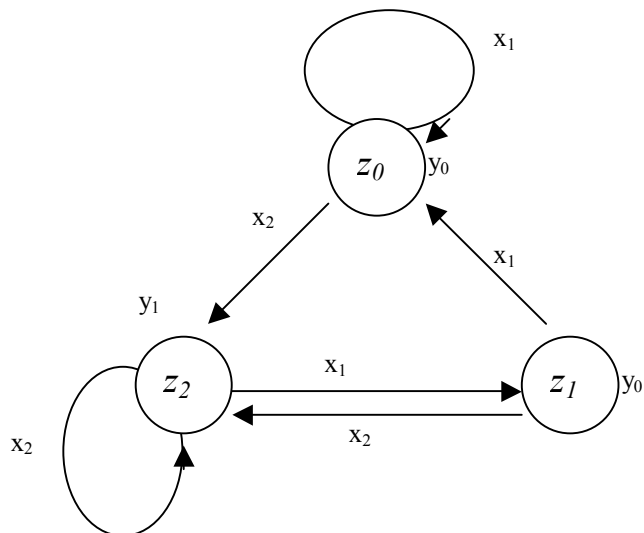


Рис. 2.2. Задание конечного автомата Мура

а) табличным способом;

б) с помощью графа.

Более подробно вопросы, связанные с построением и исследованием конечных автоматов, рассматриваются в курсе «Организация и функционирование ЭВМ».

## 2.5. Дискретно-стохастические модели (Р-схемы)

В общем виде *вероятностный автомат* (англ. probabilistic automat) можно определить как дискретный потактный преобразователь информации с памятью, функционирование которого в каждом такте зависит только от состояния памяти в нем и может быть описано статистически.

Сущность дискретизации времени при этом подходе остается аналогичной рассмотренным ранее конечным автоматам, но при этом добавляется влияние фактора стохастичности. Применение схем вероятностных автоматов (Р-схем) имеет важное значение для разработки методов проектирования дискретных систем, проявляющих статистически закономерное случайное поведение, для выяснения алгоритмических возможностей таких систем и обоснования границ целесообразности их использования, а также для решения задач синтеза по выбранному критерию дискретных стохастических систем, удовлетворяющих заданным ограничениям.

Для того, чтобы задать вероятностный автомат надо, как и для конечного автомата определить множество  $X = (x_1, x_2, \dots, x_n)$  входных сигналов, множество  $Y = (y_1, y_2, \dots, y_r)$  выходных сигналов и множество  $Z = (z_1, z_2, \dots, z_s)$  внутренних состояний. Описание процесса функционирования автомата осуществляется путем задания ряда распределений вероятностей.

Рассмотрим множество  $G$  пар  $(x_i z_j)$  и множество  $D$  пар  $(z_k y_h)$ . Для задания вероятностного автомата надо определить для каждой пары из множества  $G$  вероятности  $b_{kh}$  перехода автомата в состояние  $z_k$  и появления на выходе сигнала  $y_h$ :

Элементы из $D$	$(z_1 y_1)$	$(z_2 y_2)$	...	$(z_s y_{r-1})$	$(z_s y_r)$
$x_i z_j$	$b_{11}$	$b_{12}$	...	$b_{s\ r-1}$	$b_{sr}$

$$\text{При этом } \sum_{k=1}^s \sum_{h=1}^r b_{kh} = 1.$$

Число таких распределений, представленных в виде таблиц, равно числу элементов множества  $G$ . Обозначим множество таких распределений как  $B$ . Тогда совокупность множеств  $Z, X, Y, B$  определяет вероятностный автомат.

Это наиболее общий случай. Если распределения для нового состояния  $P$ -автомата и его выходного сигнала независимы, то это вероятностный автомат Мили. Для его задания надо определить множества распределений

элементы из $Y$	$y_1$	$y_2$	$\dots$	$y_{r-1}$	$y_r$
$x_i z_j$	$q_1$	$q_2$	$\dots$	$q_{r-1}$	$q_r$
элементы из $Z$	$z_1$	$z_2$	$\dots$	$z_{s-1}$	$z_s$
$x_i z_j$	$p_1$	$p_2$	$\dots$	$p_{s-1}$	$p_s$

Число таких распределений равно числу элементов множества  $G$ .

Говорят, что задан вероятностный автомат Мили, если заданы два распределения вероятностей

элементы из $Y$	$y_1 y_2 y_3 \dots y_Y$
$(x_i z_\xi)$	$q_1 q_2 q_3 \dots q_Y$

элементы из $Z$	$z_1 z_2 z_3 \dots z_k$
$(x_i z_\xi)$	$p_1 p_2 p_3 \dots p_k$

Здесь  $\sum_{k=1}^Y q_k = 1$ ,  $\sum_{k=1}^k p_k = 1$  и при этом распределения для  $q$  и  $p$

независимы.

Вероятностный автомат Мура имеет место, если определение выходного сигнала  $P$ -автомата зависит лишь от того состояния, в котором находится автомат в данном такте работы, а элементы из  $Z$  определяются как у автомата Мили.

Детерминированные автоматы это частный случай  $P$ -автомата. Также частными случаями являются  $Y$  – детерминированный и  $Z$  – детерминированные автоматы. Если выходной сигнал  $P$ -автомата определяется детерминированно, то такой автомат называется  $Y$ -детерминированным вероятностным автоматом. Аналогично,  $Z$ -детерминированным вероятностным автоматом называется  $P$ -автомат, у которого выбор нового состояния является детерминированным.

## 2.6. Марковский случайный процесс

Пусть имеется некоторая система, состояние которой может меняться с течением времени. Если состояние меняется случайным, непредсказуемым образом, будем говорить, что в системе протекает случайный процесс.

Случайный процесс называется процессом с дискретными состояниями, если возможные состояния системы:

$$S_1, S_2, S_3, \dots$$

можно перечислить (пронумеровать) одно за другим, а сам процесс состоит в том, что время от времени система скачком (мгновенно) переходит из одного состояния в другое.

Случайный процесс называется процессом с дискретным временем, если переходы из состояния в состояние возможны только в строго определенные моменты времени. В промежутках между этими моментами система сохраняет свое состояние.

Случайный процесс называется процессом с непрерывным временем, если переход системы из состояния в состояние возможен в любой наперед неизвестный, случайный момент.

Случайный процесс, протекающий в системе, называется *марковским процессом* (цепь Маркова) или процессом без последствия, если для каждого момента времени  $t_i$  вероятность любого последующего состояния системы зависит только от текущего состояния и не зависит от того, когда и каким путем система пришла в это состояние (т.е. от того, как развивался процесс в прошлом).

Иными словами, воздействие всей предыстории процесса на его будущее полностью сосредоточено в текущем значении процесса. Отсюда следует, что цепь Маркова должна обладать свойством отсутствия последствия. Это означает, что вероятность перехода в следующее состояние не должна зависеть от того, сколько времени процесс пребывал в текущем состоянии.

При моделировании систем, в которых случайные события, приводящие к изменению состояний, могут происходить только в моменты времени, отстоящие друг от друга на величину, кратную значению тактового интервала  $T$  (*дискретно – стохастические модели*), для описания интервалов времени между событиями используют регулярный просеянный поток. Его можно получить, удаляя в регулярном потоке события с вероятностью  $q$  и оставляя с вероятностью  $1-q$ . Просеянный поток иногда называют дискретным пуассоновским, так как его свойства аналогичны для моментов времени, кратных периоду  $T$ , свойствам простейшего потока. К просеянному регулярному потоку приводит, например регулярный поток данных, передаваемый по каналам связи с контролем наличия сбоев в передаваемом коде и исправлением путем повторной передачи.

Вероятность того, что величина интервала между событиями в просеянном потоке окажется равным  $i$  тактам:

$$P_i = q^{i-1} (1 - q).$$

Это выражение соответствует геометрическому распределению.

Математическое ожидание и среднее квадратическое отклонение для интервала времени между событиями в таком потоке равны соответственно:

$$m = \frac{T}{1 - q},$$

$$\sigma = m \sqrt{q} = \frac{T \sqrt{q}}{1 - q}.$$

Можно доказать, что, если число состояний системы конечно и из каждого состояния можно перейти (за то или иное число шагов) в любое другое, то существуют предельные (финальные) вероятности состояний, которые не зависят от времени и начального состояния системы. Сумма предельных вероятностей всех состояний системы равна единице:

$$\sum_{i=1}^n P_i = 1.$$

Это т.н. нормировочное уравнение.

Таким образом, при  $t \rightarrow \infty$  в системе устанавливается некоторый предельный *стационарный режим*, который состоит в том, что система случайным образом меняет свои состояния, но вероятность каждого из них уже не зависит от времени. Эта вероятность представляет собой не что иное, как относительное время пребывания системы в данном состоянии.

При моделировании систем процесс их функционирования удобно представлять в виде графа, вершинами которого являются состояния  $S_i$ , а направленные дуги описывают переходы между состояниями. Если процесс является марковским и известны вероятности переходов из состояния в состояние, то вероятности состояний  $P_i$  могут быть найдены исходя из того, что вероятность любого состояния  $S_i$  равна сумме произведений вероятностей состояний  $S_j$ , из которых есть переход в данное состояние на вероятности этих переходов  $p_{ji}$ , т.е

$$P_i = \sum_j P_j p_{ji}.$$

Рассмотрим на примере эту процедуру (рис 2.3).

Система может находиться в одном из трех состояний:  $S_1$ ,  $S_2$  или  $S_3$ .

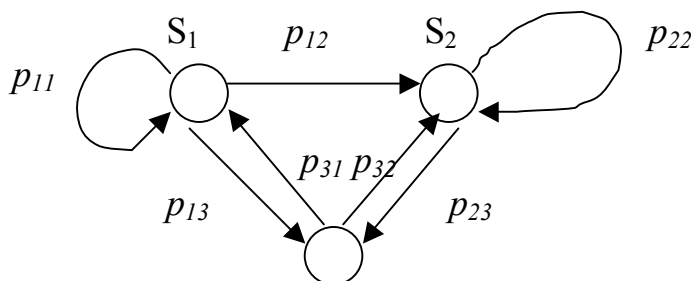


Рис. 2.3. Пример графа для случайного процесса

Запишем систему уравнений для определения вероятностей состояний:

$$\begin{aligned} P_1 &= p_{11}P_1 + p_{31}P_3; \\ P_2 &= p_{12}P_1 + p_{32}P_3 + p_{22}P_2; \\ P_3 &= p_{13}P_1 + p_{23}P_2. \end{aligned}$$

Попытка решить эту систему непосредственно неизбежно приведет к тождеству. Но, тем не менее, решение существует и может быть получено, если воспользоваться нормировочным уравнением

$$P_1 + P_2 + P_3 = 1.$$

Если подставить его вместо одной из строк системы, то можно будет получить значение вероятностей состояний.

Допустим, вероятности переходов из состояния в состояние имеют следующие значения:

$$p_{11}=0,25, \quad p_{12}=0,5, \quad p_{13}=0,25, \quad p_{22}=0,5, \quad p_{23}=0,5, \quad p_{31}=0,5, \quad p_{32}=0,5.$$

Тогда решение системы уравнений дает следующие результаты:

$$P_1 = 0,2, \quad P_2 = 0,5, \quad P_3 = 0,3.$$

Теперь рассмотрим пример исследования вычислительного узла (рис. 2.4) с использованием аппарата Марковской дискретной цепи.

Из памяти макрокоманд (МК) по синхросигналу через каждые два такта в арифметико-логическое устройство (АЛУ) считываются макрокоманды. Если АЛУ готово принять макрокоманду, она обрабатывается, иначе макрокоманда загружается в регистровую память (РП) объемом  $n$  ячеек (мест ожидания) и ожидает освобождения АЛУ. При полном заполнении РП макрокоманда блокируется в памяти макрокоманд и процесс дальнейшей выборки приостанавливается. Интервалы времени обработки макрокоманд в АЛУ случайны и имеют геометрическое распределение с параметром  $\pi$ , т.е. с вероятностью  $(1 - \pi)$  обработка макрокоманды в АЛУ по окончании очередного тактового интервала завершится, а с вероятностью  $\pi$  продлится еще на один интервал.

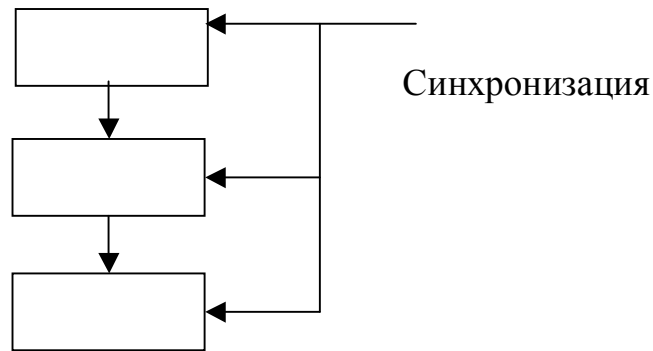


Рис. 2.4. Система АЛУ – память

При построении и исследовании модели будем пользоваться представлением данного устройства как системы массового обслуживания (СМО). Структура этой СМО изображена на рисунке 2.5.

Построим математическую модель и исследуем ее. Т.к. поток обслуживаний представляет собой просеянный регулярный поток, то процесс будет марковским, и мы можем определить финальные вероятности состояний этой системы.

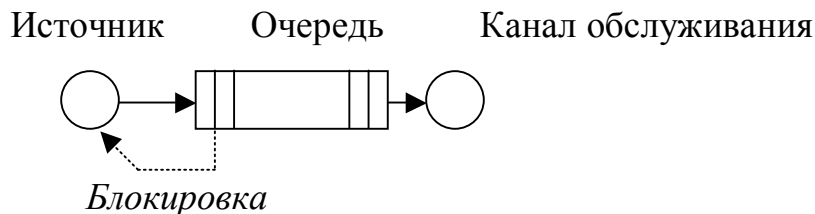


Рис. 2.5. Представление системы “АЛУ – память” как СМО

Будем определять состояние системы трехкомпонентным вектором,  $j t_1 t_2$ .

Комбинаторная составляющая этого вектора  $j$  - количество заявок, находящихся в накопителе (длина очереди),  $j = 0, 1, 2, \dots, n$ .

Временная составляющая  $t_1$  - число тактов, оставшихся до появления заявки на выходе источника ( $t_1 = 0, 1, 2$ ). Значение 0 означает, что источник заблокирован.

Составляющая  $t_2$  определяет состояние канала обслуживания (АЛУ) и может принимать два значения:

$t_2 = 0$  - канал свободен;

$t_2 = 1$  - канал занят обслуживанием заявки.

Построим граф (рис. 2.6) и систему уравнений для стационарных (финальных) вероятностей состояний  $P_{j t_1 t_2}$ .

В состоянии  $P_{020}$  система больше не вернется, поэтому  $P_{020} = 0$ .

1.  $P_{010} = (1 - \pi)P_{021} + P_{020}$
2.  $P_{021} = (1 - \pi)P_{011} + P_{010}$
3.  $P_{i21} = \pi P_{i-111} + (1 - \pi)P_{i11} \quad (i = \overline{1, n-1})$
4.  $P_{i11} = (1 - \pi)P_{i+121} + \pi P_{i21} \quad (i = \overline{0, n-1})$
5.  $P_{n21} = \pi P_{n-111} + (1 - \pi)P_{n11} + (1 - \pi)P_{n01}$
6.  $P_{n11} = \pi P_{n21}$
7.  $P_{n01} = \pi P_{n11} + \pi P_{n01}$

Обозначим:

$$p = P_{010},$$

$$\omega = \frac{\pi}{1 - \pi}.$$

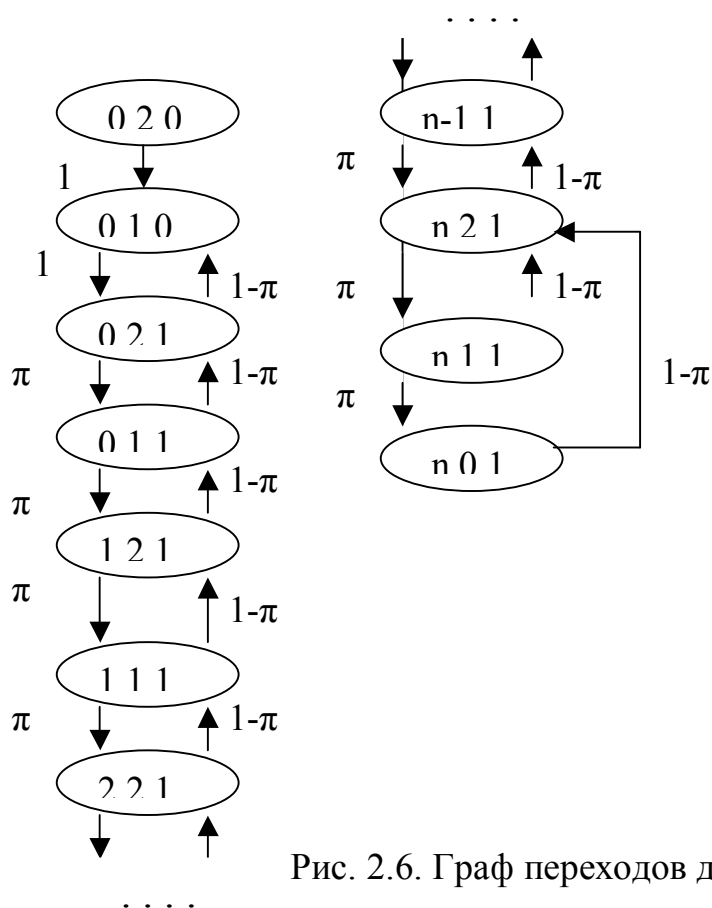


Рис. 2.6. Граф переходов для системы “АЛУ – память”

В состояние  $P_{020}$  система больше не вернется, поэтому  $P_{020} = 0$ .



$$1. P_{010} = (1 - \pi)P_{021} + P_{020}$$

$$2. P_{021} = (1 - \pi)P_{011} + P_{010}$$

$$3. P_{i21} = \pi P_{i-111} + (1 - \pi)P_{i11} \quad (i = \overline{1, n-1})$$

$$4. P_{i11} = (1 - \pi)P_{i+121} + \pi P_{i21} \quad (i = \overline{0, n-1})$$

$$5. P_{n21} = \pi P_{n-111} + (1 - \pi)P_{n11} + (1 - \pi)P_{n01}$$

$$6. P_{n11} = \pi P_{n21}$$

$$7. P_{n01} = \pi P_{n11} + \pi P_{n01}$$

Обозначим:

$$p = P_{010},$$

$$\omega = \frac{\pi}{1 - \pi}.$$

Тогда из уравнений 1 и 2 получим:

$$P_{021} = \frac{1}{1 - \pi} p, P_{011} = \frac{\omega}{1 - \pi} p.$$

Далее, проводя индукцию по  $i$ , будем иметь:

$$P_{i21} = \frac{\omega^{2i}}{1 - \pi} p \quad (i = \overline{0, n})$$

$$P_{i11} = \frac{\omega^{2i+1}}{1 - \pi} p \quad (i = \overline{0, n-1})$$

Из уравнений 6 и 7 системы определяем вероятности:

$$P_{n11} = \omega^{2n+1} p;$$

$$P_{n01} = \omega^{2n+2} p.$$

Уравнение 5 превращается в тождество. Используя уравнение нормировки:

$$\sum P_{jt_1 t_2} = 1,$$

получим:

$$p = \left[ 1 + \frac{1}{1 - \pi} \sum_{i=0}^{2\omega} \omega^i + \omega^{2n+1} + \omega^{2n+2} \right]^{-1}.$$

Отсюда, учитывая, что сумма - это сумма геометрической прогрессии, получим

$$p = \frac{2(1 - \pi) - 1}{2(1 - \pi) - \omega^{2n+2}}.$$

Используя полученное значение  $p$  (фактически, это вероятность простая АЛУ) и рассчитав вероятности всех остальных состояний, можно найти другие интересующие нас характеристики системы.

а) Среднее время обслуживания заявки системой в целом (время пребывания заявки в системе)  $S$ .

При прохождении заявки через систему, определенную долю времени она может пребывать заблокированной в источнике, часть времени находится в очереди и, наконец, некоторое время обслуживаться каналом. Для достаточно большого интервала времени работы системы  $T_p$  среднее время, затраченное на пребывание заявок в фазе обслуживания  $T^* = P_{\text{обсл.}} T_p$ , где  $P_{\text{обсл.}}$  представляет собой сумму вероятностей всех состояний, в которых происходит обслуживание заявки каналом.

Для одной заявки получаем:  $SP_{\text{обсл.}} = m$ , где

$$m = \frac{T}{1 - \pi}$$

среднее значение интервала между событиями в просеянном регулярном потоке с вероятностью просеивания  $\pi$  и тактовым интервалом  $T$ . Заметим, что единственным состоянием данной системы с отличной от нуля вероятностью, в котором канал не обслуживает заявку, является состояние 010, вероятность которого мы обозначили как  $p$ . Следовательно,  $P_{\text{обсл.}} = 1 - p$ . Отсюда:

$$S = \frac{m}{P_{\text{обсл.}}} = \frac{T}{1 - \pi} \cdot \frac{1}{1 - p}.$$

б) Интенсивность потока обработанных заявок (абсолютная пропускная способность):

$$\lambda = (1 - p)(1 - \pi) \frac{1}{T},$$

где  $(1 - p)$  – вероятность того, что канал обрабатывал заявку,  $(1 - \pi)$  – вероятность того, что обработка закончилась.

в) Средняя длина очереди  $L_{\text{оч}} = \sum_j j P_{j112}.$

## 2.7 Непрерывно – стохастические модели (Q – схемы)

Особенностью непрерывно – стохастического подхода при моделировании систем и процессов является использование в качестве типовых математических схем систем массового обслуживания (англ. Queueing system). Системы массового обслуживания представляют собой класс математических схем, разработанных в теории массового обслуживания и различных приложениях для формализации процессов функционирования систем, которые по своей сути являются процессами обслуживания.

### 2.7.1. Системы массового обслуживания. Потоки событий

Реальные системы могут быть представлены при моделировании как системы массового обслуживания (СМО), если при их функционировании можно выделить два процесса - поступление заявок на обслуживание и обслуживание заявок. Таким образом могут быть представлены различные по своей физической природе процессы – экономические, технические, производственные и т.д.

Для описания работы СМО используется понятие *поток событий* – последовательность событий, происходящих одно за другим в некоторые моменты времени. В СМО будем выделять три потока:

- входной поток: множество моментов времени поступления в систему заявок;
- поток обслуживаний: множество моментов времени окончания обработки системой заявок в предположении, что обслуживание осуществляется непрерывно;
- выходной поток: последовательность моментов времени ухода из системы обслуженных заявок.

В общем случае СМО элементарного вида может быть представлено следующим образом (рис. 2.7).

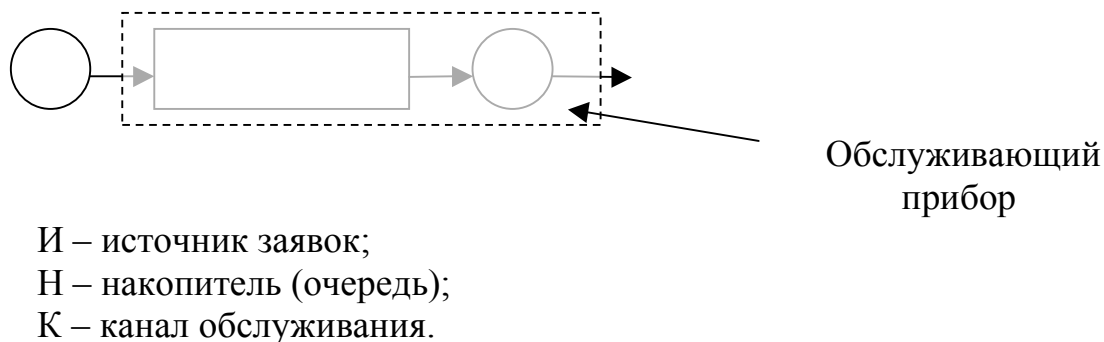


Рис. 2.7. СМО общего вида

СМО состоит из какого-то числа обслуживающих единиц, которые мы будем называть *каналами* обслуживания. В зависимости от количества каналов СМО могут быть одноканальными и многоканальными.

СМО вообще могут быть двух типов.

1. *СМО с отказами.* В таких системах заявка, поступившая в момент, когда все каналы заняты, получает отказ и покидает систему не обслуженной.

2. *СМО с ожиданием.* Заявка, заставшая все каналы занятыми, становится в очередь и ожидает освобождения одного из каналов.

Число мест в очереди  $m$  может быть как неограниченным, так и ограниченным. В первом случае заявка может стать в очередь в любой момент, в то время как во втором она вынуждена покинуть систему, если очередь превышает заданную длину в момент прихода заявки или в процессе ожидания.

При  $m = 0$  СМО с очередью превращается в СМО с отказами. Очередь может иметь ограничения не только по количеству стоящих в ней заявок (длине очереди), но и по времени ожидания (такие СМО называются «системами с нетерпеливыми клиентами»). В системах с бесконечным ожиданием (поступившая заявка ожидает обслуживания сколь угодно долго, в то время как в СМО с ограничением по длительности ожидания заявка покидает систему, если выполняются некоторые условия, например длительность ожидания превышает заданную величину).

Порядок выборки заявок из очереди определяется *дисциплиной обслуживания*. Некоторые наиболее употребляемые дисциплины:

- 1) FIFO (first in – first out) – в порядке поступления;
- 2) LIFO (last in – first out) – первой обслуживается заявка, поступившая последней;
- 3) SIRO (service in random order) – обслуживание в случайном порядке;
- 4) приоритетные системы – заявки обслуживаются в соответствии с их приоритетами.

Если в системе массового обслуживания имеется очередь заявок, то освобождающиеся каналы занимают немедленно в порядке их освобождения. В случае, когда очереди заявок нет, и имеются свободные каналы, появившаяся заявка может занимать один из свободных каналов в соответствии со специальными правилами. Наиболее часто на практике используются следующие правила.

- 1) Каналы занимают в порядке их номеров. Канал с большим номером не может быть привлечен к обслуживанию заявки, если имеется свободный канал с меньшим номером.
- 2) Каналы занимают в порядке очереди. Освободившийся канал поступает в очередь и не начинает обслуживания заявок до загрузки всех ранее освободившихся каналов.
- 3) Каналы занимают в случайном порядке в соответствии с заданными вероятностями.

Перечисленными способами выборки заявок из очереди и занятия свободных каналов, естественно, охватываются не все случаи, возникающие на практике, а лишь наиболее распространенные. Например, в первую очередь могут выбираться из очереди заявки, требующие меньшего времени обслуживания. В «системах с нетерпеливыми клиентами» заявки могут приниматься к обслуживанию по минимальному времени до получения отказа.

Реальный процесс функционирования системы массового обслуживания для удобства исследования можно представлять в виде последовательности отдельных актов (фаз) обслуживания, выполняемых различными устройствами. При этом, как правило, соблюдается такой порядок, при котором следующее устройство может

приступить к обслуживанию заявки лишь тогда, когда работа предыдущего с данной заявкой полностью закончена.

В частном случае обслуживание может быть однофазным.

Простейшим примером многофазного обслуживания является обслуживание покупателей в магазине. Сначала покупатель занимает одного из работников прилавка, демонстрирующего товары и оформляющего товарные чеки (первая фаза). Отобрав товары и получив чек, покупатель должен пройти через вторую фазу — оплатить чек в кассе. И только с оплаченным чеком покупатель может быть принят на обслуживание в отдел контроля и выдачи покупок (третья фаза).

Характеристики СМО существенно зависят от вида и параметров входного потока и потока обслуживаний.

Поток событий называется *однородным*, если он характеризуется только моментами наступления событий и задается последовательностью  $\{t_n\}$ , где

$$\{t_n\} = \{0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq \dots\}.$$

Соответственно, поток событий является *неоднородным*, если он задается последовательностью  $\{t_n, f_n\}$ , где  $t_n$  — моменты наступления событий, а  $f_n$  — набор признаков событий (приоритеты, принадлежность тому или иному источнику, тип канала для его обслуживания и т. д.).

Поток *регулярный*, если события поступают через равные промежутки времени. Соответственно, поток *нерегулярный*, если интервалы между событиями представляют собой случайные величины.

Если промежутки времени между последовательными событиями представляют собой независимые, одинаково распределенные случайные величины, поток называется *рекуррентным* (поток Пальма, поток с ограниченным последствием).

Для краткой характеристики СМО Д. Кендалл ввел символику (нотацию):

$A/B/s/m/k$ , где

$s$  - число обслуживающих приборов;

$m$  — количество мест ожидания (если не указано, то считается, что  $m=0$ , т.е. это система с отказами); при неограниченной очереди в качестве  $m$  ставят символ  $\infty$ ;

$k$  — количество источников заявок. Если  $k$  отсутствует, то по умолчанию количество источников предполагается равным одному.

$A$  и  $B$  характеризуют соответственно: поток требований и поток обслуживания, задавая функцию распределения интервалов между заявками во входном потоке и функцию распределения времен обслуживания.

$A$  и  $B$  могут принимать значения:

$D$  — детерминированное распределение;

$M$  — показательное;

$E_r$  — распределение Эрланга порядка  $r$ ;

$H_r$  - гиперпоказательное;

G – распределение общего вида.

При этом подразумевается, что потоки являются рекуррентными, т.е. интервалы между событиями независимы и имеют одинаковое распределение.

Обязательными в этой нотации являются первые три позиции.

### 2.7.2. Простейший поток

Среди всех возможных потоков при исследовании СМО особую роль играет поток, который называется *простейшим*. Такое название потока связано с тем, что математическое описание событий, связанных с простейшими потоками, оказывается наиболее простым. Даже самый простой, на первый взгляд, регулярный поток со строго постоянными интервалами между событиями доставляет больше трудностей при создании аналитической модели: он обладает ярко выраженным последствием, так как моменты появления событий в нем связаны между собой жесткой функциональной зависимостью.

Простейшим называется поток, обладающий следующими тремя свойствами: стационарность, ординарность и отсутствие последствия. Расшифруем эти понятия.

Поток является *стационарным*, если вероятность наступления заданного числа событий в течение интервала времени фиксированной длины зависит только от продолжительности интервала и не зависит от его расположения на временной оси. Иначе говоря, вероятностные характеристики и интенсивность такого потока со временем не изменяются.

Поток является *ординарным*, если вероятность появления двух или более событий в течение элементарного интервала времени  $\Delta t \rightarrow 0$  есть величина бесконечно малая по сравнению с вероятностью появления одного события на этом интервале. Другими словами, два и более событий в таком потоке произойти одновременно не могут.

Поток называется потоком *без последствия*, если для любых неперекрывающихся интервалов времени число событий, попадающих на один из них, не зависит от числа событий, попадающих на другие. Иногда это свойство формулируют следующим образом: распределение времени до ближайшего события не зависит от времени наблюдения, т.е. от того, сколько времени прошло после последнего события. Отсутствие последствия в потоке означает, что события, образующие поток появляются в последовательные моменты времени независимо друг от друга.

Для простейшего потока число событий, попадающих на любой фиксированный интервал времени подчиняется закону Пуассона, поэтому его иначе называют стационарным пуассоновским.

Вероятность того, что за интервал времени  $\tau$  в простейшем потоке произойдет ровно  $m$  событий

$$P_m(\tau) = \frac{(\lambda\tau)^m}{m!} e^{-\lambda\tau}$$

Здесь  $\lambda$  – интенсивность, т.е. среднее число событий в единицу времени.

Вероятность того, что в течении интервала времени  $\tau$  не произойдет не одного события

$$P_0(\tau) = e^{-\lambda\tau}.$$

Вероятность того, что за это время произойдет хотя бы одно событие

$$P_{\geq 1}(\tau) = 1 - e^{-\lambda\tau}.$$

При построении и анализе непрерывно – стохастических моделей исследуются вероятности появления или не появления событий в течении элементарного интервала времени  $\Delta t \rightarrow 0$ . Произведя замену  $\tau$  на  $\Delta t$ , и, разлагая  $e^{-\lambda\Delta t}$  в степенной ряд, получим, пренебрегая величинами высшего порядка малости

$$P_0(\Delta t) = e^{-\lambda\Delta t} = 1 - \lambda\Delta t + \frac{1}{2!}(\lambda\Delta t)^2 - \dots = 1 - \lambda\Delta t + 0\Delta t \cong 1 - \lambda\Delta t.$$

Вероятность появления хотя бы одной заявки

$$P_{\geq 1}(\Delta t) = 1 - P_0(\Delta t) = \lambda\Delta t.$$

Учитывая ординарность простейшего потока, можно утверждать, что последнее выражение представляет собой вероятность появления одного события.

Иногда удобней анализировать системы, рассматривая случайные интервалы  $T$  между событиями. Их величины имеют показательное распределение, для которого функция распределения вероятностей и функция плотности вероятностей имеют вид:

$$F(t) = 1 - e^{-\lambda t},$$

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

Математическое ожидание и среднее квадратическое отклонение для  $T$ :

$$m_T = \sigma_T = \frac{1}{\lambda}$$

Одной из числовых характеристик для потоков является коэффициент вариации – отношение среднего квадратичного к математическому ожиданию:

$$v = \frac{\sigma}{m}.$$

Для реальных потоков значение этого коэффициента лежит в пределах от 0 до 1. Коэффициент вариации характеризует степень

стохастичности, непредсказуемости потока. Для регулярного потока с равными интервалами между событиями  $\nu = 0$ . Для простейшего потока  $\nu = 1$ , то есть этот поток является наиболее стохастичным среди всех потоков.

Заметим, что показательное распределение интервала между событиями – недостаточное условие для того, чтобы поток был простейшим.

Рассмотрим пример. Пусть имеется простейший поток событий с интенсивностью  $\lambda$ . Из этого потока формируется другой следующим образом: точно в середину каждого интервала вставим еще одно событие. Распределение интервалов в исходном потоке показательное

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

Новый поток событий имеет интенсивность  $2\lambda$ . Можно показать, что интервалы времени в этом потоке тоже распределены по показательному закону с функцией плотности вероятностей

$$f(t) = 2\lambda e^{-2\lambda t}.$$

Тем не менее, полученный поток не является простейшим. Рассмотрим два соседних интервала между событиями в этом потоке. С вероятностью  $1/2$  они независимы, с вероятностью  $1/2$  равны друг другу, следовательно, зависимы. Таким образом, в сформированном потоке присутствует последствие.

Простейший поток обладает следующими особенностями:

1. Сумма  $M$  независимых, ординарных, стационарных потоков заявок с интенсивностями  $\lambda_i$  ( $i = 1, \dots, M$ ) сходится к простейшему потоку с интенсивностью, равной сумме интенсивностей исходных потоков при условии, что складываемые потоки оказывают приблизительно одинаково малое влияние на суммарный поток. Сходимость суммарного потока к простейшему осуществляется очень быстро. Практически можно считать, что сложение четырех-пяти стационарных, ординарных, независимых потоков, сравнимых по интенсивности, достаточно для того, чтобы суммарный поток был близок к простейшему.

Таким образом, для выяснения всех свойств суммарного потока достаточно знать лишь интенсивности суммируемых потоков и практически не требуется знать внутреннюю структуру этих потоков.

Простейший поток обладает устойчивостью, состоящей в том, что при суммировании независимых простейших потоков получается снова простейший поток, причем интенсивности складываемых потоков суммируются.

2. Поток заявок, полученный путем случайного разрежения исходного потока, когда каждая заявка с определенной вероятностью  $p$  исключается из потока независимо от того, исключены другие заявки или нет, образует простейший поток с интенсивностью

$$\lambda_p = p \lambda,$$



где  $\lambda$  — интенсивность исходного потока. В отношении исходного потока заявок делается предположение лишь об ординарности и стационарности.

3. Для простейшего потока характерно, что поступление заявок через короткие промежутки времени более вероятно, чем через длинные,— 63% промежутков времени между заявками имеют длину, меньшую среднего периода  $1/\lambda$ . Следствием этого является то, что простейший поток по сравнению с другими видами потоков создает наиболее тяжелый режим работы системы. Поэтому предположение о том, что на вход системы поступает простейший поток заявок, приводит к определению предельных значений характеристик качества обслуживания. Если реальный поток отличен от простейшего, то система будет функционировать не хуже, чем это следует из полученных оценок.

4. Интервал времени между произвольным моментом времени и моментом поступления очередной заявки имеет такое же распределение с тем же средним  $M[\tau] = 1/\lambda$ , что и интервал времени между двумя последовательными заявками. Эта особенность простейшего потока является следствием отсутствия последствия.

Важная роль простейшего потока при моделировании определяется тем, что простейшие или близкие к ним потоки часто встречаются на практике. Кроме того, при анализе СМО можно получить вполне удовлетворительные результаты, заменяя входной поток любой структуры простейшим с той же интенсивностью.

В заключение заметим, что выходные потоки обычно имеют последствие, даже если входные потоки им не обладают. Рассмотрим, например, СМО с фиксированным временем обслуживания  $\tau_{\text{обсл}}$  (рис.2.8).

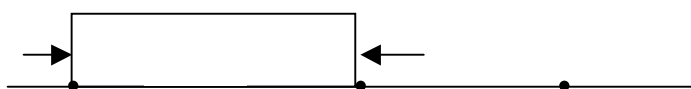


Рис. 2.8. Последствие при регулярном обслуживании

Если в момент  $t_1$  на выходе появилась обслуженная заявка, то на интервале  $(t_1, t_1 + \tau_{\text{обсл}})$  заявка не появится.

### 2.7.3. Непрерывные марковские цепи. Уравнения Колмогорова

Рассмотрим некоторую физическую систему с дискретными состояниями  $S_1, S_2, \dots, S_n$ , которая переходит из состояния в состояние под влиянием каких-то случайных событий. Если все потоки, переводящие систему из состояния в состояние, пуассоновские, то процесс,

протекающий в системе, будет марковским. Действительно, пуассоновский поток обладает отсутствием последствия, поэтому, при заданном состоянии системы в заданный момент, ее переходы в другие состояния в будущем обусловлены только появлением каких-то событий в пуассоновских потоках, а вероятности появления этих событий не зависят от предыстории процесса.

Процессы в системах с дискретными состояниями и непрерывным временем называются *непрерывными марковскими цепями*.

Рассмотрим систему, для которой при исследовании Р – схем мы полагали входной поток детерминированным и поток обслуживания просеянным (интервалы времени обслуживания подчинялись геометрическому распределению).

Теперь мы будем считать, что входной поток простейший с интенсивностью  $\lambda$ , а поток обслуживания – пуассоновский с интенсивностью  $\mu$ . Число мест ожидания в очереди ограничено и равно  $n$ . Согласно нотации Кендалла мы можем определить вид такой системы как СМО М /М/ 1/ $n$  с блокировкой источника. Ее структура представлена на рисунке 3.3.

Как и прежде предполагаем, что дисциплина обслуживания FIFO с блокировкой источника.

Состояние системы будем определять числом заявок, находящихся в данный момент в системе. Таким образом, всего возможны  $n+3$  состояния: от 0 до  $n+2$ . Напомним, что, если число состояний системы конечно и из каждого состояния можно перейти (за то или иное число шагов) в любое другое, то существуют предельные (финальные) вероятности состояний, которые не зависят от времени и начального состояния системы.

Обозначим  $R_i(\Delta t)$  – вероятность прихода за время  $\Delta t$   $i$  заявок, а  $R'_i(\Delta t)$  – вероятность обслуживания за  $\Delta t$   $i$  заявок.

Тогда вероятность перехода системы из состояния  $s$  в состояние  $s+1$  за время  $\Delta t$  (увеличения числа заявок в системе на 1) будет равняться

$$\omega_{s,s+1}(\Delta t) = R_1(\Delta t) \cdot R'_0(\Delta t) + \underbrace{R_2(\Delta t)R'_1(\Delta t) + \dots}_{0(\Delta t)}$$

Ввиду ординарности потоков слагаемыми, начиная со второго, можно пренебречь и, подставив в формулу соответствующие вероятности, получаем

$$\omega_{s,s+1}(\Delta t) = \lambda \Delta t (1 - \mu \Delta t) = \lambda \Delta t + \lambda \mu (\Delta t)^2 \approx \lambda \Delta t.$$

Аналогично получим вероятности других переходов:

$$\omega_{s,s-1}(\Delta t) = R_0(\Delta t) \cdot R'_1(\Delta t) + \underbrace{R_1(\Delta t)R'_2(\Delta t) + \dots}_{0(\Delta t)} = (1 - \lambda \Delta t) \mu \Delta t \approx \mu \Delta t,$$

$$\omega_{s,s}(\Delta t) = R_0(\Delta t) \cdot R'_0(\Delta t) + R_1(\Delta t)R'_1(\Delta t) + \dots = (1 - \lambda \Delta t)(1 - \mu) \Delta t + \lambda \mu (\Delta t)^2 + \dots \approx 1 - (\lambda + \mu) \Delta t.$$

Вероятность подтверждения состояния 0 равна вероятности того, что за время  $\Delta t$  в систему не поступит ни одной новой заявки:

$$\omega_{0,0}(\Delta t) = (1 - \lambda \Delta t).$$

В состоянии  $n+2$  источник заблокирован и заявки системе не поступают, поэтому вероятность того, что за время  $\Delta t$  изменения состояния не произойдет, равна вероятности необслуживания заявки:

$$\omega_{n+2,n+2}(\Delta t) = (1 - \mu \Delta t).$$

Теперь можно построить граф состояний для нашей системы (рис. 2.9).

Построим по графу систему уравнений для вероятностей состояний.

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda \Delta t) + P_1(t)\mu \Delta t;$$

$$P_i(t + \Delta t) = P_i(t)[1 - (\lambda + \mu)\Delta t] + P_{i-1}(t)\lambda \Delta t + P_{i+1}(t)\mu \Delta t, \quad (i = \overline{1, n+1});$$

$$P_{n+2}(t + \Delta t) = P_{n+2}(t)[1 - \mu \Delta t] + P_{n+1}(t)\lambda \Delta t.$$

Преобразуем эту систему. Рассмотрим эту процедуру на примере первого уравнения. Раскроем скобки в правой части, перенесем в левую часть  $P_0(t)$  и разделим обе части на  $\Delta t$ :

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t) + \mu P_1(t).$$

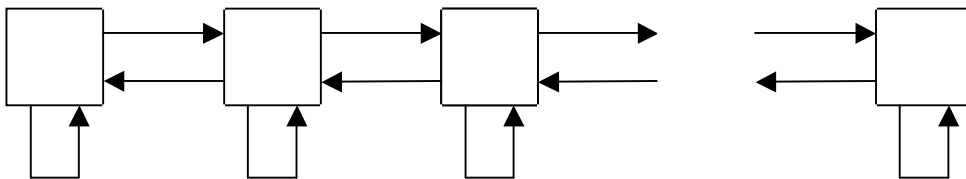


Рис. 2.9. Граф состояний системы М /М/ 1/п с блокировкой

При  $\Delta t \rightarrow 0$  левая часть есть не что иное, как производная от  $P_0(t)$ :

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t).$$

Таким образом, наша система превратилась в систему дифференциальных уравнений, которая носит название *системы уравнений Колмогорова*.

Напомним, что, если число состояний системы конечно и из каждого состояния можно перейти (за то или иное число шагов) в любое другое, то существуют предельные (финальные) вероятности состояний, которые не зависят от времени и начального состояния системы. Таким образом, получаем

$$P_0(t) = P_0, \quad P_1(t) = P_1 \quad \text{и, следовательно,} \quad \frac{dP_0(t)}{dt} = 0.$$

Теперь первая строка нашей системы примет вид:

$$-\lambda P_0 + \mu P_1 = 0.$$

После проведения аналогичных преобразований для остальных строк, система будет выглядеть так:

$$\begin{aligned} &-\lambda P_0 + \mu P_1 = 0; \\ &-(\lambda + \mu)P_i + \lambda P_{i-1} + \mu P_{i+1} = 0, \quad (i = \overline{1, n+1}); \\ &-\lambda P_{n+1} + \mu P_{n+2} = 0. \end{aligned}$$

Важность полученного результата заключается в том, что от системы дифференциальных уравнений мы перешли к системе алгебраических уравнений, которую можно достаточно просто решить и получить значения вероятностей состояний. Но предварительно преобразуем её, начиная со второго и заканчивая предпоследним уравнением – новое уравнение получаем сложением старого с новым предыдущим и заменяя им старое значение.

В результате, новое предпоследнее уравнение будет совпадать со старым последним уравнением, и все строки будут иметь вид:

$$-\lambda P_i + \mu P_{i+1} = 0, \quad (i = \overline{1, n+1}).$$

Введем обозначения

$$\frac{\lambda}{\mu} = \omega, \quad P_0 = p.$$

Проанализируем полученную систему.

$$\begin{aligned} &-\lambda P_0 + \mu P_1 = 0, \quad P_1 = \frac{\lambda}{\mu} P_0 = \omega p; \\ &-\lambda P_1 + \mu P_2 = 0, \quad P_2 = \frac{\lambda}{\mu} P_1 = \omega^2 p; \end{aligned}$$

$$\dots$$

$$P_i = \omega^i p, \quad i = \overline{0, n+2};$$

Используем уравнение нормировки:

$$\sum_{i=0}^{n+2} P_i = 1;$$

или 
$$p \sum_{i=0}^{n+2} \omega^i = 1;$$

Это сумма геометрической прогрессии. Напомним, что сумма  $k$  элементов конечной геометрической прогрессии равна

$$S_k = \frac{b_1(q^k - 1)}{q - 1}, \text{ где } b_1 - \text{первый член прогрессии, а } q - \text{ее множитель.}$$

Отсюда:

$$p = P_0 = \frac{\omega - 1}{(\omega^{n+3} - 1)}.$$

Зная  $P_0$ , можно рассчитать значения вероятности всех остальных состояний.

Определим некоторые показатели эффективности работы системы так, как это было сделано для дискретно-стохастической модели.

Среднее время пребывания заявки в системе

$$S = \frac{m}{P_{\text{обсл}}},$$

где  $m$  – среднее время обслуживания заявки каналом, равное  $1/\mu$ , а  $P_{\text{обсл}}$  – вероятность того, что канал обслуживает заявку, равная  $1-p$ .

Отсюда

$$S = \frac{1}{\mu(1-p)} = \frac{1 - \omega^{n+3}}{\mu\omega(1 - \omega^{n+2})} = \frac{1 - \omega^{n+3}}{\lambda(1 - \omega^{n+2})}.$$

Средняя длина очереди

$$L_{\text{оч}} = \sum_{i=1}^{n+1} (i-1)P_i + nP_{n+2}.$$

#### 2.7.4. Диаграмма интенсивностей переходов

Исследуя в предыдущем разделе систему М/М/1/п с блокировкой, мы получили искомые результаты, используя следующую последовательность действий:

- построение графа Марковской цепи;
- определение вероятностей переходов;
- построение системы дифференциальных уравнений;
- переход к системе алгебраических уравнений;
- решение системы и получение вероятностей состояний.

Рассмотрим теперь более простой способ непосредственного выписывания алгебраических уравнений.

Построим вместо графа марковского процесса (рис. 2.9) *диаграмму интенсивностей переходов* (ДИП), в которой овалами обозначим состояния, а дугами переходы между ними. Как и прежде, будем кодировать состояния числом заявок, находящихся в данный момент в системе. Дугам приписываются числа, соответствующие интенсивностям

переходов (а не вероятностям, как это было ранее). В такой диаграмме нет петель, т.е. дуг из состояния  $S_k$  в это же состояние (рис. 2.10).

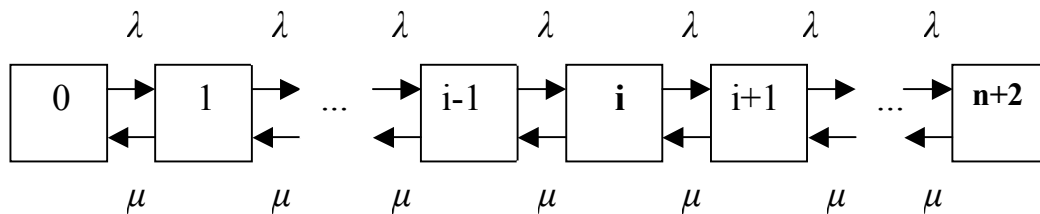


Рис. 2.10. ДИП для системы М /М/ 1/п с блокировкой

Когда переходы осуществляются только в соседние состояния, говорят о процессе «размножения и гибели» и соответственно об интенсивностях, с которыми может увеличиваться или уменьшаться количество заявок в системе.

Эта диаграмма содержит, по сути, ту же информацию, что и граф сети Маркова. Построим для этой диаграммы алгебраическое уравнение для вероятностей состояний, исходя из закона сохранения потоков вероятностей: сумма входящих потоков вероятностей для любого состояния равна сумме исходящих из него потоков. В данном случае мы будем понимать под термином “поток вероятности” произведение вероятности состояния на интенсивность перехода из этого состояния.

Так, для состояния с номером  $i$  ДИП на рис. 2.10 два входных потока вероятностей:

- от состояния  $i-1$ , равный  $\lambda P_{i-1}$ ;
- от состояния  $i+1$ , равный  $\mu P_{i+1}$ .

Выходные потоки для этого состояния:

- к состоянию  $i-1$ , равный  $\mu P_i$ ;
- к состоянию  $i+1$ , равный  $\lambda P_i$ ;

Так как сумма входящих потоков вероятностей для состояния равна сумме исходящих из него потоков, получаем

$$(\lambda + \mu)P_i = \lambda P_{i-1} + \mu P_{i+1}.$$

Вспомним рассмотренную в разделе 2.7.3 систему М/М/1/п. Уравнение для  $i$ -го состояния, которое мы получили, исследуя ее:

$$-(\lambda + \mu)P_i + \lambda P_{i-1} + \mu P_{i+1} = 0.$$

Сравнивая эти два выражения, убеждаемся в полной их идентичности.

Для процессов «размножения и гибели» выписывание алгебраических уравнений для вероятностей состояний можно еще больше упростить, если воспользоваться следующим правилом: встречные потоки вероятностей через сечение диаграммы равны между

собой. Например, для сечения между состояниями  $i$  и  $i+1$  это выражается равенством

$$\lambda P_i = \mu P_{i+1}.$$

Этот результат мы уже получили ранее в разделе 2.7.3, преобразовав исходную систему алгебраических уравнений.

### 2.7.5 Формула Литтла

Прежде чем перейти к анализу СМО различной конфигурации, выведем для системы с неограниченной очередью важное соотношение между некоторыми характеристиками системы, которое называется формула Литтла.

Эта формула связывает для предельного (стационарного режима работы) среднее число заявок, находящихся в СМО  $L_c$  и среднее время пребывания заявки в системе  $W_c$ .

Рассмотрим произвольную СМО (одноканальную, многоканальную, марковскую, немарковскую и т.д.). Если интенсивность поступления заявок, поступающих в систему будет меньше, чем интенсивность обработки этих заявок каналом, то установится стационарный (предельный) режим работы системы, когда среднее число прибывающих за единицу времени заявок равно среднему числу заявок, покидающих СМО (т.е. оба потока – входной и выходной будут иметь одну и ту же интенсивность  $\lambda$ ).

Введем в рассмотрение две функции:  $X(t)$  и  $Y(t)$ , значениями которых будут являться соответственно число заявок, поступивших к моменту  $t$  в систему и число заявок, покинувших систему к этому моменту (рис.2.11).

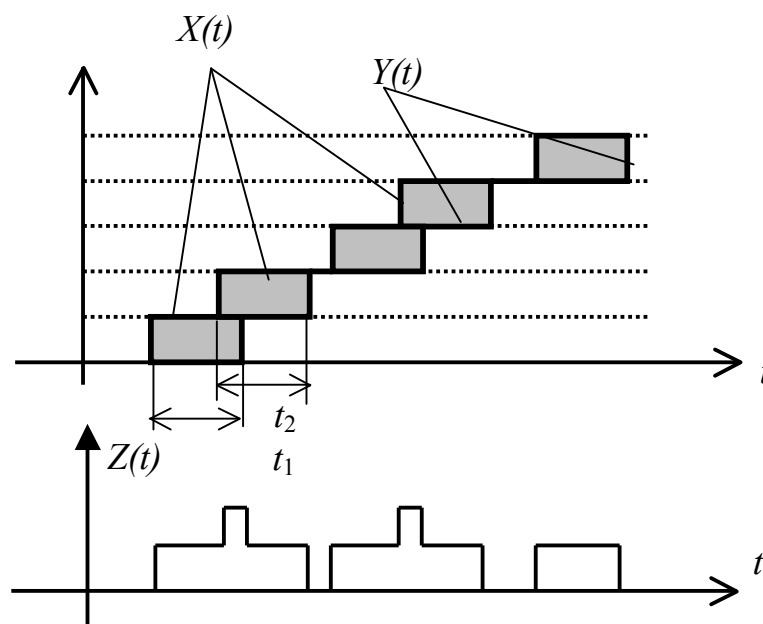


Рис. 2.11. К выводу формулы Литтла.

Обе функции случайны, меняются скачком на 1 при приходе или окончании обслуживания очередной заявки соответственно.. Число заявок, находящихся в СМО в момент  $t$  равно:

$$Z(t) = X(t) - Y(t).$$

Рассмотрим очень большой промежуток времени  $T$  и вычислим для него среднее число заявок, находящихся в СМО. Это среднее значение функции  $Z(t)$  интервале  $T$ :

$$L_c = \frac{1}{T} \int_0^T Z(t) dt .$$

Интеграл представляет собой площадь заштрихованной фигуры, состоящей из прямоугольников высотой 1 и длиной, равной времени пребывания в системе соответствующей заявки  $t_1, t_2 \dots t_i \dots$ . Следовательно, мы можем заменить этот интеграл суммой площадей и записать:

$$L_c = \frac{1}{T} \sum_i t_i .$$

Умножим и разделим это выражение на  $\lambda$ :

$$L_c = \frac{\lambda}{T\lambda} \sum_i t_i$$

Но  $T\lambda$  - это количество заявок, поступивших в систему за время  $T$ , а  $\sum_i t_i$  представляет собой суммарное время пребывания всех этих заявок в системе. Следовательно, отношение  $\sum_i t_i$  к  $T\lambda$  является средним временем пребывания заявки в СМО, т.е.  $W_c$ . Окончательно:

$$L_c = \lambda W_c$$

Аналогично выводится соотношение

$$L_{оч} = \lambda W_{оч} ,$$

где  $L_{оч}$  - средняя длина очереди,

$W_{оч}$  - среднее время пребывания заявки в очереди.

Эти формулы носят название формул Литтла и позволяют при анализе систем по значению  $L$  находить  $W$  и наоборот.



### 2.7.6. Исследование СМО с помощью диаграмм интенсивностей переходов

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

- среднее число заявок  $A$ , обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО;
- вероятность обслуживания поступившей заявки  $Q$  или *относительная пропускная способность* СМО:

$$Q = A/\lambda;$$

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

$$P_{отк} = 1 - Q;$$

- среднее число заявок в СМО (обслуживаемых или ожидающих в очереди)  $L_c$ ;

- среднее число заявок в очереди  $L_{оч}$ ;

- среднее время пребывания заявки в СМО (в очереди или под обслуживанием)  $W_c$ ;

- среднее время пребывания заявки в очереди  $W_{оч}$ ;

- среднее число занятых каналов  $\bar{k}$ .

В общем случае все эти характеристики зависят от времени. Но многие СМО работают в неизменных условиях достаточно долгое время, и поэтому для них успевает установиться режим, близкий к стационарному. В дальнейшем, не оговаривая этого каждый раз специально, будем вычислять финальные вероятности состояний и финальные характеристики эффективности СМО, относящиеся к предельному, стационарному режиму ее работы.

Рассмотрим наиболее часто встречающиеся на практике конфигурации СМО и определим показатели эффективности их функционирования.

**1) Одноканальная СМО с фиксированным числом мест ожидания (с ограниченной очередью) – М/М/1/ $m$ .** Входной поток заявок простейший с интенсивностью  $\lambda$ , поток обслуживаний простейший с интенсивностью  $\mu$ , количество мест ожидания  $m$ . Заявка, заставшая очередь полностью заполненной, теряется. Максимальное количество заявок, присутствующих в системе –  $m+1$  ( $m$  заявок в очереди и одна заявка в канале). ДИП для этой системы представлена на рис. 2.12.

$\lambda \quad \lambda \quad \lambda \quad \lambda \quad \lambda \quad \lambda$

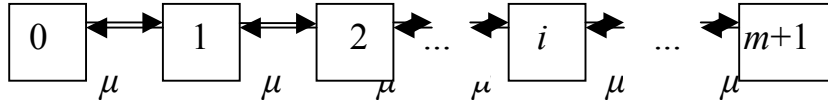


Рис. 2.12. ДИП для системы **M/M/1/m**

Воспользовавшись правилом равенства встречных потоков вероятностей через сечение диаграммы и, введя обозначения  $P_0 = p$  и  $\lambda/\mu = \omega$ , получим:

$$\begin{aligned}\lambda P_0 &= \mu P_1, & P_1 &= (\lambda/\mu)P_0 = \omega p; \\ \lambda P_1 &= \mu P_2, & P_2 &= (\lambda/\mu)P_1 = \omega^2 p; \\ \lambda P_2 &= \mu P_3, & P_3 &= (\lambda/\mu)P_2 = \omega^3 p;\end{aligned}$$

...

По индукции получаем  $P_i = \omega^i p$ .

Для определения значения  $P_0 = p$  воспользуемся нормировочным уравнением:

$$\sum_{i=0}^{m+1} P_i = 1, \text{ или } \sum_{i=0}^{m+1} \omega^i p = p \sum_{i=0}^{m+1} \omega^i = 1.$$

Учитывая, что полученная сумма представляет собой сумму геометрической прогрессии, в которой  $n+2$  слагаемых, получим:

$$\frac{p(1 - \omega^{m+2})}{1 - \omega} = 1.$$

$$\text{Окончательно: } p = \frac{1 - \omega}{1 - \omega^{m+2}}. \quad (2.7.1)$$

#### **Характеристики эффективности:**

— относительная пропускная способность

$$Q = 1 - P_{m+1};$$

— абсолютная пропускная способность

$$A = \lambda(1 - P_{m+1});$$

— вероятность отказа

$$P_{\text{отк}} = P_{m+1} = \omega^{m+1} p;$$

— среднее число занятых каналов (вероятность того, что канал занят)

$$\bar{k} = 1 - P_0 = 1 - \frac{1 - \omega}{1 - \omega^{m+2}}.$$

— среднее число заявок в очереди

$$L_{\text{оч}} = \sum_{i=1}^m i P_{i+1} = \sum_{i=1}^m i \omega^{i+1} p = \omega^2 p \sum_{i=1}^m i \omega^{i-1} = \omega^2 p \sum_{i=1}^m \frac{d}{d\omega} (\omega^i) = \omega^2 p \frac{d}{d\omega} \sum_{i=1}^m \omega^i.$$

Сумма в этом выражении представляет собой сумму  $m$  элементов геометрической прогрессии и равна  $\frac{\omega(1-\omega^m)}{1-\omega}$ .

Производная от этого выражения  $\frac{1-\omega^n(m+1-m\omega)}{(1-\omega)^2}$ .

Учитывая выражение для  $p$ , окончательно имеем:

$$L_{оч} = \frac{\omega^2[1-\omega^m(m+1-m\omega)]}{(1-\omega)(1-\omega^{m+2})}.$$

— среднее число заявок в СМО

$$L_c = L_{оч} + \bar{k}.$$

По формуле Литтла можно получить значение среднего времени пребывания заявки в очереди:

$$W_{оч} = L_{оч} / \lambda.$$

Для определения *среднего времени пребывания заявки в системе* будем исходить из того, что

$$W_{сис} = W_{оч} + \bar{t}_{обсл},$$

где  $\bar{t}_{обсл}$  — среднее время обслуживания заявки каналом, которое может принимать значение  $\frac{1}{\mu}$ , если заявка попала в систему (вероятность этого равна  $1-P_{отк}$ ), или 0, если заявка получила отказ (с вероятностью  $P_{отк}$ ). Отсюда

$$\bar{t}_{обсл} = 0 \cdot P_{отк} + \frac{1}{\mu}(1-P_{отк}) = \frac{Q}{\mu}.$$

Окончательно:

$$W_{сис} = W_{оч} + \frac{Q}{\mu}.$$

Обратим внимание на то, что формула (2.7.1) справедлива только при  $\omega \neq 0$  (при  $\omega=0$  она превращается в неопределенность вида  $0/0$ ). Но, учитывая, что сумма геометрической прогрессии  $\sum_{i=0}^{m+1} \omega^i$  при  $\omega=0$  равна  $m+2$ , получим в этом случае

$$p = \frac{1}{m+2}.$$

Исходя из этого, характеристики эффективности системы могут быть определены следующим образом:

$$L_{оч} = \frac{m(m+1)}{2(m+2)}; \quad L_c = L_{оч} + (1-p) = \frac{m+1}{2}.$$

**Пример.** Автозаправочная станция (АЗС) представляет собой СМО с одним каналом обслуживания (одной колонкой). Площадка при станции допускает пребывание в очереди на заправку не более трех

машин одновременно ( $m=3$ ). Если в очереди уже находится три машины, очередная машина, прибывшая к станции, в очередь не становится, а проезжает мимо. Поток машин, прибывающих для заправки, имеет интенсивность  $\lambda = 1$  (машина в минуту). Процесс заправки продолжается в среднем 1,25 мин. Определить:

- вероятность отказа;
- относительную и абсолютную пропускную способности СМО;
- среднее число машин, ожидающих заправки;
- среднее число машин, находящихся на АЗС (включая и обслуживаемую);
- среднее время ожидания машины в очереди;
- среднее время пребывания машины на АЗС (включая обслуживание).

**Решение.** Находим приведенную интенсивность потока заявок:

$$\mu = 1/1,25 = 0,8; \quad \omega = \lambda / \mu = 1/0,8 = 1,25.$$

По формулам для системы М/М/1/м получаем:

$$p = \frac{1 - 1,25}{1 - 1,25^5} = 0,122.$$

Вероятность отказа

$$P_{отк} = P_{m+1} = 1,25^4 \cdot 0,122 = 0,297.$$

Относительная пропускная способность СМО

$$Q = 1 - P_{отк} = 0,703 = 0,703.$$

Абсолютная пропускная способность СМО

$$A = \lambda Q = 0,703 \text{ (машины в мин.)}$$

Среднее число машин в очереди

$$L_{оч} = \frac{1,25^2 [1 - 1,25^3 (3 + 1 - 3 \cdot 1,25)]}{(1 - 1,25)(1 - 1,25^5)} \approx 1,56.$$

Прибавляя к этой величине среднее число машин, находящихся под обслуживанием

$$\bar{k} = 1 - \frac{1 - 1,25}{1 - 1,25^5} \approx 0,88,$$

получаем среднее число машин, связанных с АЗС:

$$L_c = L_{оч} + \bar{k} = 2,44.$$

Среднее время ожидания машины в очереди, по формуле Литтла равно

$$W_{оч} = L_{оч} / \lambda = 1,56 \text{ (мин.)}$$

Среднее время пребывания машины на АЗС

$$W_{сис} = W_{оч} + \frac{Q}{\mu} = 1,56 + \frac{0,703}{0,8} = 2,44 \text{ (мин.)}$$

## 2) Многоканальная СМО с отказами М/М/п (задача Эрланга).

На п-канальную СМО с отказами поступает простейший поток заявок с интенсивностью  $\lambda$ , время обслуживания - показательное с параметром

$\mu = 1/\sqrt{t_{\text{обсл}}}$ . Состояния СМО нумеруются по числу заявок, находящихся в СМО (в силу отсутствия очереди, оно совпадает с числом занятых каналов, см. ДИП на рис. 2.13):

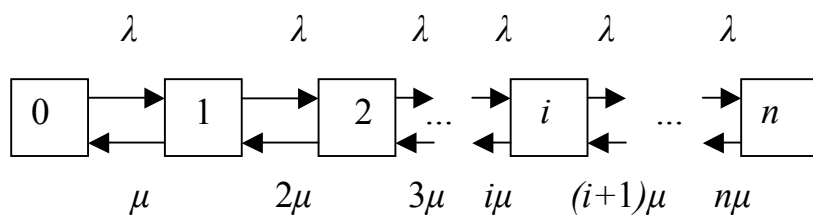


Рис. 2.13. ДИП для системы **M/M/n**

Воспользуемся правилом равенства встречных потоков через сечения диаграммы:

$$\lambda P_0 = \mu P_1, \quad P_1 = \frac{\lambda}{\mu} P_0 = \omega p;$$

$$\lambda P_1 = 2\mu P_2, \quad P_2 = \frac{\lambda}{2\mu} P_1 = \frac{\omega^2}{2} p;$$

$$\lambda P_2 = 3\mu P_3, \quad P_3 = \frac{\lambda}{3\mu} P_2 = \frac{\omega^3}{3!} p;$$

...

По индукции получаем  $P_i = \frac{\omega^i}{i!} p$ .

Финальные вероятности состояний выражаются формулами Эрланга:

$$P_0 = p = \left\{ 1 + \frac{\omega}{1!} + \frac{\omega^2}{2!} + \dots + \frac{\omega^n}{n!} \right\}^{-1}; \quad P_k = \frac{\omega^k}{k!} P_0 \quad (k = 1, 2, \dots, n).$$

**Характеристики эффективности:**

— вероятность отказа (вероятность занятости всех каналов)

$$P_{\text{отк}} = P_n = \frac{\omega^n}{n!} p;$$

— относительная пропускная способность

$$Q = 1 - P_n = 1 - \frac{\omega^n}{n!} p;$$

— абсолютная пропускная способность

$$A = \lambda Q = \lambda \left( 1 - \frac{\omega^n}{n!} p \right);$$

— среднее число занятых каналов можно вычислить непосредственно через вероятности  $P_0, P_1, \dots, P_n$  по формуле:

$$\bar{k} = 0P_0 + 1P_1 + \dots + nP_n,$$

но, учитывая, что абсолютная пропускная способность  $A$  есть не что иное, как среднее число заявок, обслуживаемых системой в единицу времени, а один занятый канал обслуживает за единицу времени в среднем  $\mu$  заявок, получаем:

$$\bar{k} = \frac{A}{\mu} = \omega \left(1 - \frac{\omega^n}{n!} p\right).$$

**Пример.** Имеется двухканальная простейшая СМО с отказами. На ее вход поступает поток заявок с интенсивностью  $\lambda = 4$  заявки/ч. Среднее время обслуживания одной заявки  $\bar{t}_{об} = 0,8$  ч. Каждая обслуженная заявка приносит доход  $c = 4$  руб. Содержание каждого канала обходится 2 руб./ч. Решить: выгодно или невыгодно в экономическом отношении увеличить число каналов СМО до трех?

**Решение.**  $\omega = \lambda/\mu = 3,2$ .

По формулам Эрланга получаем:

$$P_0 = p \left\{ 1 + \frac{3,2}{1!} + \frac{3,2^2}{2!} \right\}^{-1} = (9,32)^{-1} \approx 0,107, \quad P_2 = \frac{3,2^2}{2!} \cdot 0,107 \approx 0,55.$$

Относительная пропускная способность  $Q=1-P_2=0,450$ , абсолютная пропускная способность  $A = \lambda Q = 1,8$  заявки/ч.

Доход от заявок, приносимый СМО в данном варианте, равен  $D = A \cdot c \approx 7,2$  руб/ч.

Подсчитаем те же характеристики для трехканальной СМО (отмечая их штрихом вверху):

$$P'_0 = p' \left\{ 1 + \frac{3,2}{1!} + \frac{3,2^2}{2!} + \frac{3,2^3}{3!} \right\}^{-1} \approx 0,0677; P'_3 = \frac{3,2^3}{3!} \cdot 0,0677 \approx 0,371.$$

$$Q' = 1 - P'_3 \approx 0,629; A' \approx 2,52; D' = A' \cdot c \approx 10,08 \text{ руб./ч.}$$

Увеличение дохода равно  $D' - D = 2,88$  руб./ч; увеличение расхода равно 2 руб./ч; из этого видно, что переход от  $n = 2$  к  $n = 3$  экономически выгоден.

**3) Многоканальная СМО с фиксированным числом мест ожидания (с ограниченной очередью) – М/М/п/м.** На  $n$ -канальную СМО поступает простейший поток заявок с интенсивностью  $\lambda$ , время обслуживания каждого канала показательное с параметром  $\mu = 1/\sqrt{t_{\text{обсл}}}$ .

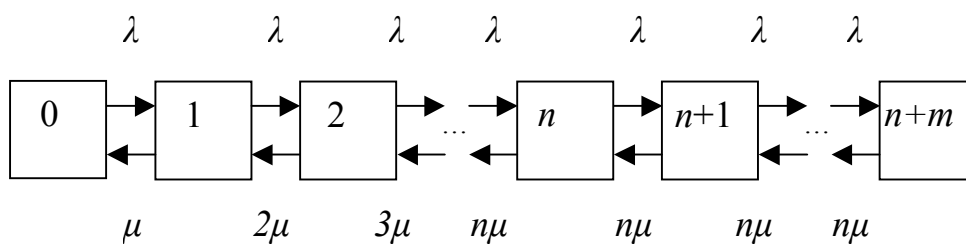


Рис. 2.14. ДИП для системы **М/М/п/м**

Диаграмма на рис. 2.14 для состояний  $0 - n$  полностью совпадает с диаграммой для системы М/М/п с отказами (см. рис. 3.10). Значит, и вероятности для этих состояний тоже совпадут:

$$P_k = \frac{\omega^k}{k!} P_0 \quad (k = 1, 2, \dots, n).$$

Получим вероятности для состояний с номерами больше  $n$ .

$$\lambda P_n = n\mu P_{n+1}, \quad P_{n+1} = \frac{\lambda}{n\mu} P_0 = \frac{\omega}{n} P_n;$$

$$\lambda P_{n+1} = n\mu P_{n+2}, \quad P_{n+2} = \frac{\lambda}{n\mu} P_n = \frac{\omega^2}{n^2} P_n;$$

$$\lambda P_{n+2} = n\mu P_{n+3}, \quad P_{n+3} = \frac{\lambda}{n\mu} P_1 = \frac{\omega^3}{n^3} P_n;$$

.....

По индукции получаем  $P_{n+i} = \frac{\omega^i}{n^i} P_n, \quad i = 1, 2, \dots, m$ .

Суммарная вероятность

$$\sum_{i=1}^m P_{n+i} = \sum_{i=1}^m \frac{\omega^i}{n^i} P_n = P_n \sum_{i=1}^m \frac{\omega^i}{n^i}$$

представляет собой сумму геометрической прогрессии со знаменателем  $\frac{\omega}{n}$  и,

учитывая, что  $P_n = \frac{\omega^n}{n!} p$ , получаем:

$$\sum_{i=1}^m P_{n+i} = \frac{\omega^n p}{n!} \cdot \frac{\frac{\omega}{n} (1 - \frac{\omega^m}{n^m})}{1 - \frac{\omega}{n}} = \frac{\omega^{n+1} (1 - \frac{\omega^m}{n^m})}{n! (n - \omega)} p.$$

Из уравнения нормировки можно найти значение  $P_0 = p$ :

$$p = [1 + \frac{\omega}{1!} + \frac{\omega^2}{2!} + \dots + \frac{\omega^n}{n!} + \frac{\omega^{n+1} (1 - \frac{\omega^m}{n^m})}{n! (n - \omega)}]^{-1}.$$

Вычислив по этой формуле значение  $p$ , можно рассчитать и вероятности всех остальных состояний.

При  $\omega = n$  последний элемент суммы в скобках выражения для  $p$  превращается в неопределенность вида  $0/0$ . Раскрыв эту неопределенность, получим, что значение  $p$  в этом случае равно:

$$p = [1 + \frac{n}{1!} + \frac{n^2}{2!} + \dots + \frac{n^n}{n!} + \frac{mn^n}{n!}]^{-1}.$$

**Характеристики эффективности:**

— *вероятность отказа (вероятность занятости всех каналов и мест в очереди)*



$$P_{отк} = P_{n+m} = \frac{\omega^{n+m}}{n^m n!} p;$$

— *относительная пропускная способность*

$$Q = 1 - P_{отк} = 1 - \frac{\omega^{n+m}}{n^m n!} p;$$

— *абсолютная пропускная способность*

$$A = \lambda Q = \lambda \left(1 - \frac{\omega^{n+m}}{n^m n!} p\right);$$

— *среднее число занятых каналов* можно вычислить исходя из того, что вся система обслуживает за единицу времени в среднем  $A$  заявок, а один канал —  $\mu$ :

$$\bar{k} = \frac{A}{\mu} = \omega \left(1 - \frac{\omega^{n+m}}{n^m n!} p\right).$$

— *среднее число заявок в очереди* можно вычислить непосредственно, как математическое ожидание дискретной случайной величины, умножая возможное значение числа заявок в очереди на вероятность этого значения и складывая результаты:

$$\begin{aligned} L_{оч} &= 1P_{n+1} + 2P_{n+2} + \dots + mP_{n+m} = 1 \frac{\omega^{n+1}}{nn!} p + 2 \frac{\omega^{n+2}}{n^2 n!} p + \dots + m \frac{\omega^{n+m}}{n^m n!} p = \\ &= \frac{\omega^{n+1}}{nn!} p \left[1 + 2 \frac{\omega}{n} + 3 \left(\frac{\omega}{n}\right)^2 + \dots + m \left(\frac{\omega}{n}\right)^{m-1}\right] = \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^m i \left(\frac{\omega}{n}\right)^{i-1}. \end{aligned}$$

Обозначим  $\frac{\omega}{n} = \rho$ . Тогда

$$L_{оч} = \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^m i \rho^{i-1} = \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^m \frac{d}{d\rho} (\rho^i) = \frac{\omega^{n+1}}{nn!} p \frac{d}{d\rho} \sum_{i=1}^m (\rho^i).$$

Учитывая, что сумма в этом выражении является суммой  $m$  элементов геометрической прогрессии, первый член и множитель которой равны  $\rho$ , получаем

$$L_{оч} = \frac{\omega^{n+1}}{nn!} p \frac{d}{d\rho} \left( \frac{\rho(1-\rho^m)}{1-\rho} \right) = \frac{\omega^{n+1} [1 - (m+1)\rho^m + m\rho^{m+1}]}{nn!(1-\rho)^2} p.$$

— *среднее число заявок в системе*

$$L_c = L_{оч} + \bar{k}.$$

— *среднее время пребывания заявки в очереди* можно получить по формуле Литтла:

$$W_{оч} = L_{оч} / \lambda.$$

— *среднее время пребывания заявки в системе* так же, как и для одноканальной системы, отличается от среднего времени ожидания на среднее время обслуживания, умноженное на относительную пропускную способность:

$$W_c = W_{оч} + \frac{Q}{\mu}.$$

При  $\frac{\omega}{n} = \rho = 1$  раскрытие неопределенности в формуле для  $L_{оч}$  приводит к следующему результату:

$$L_{оч} = \frac{\omega^{n+1} m(m+1)}{2nn!} p.$$

**Пример.** Автозаправочная станция (АЗС) с двумя колонками ( $n = 2$ ) предназначена для обслуживания машин. Поток машин, прибывающих на АЗС, имеет интенсивность  $\lambda = 2$  (машины в минуту); среднее время обслуживания одной машины

$$\bar{t}_{об} = \frac{1}{\mu} = 2 \text{ (мин)}.$$

Площадка у АЗС может вместить очередь не более  $m = 3$  (машин). Машина, прибывшая в момент, когда все три места в очереди заняты, покидает АЗС (получает отказ). Найти характеристики СМО:

- вероятность отказа,
- относительную и абсолютную пропускную способности,
- среднее число занятых колонок,
- среднее число машин в очереди,
- среднее время ожидания и пребывания машины на АЗС.

**Решение.** Имеем:  $n = 2$ ,  $m = 3$ ,  $\lambda = 2$ ,  $\mu = 0,5$ ,  $\omega = 4$ ,  $\rho = \omega/n = 2$ .

По формулам для СМО **М/М/п/п** находим:

$$p = \left[ 1 + \frac{4}{1!} + \frac{4^2}{2!} + \frac{4^3(1 - \frac{4^3}{2^3})}{2!(2-4)} \right]^{-1} \frac{1}{125} = 0,008.$$

Вероятность отказа:

$$P_{отк} = P_{n+m} = \frac{4^5}{2^3 2!} p = 64 \cdot 0,008 = 0,512.$$

Относительная пропускная способность:

$$Q = 1 - P_{отк} = 0,488.$$

Абсолютная пропускная способность:

$$A = \lambda Q = 0,976 \text{ (машины в минуту)}.$$

Среднее число занятых каналов (колонок):

$$\bar{k} = \frac{A}{\mu} = \frac{0,976}{0,5} = 1,952$$

(т. е. обе колонки почти все время заняты).

Среднее число машин в очереди:

$$L_{оч} = \frac{4^3 [1 - (3+1) \cdot 2^3 + 3 \cdot 2^4]}{2 \cdot 2! (1-2)^2} \cdot 0,008 = 2,18.$$

Среднее время пребывания в очереди:

$$W_{оч} = L_{оч} / \lambda = \frac{2,18}{2} = 1,09 \text{ (мин)}.$$

Среднее время пребывания машины на АЗС (включая время обслуживания)

$$W_c = W_{оч} + \frac{Q}{\mu} = 1,09 + 0,976 = 2,07 \text{ (мин)}.$$

#### 4) Одноканальная СМО с неограниченной очередью – М/М/1/∞.

Входной поток заявок простейший с интенсивностью  $\lambda$ , поток обслуживаний простейший с интенсивностью  $\mu$ , количество мест ожидания не ограничено (см. ДИП на рис. 2.15).

Финальные вероятности существуют и могут быть определены, если интенсивность поступления заявок ниже интенсивности их обслуживания (т.е.  $\omega = \frac{\lambda}{\mu} < 1$ ). При  $\omega=1$  очередь будет бесконечно расти.

Воспользовавшись правилом равенства встречных потоков вероятностей через сечение диаграммы и, используя, как и ранее, обозначения  $P_0 = p$  и  $\lambda/\mu = \omega$ , получим:

$$\begin{aligned} \lambda P_0 &= \mu P_1, & P_1 &= (\lambda/\mu)P_0 = \omega p; \\ \lambda P_1 &= \mu P_2, & P_2 &= (\lambda/\mu)P_1 = \omega^2 p; \\ \lambda P_2 &= \mu P_3, & P_3 &= (\lambda/\mu)P_2 = \omega^3 p; \end{aligned}$$

...

По индукции получаем  $P_i = \omega^i p$ .

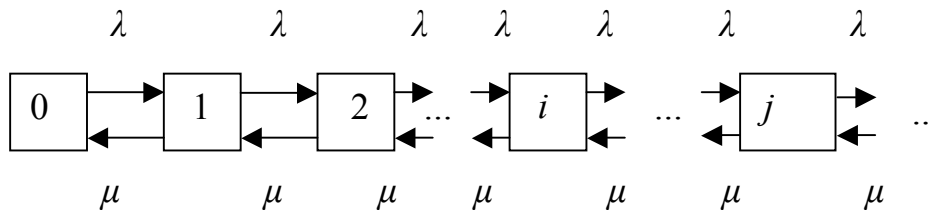


Рис. 2.15. ДИП для системы М/М/1/∞

Для определения значения  $P_0 = p$  воспользуемся нормировочным уравнением:

$$\sum_{i=0}^{\infty} P_i = 1, \text{ или } \sum_{i=0}^{\omega} \omega^i p = p \sum_{i=0}^{\infty} \omega^i = 1.$$

Учитывая, что полученная сумма представляет собой сумму бесконечной геометрической прогрессии, получим:

$$\frac{p}{1-\omega} = 1.$$

Окончательно:  $p = 1 - \omega$ .

**Характеристики эффективности:**

При отсутствии ограничений по длине очереди каждая заявка, пришедшая в систему, будет обслужена, поэтому

— *относительная пропускная способность*

$$Q = 1;$$

— *абсолютная пропускная способность*

$$A = \lambda Q = \lambda;$$

— *среднее число занятых каналов (вероятность того, что канал занят)*

$$\bar{k} = 1 - P_0 = \omega;$$

— *среднее число заявок в очереди*

$$L_{оч} = \sum_{i=1}^{\infty} i P_{i+1} = \sum_{i=1}^{\infty} i \omega^{i+1} p = \omega^2 p \sum_{i=1}^{\infty} i \omega^{i-1} = \omega^2 p \sum_{i=1}^{\infty} \frac{d}{d\omega} (\omega^i) = \omega^2 p \frac{d}{d\omega} \sum_{i=1}^{\infty} \omega^i.$$

Сумма в этом выражении представляет собой сумму бесконечной геометрической прогрессии и равна  $\frac{\omega}{1-\omega}$ .

Производная от этого выражения  $\frac{1}{(1-\omega)^2}$ .

Учитывая выражение для  $p$ , окончательно имеем:

$$L_{оч} = \frac{\omega^2}{1-\omega}.$$

— *среднее число заявок в СМО*

$$L_c = L_{оч} + \bar{k} = \frac{\omega^2}{1-\omega} + \omega = \frac{\omega}{1-\omega}.$$

По формуле Литтла можно получить значение *среднего времени пребывания заявки в очереди*

$$W_{оч} = L_{оч} / \lambda$$

и *среднего времени пребывания заявки в системе*

$$W_c = L_c / \lambda.$$

**5) Многоканальная СМО с неограниченной очередью) – М/М/п/∞.** На  $n$ -канальную СМО поступает простейший поток заявок с интенсивностью  $\lambda$ , время обслуживания каждого канала показательное с параметром  $\mu = 1 / \sqrt{t_{обсл}}$ .

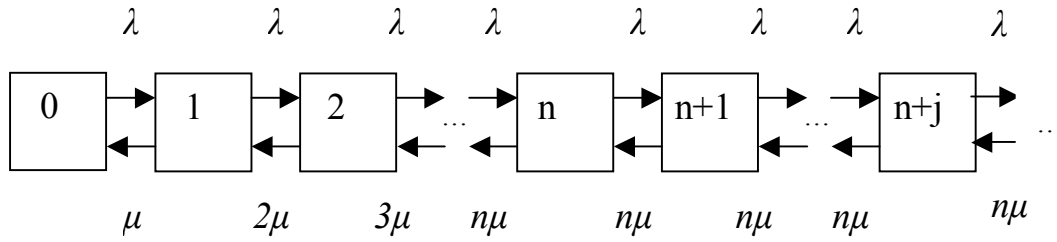


Рис. 2.16. ДИП для системы  $M/M/n/\infty$

Диаграмма для этой системы ( рис. 2.16 ) отличается диаграммы для многоканальной системы с ограниченной очередью  $M/M/n/m$  (см. рис. 2.13) только бесконечным числом состояний. Значит, и вероятности для этих состояний тоже совпадут:

$$P_k = \frac{\omega^k}{k!} P_0 \quad (k = 1, 2, \dots, n),$$

$$P_{n+i} = \frac{\omega^i}{n^i} P_n, \quad i = 1, 2, \dots, m.$$

Суммарная вероятность для всех состояний с номерами больше  $n$ :

$$\sum_{i=1}^{\infty} P_{n+i} = \sum_{i=1}^{\infty} \frac{\omega^i}{n^i} P_n = P_n \sum_{i=1}^{\infty} \frac{\omega^i}{n^i}$$

представляет собой сумму бесконечной геометрической прогрессии со знаменателем  $\frac{\omega}{n}$  и, учитывая, что  $P_n = \frac{\omega^n}{n!} p$ , получаем:

$$\sum_{i=1}^m P_{n+i} = \frac{\omega^n p}{n!} \cdot \frac{\frac{\omega}{n}}{1 - \frac{\omega}{n}} = \frac{\omega^{n+1}}{n!(n-\omega)} p.$$

Из уравнения нормировки можно найти значение  $P_0 = p$ :

$$p = \left[ 1 + \frac{\omega}{1!} + \frac{\omega^2}{2!} + \dots + \frac{\omega^n}{n!} + \frac{\omega^{n+1}}{n!(n-\omega)} \right]^{-1}.$$

### Характеристики эффективности:

Как и для одноканальной СМО при отсутствии ограничений по длине очереди каждая заявка, пришедшая в систему, будет обслужена, поэтому

— относительная пропускная способность

$$Q = 1;$$

— абсолютная пропускная способность

$$A = \lambda Q = \lambda;$$

— среднее число занятых каналов определим следующим образом. Так как система работает в стационарном режиме, все заявки, поступившие в систему, в конечном итоге будут обслужены. В единицу времени поступает в среднем  $\lambda$  заявок, один канал в единицу времени обрабатывает в среднем  $\mu$  заявок, следовательно, для того, чтобы были обслужены все поступившие заявки, требуется

$$\bar{k} = \frac{\lambda}{\mu} = \omega$$

каналов.

— среднее число заявок в очереди вычислим как математическое ожидание дискретной случайной величины, умножая возможное значение числа заявок в очереди на вероятность этого значения и складывая результаты:

$$\begin{aligned} L_{oc} &= 1P_{n+1} + 2P_{n+2} + \dots + iP_{n+i} + \dots = \sum_{i=1}^{\infty} iP_{n+i} = \sum_{i=1}^{\infty} i \frac{\omega^i}{n^i} P_n = \sum_{i=1}^{\infty} i \frac{\omega^i \omega^n}{n^i n!} p = \\ &= \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^{\infty} i \left(\frac{\omega}{n}\right)^{i-1}. \end{aligned}$$

Используя обозначение  $\frac{\omega}{n} = \rho$ , получим

$$L_{oc} = \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^{\infty} i \rho^{i-1} = \frac{\omega^{n+1}}{nn!} p \sum_{i=1}^{\infty} \frac{d}{d\rho} (\rho^i) = \frac{\omega^{n+1}}{nn!} p \frac{d}{d\rho} \sum_{i=1}^{\infty} (\rho^i).$$

Учитывая, что сумма в этом выражении является суммой  $m$  элементов геометрической прогрессии, первый член и множитель которой равны  $\rho$ , получаем

$$L_{oc} = \frac{\omega^{n+1}}{nn!} p \frac{d}{d\rho} \left( \frac{\rho}{1-\rho} \right) = \frac{\omega^{n+1}}{nn!(1-\rho)^2} p.$$

— среднее число заявок в системе

$$L_c = L_{oc} + \bar{k}.$$

— среднее время пребывания заявки в очереди и среднее время пребывания заявки в системе и можно получить по формулам Литтла:

$$W_{oc} = L_{oc} / \lambda, \quad W_c = L_c / \lambda.$$

**Пример.** Автозаправочная станция с двумя колонками ( $n = 2$ ) обслуживает поток машин с интенсивностью  $\lambda = 0,8$  (машин в минуту). Среднее время обслуживания одной машины  $\bar{t}_{об} = \frac{1}{\mu} = 2$  (мин).

В данном районе нет другой АЗС, так что очередь машин перед АЗС может расти практически неограниченно. Найти характеристики СМО.

**Решение** Имеем:  $n = 2$ ,  $m = 3$ ,  $\lambda = 2$ ,  $\mu = 0,5$ ,  $\omega = 1,6$ ,  $\rho = \omega/n = 0,8$ . Поскольку  $\rho < 1$ , очередь не растёт безгранично и имеет смысл говорить о предельном стационарном режиме работы СМО. По формулам для системы **M/M/n/∞** находим :

$$p = \left[ 1 + \frac{1,6}{1!} + \frac{1,6^2}{2!} + \frac{1,6^3}{2!(2-1,6)} \right]^{-1} \approx 0,111.$$

Среднее число занятых каналов:

$$\bar{k} = \frac{\lambda}{\mu} = 1,6.$$

Вероятность отсутствия очереди у АЗС будет:

Среднее число машин в очереди:

$$L_{оч} = \frac{1,6^3}{2 \cdot 2! \cdot (1 - 0,8)^2} \cdot 0,111 \approx 2,84.$$

Среднее число машин на АЗС:

$$L_c = L_{оч} + \bar{k} \approx 2,84 + 1,6 = 4,44.$$

Среднее время ожидания в очереди:

$$W_{оч} = L_{оч} / \lambda = 3,55 \text{ (мин)}.$$

Среднее время пребывания машины на АЗС:

$$W_c = L_c / \lambda = 5,55 \text{ (мин)}.$$

**6. Простейшая многофазовая СМО с очередью.** Анализ многофазовых СМО в общем случае затруднен, тем что входящий поток каждой последующей фазы является выходным потоком предыдущей и в общем случае имеет последствие. Однако если на вход СМО с неограниченной очередью поступает простейший поток заявок, а время обслуживания показательное, то выходной поток этой СМО — простейший, с той же интенсивностью  $\lambda$ , что и входящий. Из этого следует, что многофазовую СМО с неограниченной очередью перед каждой фазой, простейшим входящим потоком заявок и показательным временем обслуживания на каждой фазе можно анализировать как простую последовательность простейших СМО,

Если очередь к фазе ограничена, то выходной поток этой фазы перестает быть простейшим и вышеуказанный прием может применяться только в качестве приближенного.

### 2.7.7. Замкнутые системы массового обслуживания (СМО с ожиданием ответа)

До сих пор мы рассматривали такие системы массового обслуживания, где заявки приходили откуда-то извне и интенсивность потока заявок не зависела от состояния самой системы. Сейчас 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.

Примером замкнутой СМО может служить, например, следующая система. Рабочий-наладчик обслуживает  $m$  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  $t_{\text{обсл}}=1/\mu$ , 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.

Иными словами, мы имеем СМО с  $m$  источниками, каждый из которых может выдать заявку и после этого ожидает окончания обслуживания этой заявки. Если в приведенном примере станки обслуживаются бригадой из  $n$  наладчиков, то СМО становится многоканальной.

Другим примером может быть система с центральным процессором и удаленными терминалами. Пользователь, отправивший запрос с терминала не может выдать новый запрос, пока не получит сообщения от процессора об окончании обработки предыдущего.

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, в сущности, 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.

Построим аналитическую модель такой системы.

Сначала рассмотрим случай с одним обслуживающим прибором и количеством источников равным  $m$ . Состояния будем кодировать числом выданных заявок. Диаграмма интенсивностей переходов для данной



системы изображена на рис.2.17.

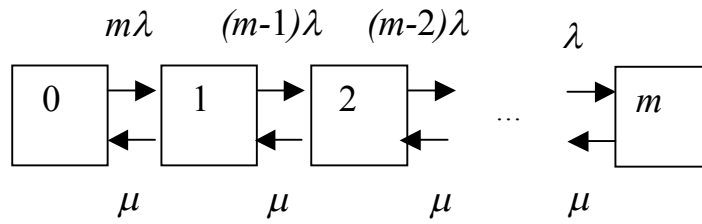


Рис. 2.17. ДИП для одноканальной замкнутой СМО

Определим вероятности состояний.

$$P_0 m \lambda = P_1 \mu; \quad P_1 = \frac{m \lambda}{\mu} P_0 = m \omega p;$$

$$P_1 (m-1) \lambda = P_2 \mu; \quad P_2 = \frac{(m-1) \lambda}{\mu} P_1 = m(m-1) \omega^2 p;$$

...

$$P_i = m(m-1) \dots (m-i+1) \omega^i p;$$

...

$$P_m = m(m-1) \dots 1 \cdot \omega^m p;$$

Исходя из уравнения нормировки, получаем

$$p = \frac{1}{1 + m \omega + m(m-1) \omega^2 + m(m-1)(m-2) \omega^3 + \dots + m(m-1) \dots 1 \omega^m}.$$

Найдем характеристики эффективности замкнутой СМО.

Абсолютная пропускная способность – это среднее количество заявок, обслуживаемых каналом в единицу времени. Вычислим эту характеристику. Канал занят обслуживанием заявок с вероятностью

$$P_{зан} = 1 - P_0 = 1 - p.$$

Если он занят, то обслуживает в среднем  $\mu$  заявок в единицу времени. Таким образом, абсолютная пропускная способность системы

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

Так как каждая заявка, в конце концов, будет обслужена, то относительная пропускная способность  $q = 1$ .

Вычислим среднее число заявок, ожидающих обслуживания, иначе — среднее число источников, выдавших заявку и ожидающих ответа. По сути, это количество заявок, находящихся в данный момент в системе  $L_c$ . Вообще говоря, эту величину можно вычислить непосредственно, по формуле

$$L_c = 1 \cdot P_1 + 2 \cdot P_2 + \dots + m \cdot P_m,$$

но проще будет найти ее через абсолютную пропускную способность  $A$ .

Каждый источник, еще не выдавший заявку, порождает поток заявок с интенсивностью  $\lambda$ . Таких источников  $m - L_c$ , а поток, порождаемый ими имеет интенсивность  $(m - L_c) \lambda$ . Все эти заявки будут обслужены каналом,

следовательно,

$$(m - L_c)\lambda = (1-p)\mu,$$

откуда

$$L_c = m - \frac{\mu}{\lambda}(1-p) = m - \frac{1-p}{\omega}.$$

Определим теперь среднее число источников, выдавших заявки, обслуживание которых еще не началось. Фактически, это количество заявок в очереди  $L_{оч}$ .

Среднее число заявок в системе  $L_c$  складывается из числа заявок в очереди  $L_{оч}$  и среднего числа заявок, находящихся под обслуживанием в канале  $\bar{k}$ :

$$L_c = L_{оч} + \bar{k}.$$

В канале может находиться 0 заявок с вероятностью  $p$  или 1 с вероятностью  $1-p$ , следовательно,

$$\bar{k} = 0 \cdot p + 1 \cdot (1-p) = 1-p.$$

Отсюда

$$L_{оч} = m - \frac{1-p}{\omega} - (1-p) = m - (1-p)\left(1 + \frac{1}{\omega}\right).$$

Теперь перейдем к случаю с несколькими каналами обслуживания. Будем кодировать состояния общим числом выданных источниками и еще не обслуженных заявок. Так как источник не может выдать новую заявку до окончания обслуживания предыдущей, то интенсивность общего потока заявок зависит от того, сколько заявок связано с процессом обслуживания (непосредственно обслуживается или стоит в очереди). Количество источников –  $m$ , число каналов –  $n$  ( $n < m$ ). Диаграмма интенсивностей переходов для данной системы изображена на рис. 2.18.

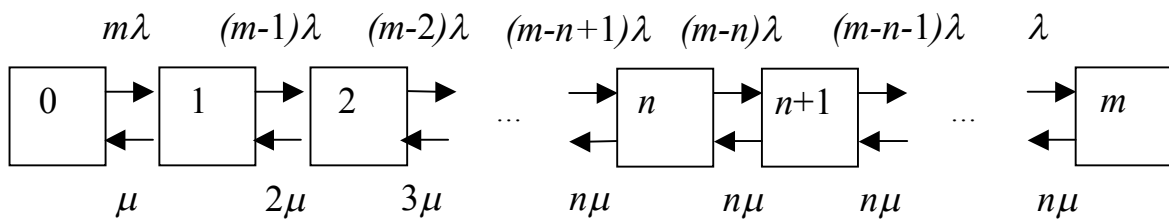


Рис. 2.18. ДИП для многоканальной замкнутой СМО

Требуется найти вероятности состояний данной системы и ее характеристики.

$$P_0 m \lambda = P_1 \mu; \quad P_1 = \frac{m \lambda}{\mu} P_0 = m \omega p;$$

$$P_1(m-1)\lambda = P_2 2\mu; \quad P_2 = \frac{(m-1)\lambda}{2\mu} P_1 = \frac{m(m-1)}{2} \omega^2 p;$$

$$P_2(m-2)\lambda = P_3 3\mu; \quad P_3 = \frac{(m-2)\lambda}{3\mu} P_2 = \frac{m(m-1)(m-2)}{3!} \omega^3 p;$$

$$P_i = \frac{m(m-1)(m-2)\dots(m-i+1)}{i!} \omega^i p, \quad i=1, 2, \dots, n.$$

$$P_n = \frac{m(m-1)(m-2)\dots(m-n+1)}{n!} \omega^n p.$$

При  $i > n$  возникает ситуация, когда  $n$  заявок обслуживаются, а  $i-n$  ожидают обслуживания.

$$P_n(m-n)\lambda = P_{n+1} n\mu; \quad P_{n+1} = \frac{(m-n)\lambda}{n\mu} P_n = \frac{(m-n)}{n} \omega P_n;$$

$$P_{n+1}(m-n-1)\lambda = P_{n+2} n\mu; \quad P_{n+2} = \frac{(m-n-1)\lambda}{n\mu} P_{n+1} = \frac{(m-n)(m-n-1)}{n^2} \omega^2 P_n;$$

$$P_{n+i} = \frac{(m-n)(m-n-1)\dots(m-n-i+1)}{n^i} \omega^i P_n;$$

$$P_m = \frac{(m-n)(m-n-1)\dots 2 \cdot 1}{n^{m-n}} \omega^{m-n} P_n.$$

Подставляя в выражения для  $P_{n+i}$  и  $P_m$  полученное выше значение  $P_n$ , получаем

$$P_{n+i} = \frac{m(m-1)\dots(m-n-i+1)}{n! n^i} \omega^{n+i} p;$$

$$P_m = \frac{m!}{n! n^{m-n}} \omega^m p.$$

Исходя из уравнения нормировки, определим значение  $p$ :

$$p = \left[ 1 + \frac{m}{1!} \omega + \frac{m(m-1)}{2!} \omega^2 + \dots + \frac{m(m-1)(m-2)\dots(m-n+1)}{n!} \omega^n + \right. \\ \left. + \frac{m(m-1)\dots(m-n)}{n! n} \omega^{n+1} + \dots + \frac{m!}{n! n^{m-n}} \omega^m \right]^{-1}.$$

Через вычисленные вероятности может быть определено среднее число занятых каналов:

$$\bar{k} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 + \dots + n \cdot (P_n + P_{n+1} + \dots P_m) = \\ = P_1 + 2P_2 + \dots + (n-1)P_{n-1} + n(1 - P_0 - P_1 - \dots P_{n-1}).$$

Абсолютная пропускная способность системы

$$A = \bar{k}\mu.$$

**Пример.** Рабочий обслуживает группу из трех станков. Каждый станок останавливается в среднем 2 раза в час. Процесс наладки занимает у рабочего, в среднем, 10 минут. Определить характеристики замкнутой СМО: вероятность занятости рабочего; его абсолютную пропускную способность  $A$ ; среднее количество неисправных станков  $L_c$ . Все потоки полагаем простейшими.

**Решение.** Имеем  $n = 3$ ,  $\lambda = 2$ ,  $\mu = \frac{1}{1/6} = 6$ ,  $\omega = \lambda / \mu = 1/3$ .

Определяем по формулам для одноканальной замкнутой СМО

$$p = \frac{1}{1 + 3 \cdot \left(\frac{1}{3}\right) + 3 \cdot 2 \cdot \left(\frac{1}{3}\right)^2 + 3 \cdot 2 \cdot 1 \cdot \left(\frac{1}{3}\right)^3} \approx 0,346.$$

Вероятность занятости рабочего:

$$P_{\text{зан}} = 1 - P_0 = 1 - p = 0,654.$$

Абсолютная пропускная способность (среднее число неисправностей, которое рабочий ликвидирует в час):

$$A = (1 - p)\mu = 0,654 \cdot 6 = 3,94.$$

Среднее число неисправных станков:

$$L_c = m - \frac{1 - p}{\omega} = 3 - \frac{0,654}{1/3} = 1,04.$$

### 2.7.8. Распределение Эрланга. Метод этапов

На практике марковские процессы в чистом виде почти никогда не встречаются: реальные процессы почти всегда обладают тем или другим последствием. Для марковского процесса время пребывания системы подряд в каком-либо состоянии распределено по показательному закону; на самом деле это далеко не всегда бывает так. Например, если поток событий, переводящий систему из состояния в состояние является потоком отказов какого-то узла, то более естественно предположить, что оставшееся время безотказной работы узла зависит от того, сколько времени узел уже работал. При этом время пребывания узла в рабочем состоянии представляет собой случайную величину, распределенную не по показательному, а по какому-то иному закону. Возникает вопрос о том, можно ли приближенно заменять непуассоновские потоки — пуассоновскими и к каким ошибкам в предельных вероятностях состояний может привести подобная замена. Для этого необходимо уметь

хотя бы приближенно исследовать случайные процессы, протекающие в системах с последствием.

Рассмотрим некоторую физическую систему, в которой протекает случайный процесс, направляемый какими-то непугассоновскими потоками событий. Если мы попробуем для этого процесса написать уравнения, выражающие вероятности состояний как функции времени, мы увидим, что в общем случае это нам не удастся. Действительно, для марковской системы мы вычисляли вероятность того, что в момент  $t + \Delta t$  система будет в состоянии  $S_i$  учитывая только то, в каком состоянии система была в момент  $t$ , и не учитывая, сколько времени она была в этом состоянии. Для немарковской системы этот прием уже непригоден: вычисляя вероятность перехода из одного состояния в другое за время  $\Delta t$ , мы должны будем учитывать, сколько времени система уже провела в данном состоянии. Это приводит, вместо обыкновенных дифференциальных, уравнений, к уравнениям с частными производными, то есть к гораздо более сложному математическому аппарату, с помощью которого только в редких случаях можно получить нужные результаты.

Однако, в некоторых случаях можно немарковский процесс свести искусственно (хотя бы приближенно) к марковскому – если число состояний системы не очень велико, а отличающиеся от простейших потоки событий, участвующие в задаче, представляют собой (точно или приближенно) потоки Эрланга. Тогда, вводя в схему возможных состояний системы некоторые фиктивные «псевдосостояния» (этапы), удастся свести немарковский процесс к марковскому и описать его с помощью алгебраических уравнений для предельных вероятностей состояний.

Для начала выясним, что представляет собой поток Эрланга.

Рассмотрим простейший поток с интенсивностью  $\lambda$ . Просеим этот поток, оставив в нем каждое  $k$ -е событие и удалив все остальные. В результате образуется новый поток событий, который называется *потоком Эрланга  $k$ -го порядка*.

Интервалы времени между событиями в таком потоке представляют собой сумму  $k$  интервалов с показательным распределением

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

и распределены по закону Эрланга  $k$ -го порядка:

$$f_k(t) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}.$$

Математическое ожидание, дисперсия и среднее квадратическое отклонение в потоке Эрланга  $k$ -го порядка:

$$m_k = \frac{k}{\lambda}, \quad D_k = \frac{k}{\lambda^2}, \quad \sigma_k = \frac{\sqrt{k}}{\lambda}.$$

Заметим, что как закон распределения  $f_k(t)$ , так и все его характеристики выражены через интенсивность  $\lambda$  порождающего простейшего потока, а не через интенсивность самого потока Эрланга  $\Lambda_k = \lambda/k$ .

Подставив вместо  $\lambda$  в выражения для закона распределения и его характеристик  $k = \Lambda_k$ , получим:

$$f_k(t) = \frac{k\Lambda_k (k\Lambda_k t)^{k-1}}{(k-1)!} e^{-k\Lambda_k t},$$

$$m_k = \frac{1}{\Lambda_k}, \quad D_k = \frac{1}{k\Lambda_k} \cdot \frac{1}{2}, \quad \sigma_k = \frac{1}{\sqrt{k}\Lambda_k}.$$

Предположим теперь, что мы будем менять порядок закона Эрланга  $k$ , сохраняя неизменной интенсивность потока  $\Lambda_k = \Lambda = \text{const}$ . Его математическое ожидание будет при этом оставаться постоянным, а дисперсия и среднее квадратическое отклонение будут меняться:

$$m = \frac{1}{\Lambda}, \quad D_k = \frac{1}{k\Lambda} \cdot \frac{1}{2}, \quad \sigma_k = \frac{1}{\sqrt{k}\Lambda}.$$

Таким образом мы приходим к так называемому *нормированному* потоку Эрланга. Видно, что при  $k \rightarrow \infty$  и дисперсия и среднее квадратическое отклонение стремятся к 0. Следовательно, при этом поток Эрланга неограниченно приближается к регулярному потоку с постоянным интервалом между событиями:

$$T = \frac{1}{\Lambda}.$$

Это свойство потоков Эрланга удобно в практических целях: оно дает возможность, задаваясь различными  $k$ , получать потоки, обладающие различным последствием – от полного отсутствия последствия (при  $k=1$  поток Эрланга превращается в простейший поток) до жесткой функциональной связи между моментами появления событий ( $k = \infty$ ).

Используя это обстоятельство, можно аппроксимировать полученные эмпирическим путём распределения распределениями Эрланга. Это делают, согласовывая характеристики реального потока (математическое ожидание и дисперсию интервала времени между событиями) с теми же характеристиками заменяющего потока Эрланга.

Например, если проведены измерения интервалов между событиями в потоке обслуживания или поступления заявок, и полученные данные достаточны только для оценок среднего значения  $\tilde{m}$  и дисперсии  $\tilde{D}$ , то можно подобрать одно распределение Эрланга с такими же значениями математического ожидания и дисперсии. Для этого определяют значения  $k$  и  $\Lambda$ :

$$\Lambda = 1/\tilde{m}; \quad k = \tilde{m}^2/\tilde{D}.$$

**Пример.** В результате статистической обработки интервалов времени между событиями в некотором потоке получены следующие характеристики:

- среднее значение интервала  $\tilde{m}=2$  мин,
- среднее квадратическое отклонение интервала  $\tilde{\sigma}=0,9$  мин.

Надо подобрать поток Эрланга, обладающий приблизительно теми же характеристики.

**Решение.**

$$\Lambda = 1/\tilde{m} = 1/2 = 0,5 (\text{соб} / \text{мин}); \quad k = \tilde{m}^2 / \tilde{D} = 4 / 0,9^2 \approx 4,9.$$

Выбираем в качестве  $k$  ближайшее целое число  $k=5$ .

Таким образом, данный поток можно приближенно заменить потоком Эрланга 5-го порядка с плотностью вида

$$f_5(t) = \frac{5 \cdot 0,5 (5 \cdot 0,5 t)^4}{4!} e^{-5 \cdot 0,5 t}, \text{ или}$$

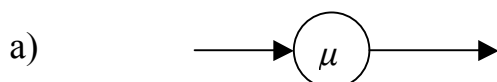
$$f_5(t) = 4,1 t^4 e^{-2,5 t}.$$

Простая замена реального немарковского потока на поток Эрланга еще не дает возможности использовать математический аппарат моделирования марковских процессов, рассмотренный нами ранее, так как поток Эрланга в большей или меньшей степени обладает последствием. Для того, чтобы немарковский процесс свести к марковскому применяют так называемый *метод этапов (метод псевдосостояний)*. Основой метода этапов является свойство отсутствия последствия у показательного распределения. Рассмотрим суть этого метода.

Пусть имеется устройство (рис. 2.17 а), процесс обслуживания заявок в котором определяется потоком событий, интервалы между которыми имеют распределение, отличающееся от показательного. Были определены характеристики этого потока: среднее значение интервала  $\tilde{m} = \frac{1}{\mu}$  и среднее

квадратическое отклонение интервала  $\tilde{\sigma}$ . В соответствии с приведенной выше методикой были выбраны параметры  $k$  и  $\Lambda$  распределения Эрланга, с помощью которого аппроксимируется реальное распределение. Заменим наше устройство другим (рис. 2.19 б), в котором процесс обслуживания заявок осуществляется поэтапно, то есть для полного обслуживания каждая заявка должна пройти  $k$  этапов, на каждом из которых интервал времени обслуживания имеет показательное распределение со средним временем  $\bar{t} = \frac{1}{k\mu}$ . При этом очередная

заявка не может поступить в прибор, пока предыдущая не покинет его, пройдя все  $k$  этапов обслуживания.



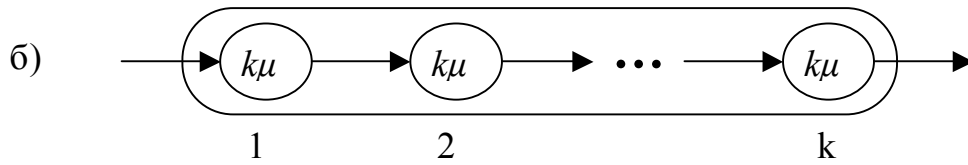


Рис. 2.19 Метод этапов

Полное время обслуживания заявки представляет собой сумму  $k$  показательного распределенных интервалов, и, следовательно, распределено по закону Эрланга порядка  $k$  со средним значением

$$\frac{1}{\Lambda} = \frac{k}{k\mu} = \frac{1}{\mu} = \tilde{m}.$$

Значение  $k$  выбиралось таким образом, чтобы среднее квадратическое отклонение интервала распределения Эрланга равнялось  $\tilde{\sigma}$ . Следовательно, характеристики процессов в обоих этих устройствах совпадают. Но, и это главное, в «поэтапном» устройстве все процессы марковские и для их исследования можно использовать все рассмотренные ранее методы. Правда, кодировать состояния просто числом заявок, находящихся в данный момент в системе теперь нельзя, так как состояния будут зависеть от того, на каком из этапов обслуживания находится в этот момент процесс.

Поясним идею метода этапов на конкретном примере.

**Пример.** Рассматривается система  $S$  — техническое устройство, которое может выходить из строя под влиянием простейшего потока неисправностей с интенсивностью  $\lambda$ . Отказавшее устройство немедленно начинает восстанавливаться. Время восстановления (ремонта)  $T$  распределено не по показательному закону (как надо было бы для того, чтобы процесс был марковским), а по закону Эрланга 3-го порядка:

$$f_3(t) = \frac{\mu(\mu t)^2}{2} e^{-\mu t}$$

Требуется свести данный немарковский процесс к марковскому и найти для него предельные вероятности состояний.

**Решение.** Случайная величина  $T$  — время восстановления — распределена по закону Эрланга и, значит, представляет собой сумму трех случайных величин  $T_1, T_2, T_3$ , распределенных по показательному закону с параметром  $\mu$ :

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

Истинных состояний системы всего два:

$S_1$  — устройство исправно;

$S_2$  — устройство восстанавливается.

Граф этих состояний показан на рис. 2.20а.

Однако в виду того, что переход по стрелке  $S_2 \rightarrow S_1$  происходит под влиянием не простейшего, а эрланговского потока событий, процесс,



происходящий в системе, марковским не является, и для него мы не можем написать алгебраические уравнения.

Чтобы искусственно свести этот процесс к марковскому, введем в цепочку состояний, вместо одного состояния  $S_2$ , три последовательных «псевдосостояния»:

$S_{21}$  — ремонт начинается;

$S_{22}$  — ремонт продолжается;

$S_{23}$  — ремонт заканчивается,

т. е. разделим ремонт на три этапа, причем время пребывания системы в каждой из фаз будем считать распределенным по показательному закону. Граф состояний будет иметь вид, показанный на рис.2.20б, где роль одного состояния  $S_2$  будут играть три псевдосостояния  $S_{21}$ ,  $S_{22}$  и  $S_{23}$ . Процесс, протекающий в такой системе, уже будет марковским.

Обозначим  $P_{21}$ ,  $P_{22}$  и  $P_{23}$  — предельные вероятности пребывания системы в псевдосостояниях  $S_{21}$ ,  $S_{22}$  и  $S_{23}$ ; тогда

$$P_2 = P_{21} + P_{22} + P_{23}.$$

Среднее время пребывания процесса ремонта для каждой из фаз одинаково, следовательно:

$$P_{21} = P_{22} = P_{23}.$$

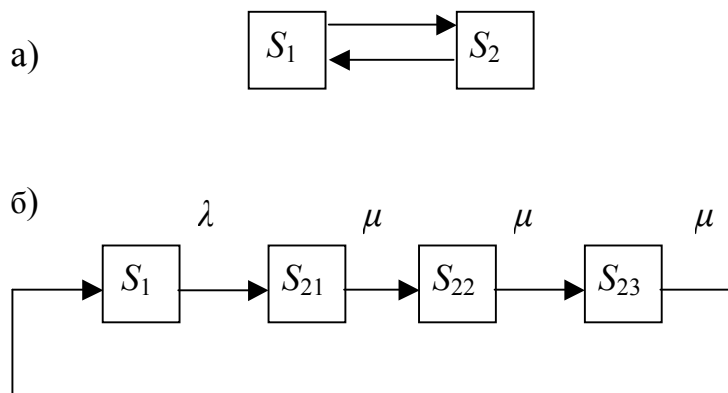


Рис 2.20. Пример использования метода псевдосостояний

Запишем уравнение для предельной вероятности состояния  $P_1$ :

$$\lambda P_1 = \mu P_{23},$$

$$P_1 = \frac{\mu}{\lambda} P_{23}.$$

Нормировочное уравнение:

$$P_1 + P_{21} + P_{22} + P_{23} = 1;$$

$$\frac{\mu}{\lambda} P_{23} + 3P_{23} = 1;$$

$$P_{23} = \frac{\lambda}{3\lambda + \mu}.$$

Отсюда:

$$P_1 = \frac{\mu}{\lambda} P_{23} = \frac{\mu}{3\lambda + \mu}, \quad P_2 = 3P_{23} = \frac{3\lambda}{3\lambda + \mu}.$$

Заметим, что метод псевдосостояний допускает сравнительно простое решение задачи только в самых простых случаях, когда число состояний исходной системы невелико. Однако, иногда удастся применить этот метод и к задачам, где число состояний не очень мало; во всяком случае, получить если не буквенное, то численное приближенное решение соответствующей системы линейных алгебраических уравнений. Рассмотрим в связи с этим процедуру построения модели с неограниченной очередью при аппроксимации потока обслуживания или входного потока потоком Эрланга.

**1. Система М/Е<sub>k</sub>/1.** Входной поток – простейший с интенсивностью  $\lambda$ , поток обслуживания – поток Эрланга порядка  $k$  с интенсивностью  $\Lambda = \frac{\mu}{k}$  ( $\mu$  – интенсивность порождающего простейшего потока). Очередь не ограничена. Структура этой системы представлена на рис. 2.21.

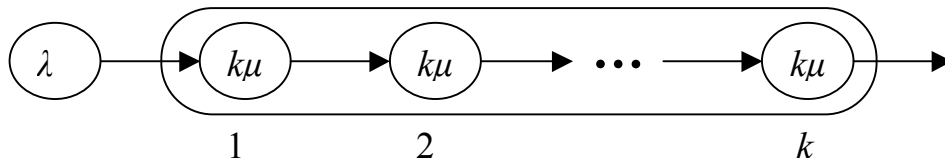


Рис 2.21. Система М/Е<sub>k</sub>/1/∞

Будем описывать состояние системы в определённый момент времени общим числом этапов обслуживания, через которое должны пройти все находящиеся в этот момент в системе заявки до полного завершения их обслуживания.

Если в системе  $n$  заявок, а обслуживаемая находится на  $i$ -ом этапе обслуживания, то общее число этапов обслуживания для всех заявок, то есть текущее состояние:

$$j = (n - 1)k + (k - i + 1) = nk - i + 1.$$

ДИП для этой системы представлена на рис. 2.22.

$$\begin{cases} \lambda P_0 = k\mu P_1; \\ (k\mu + \lambda)P_j = k\mu P_{j+1}, & 1 \leq j \leq k-1; \\ (k\mu + \lambda)P_j = k\mu P_{j+1} + \lambda P_{j-k}, & j \geq k. \end{cases}$$

стационарности  $\frac{\lambda}{\mu} < 1$ , предельные вероятности состояний существуют и могут быть найдены.

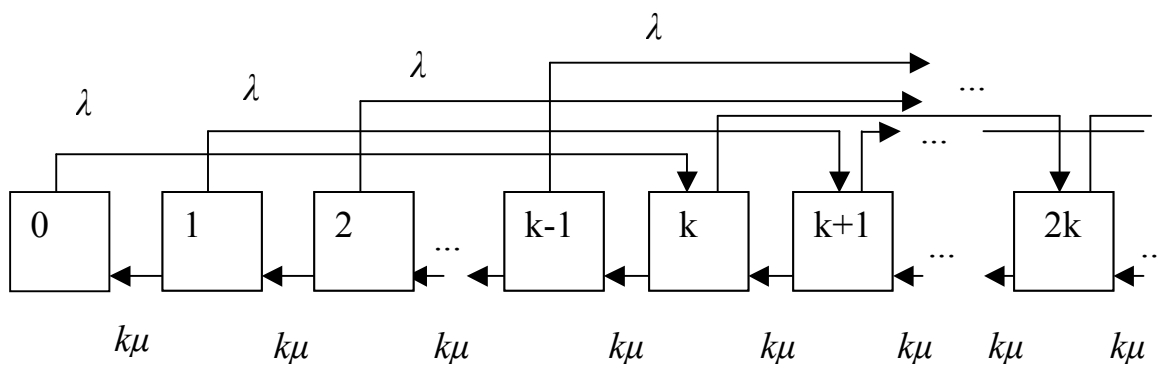


Рис 2.22. ДИП для системы  $\mathbf{M/E_k/1/\infty}$

Для этого можно использовать следующий прием. Переходим к системе М/М/1/∞ с интенсивностью потока обслуживания  $\mu$ . Задаваясь определенной точностью, находим максимальное количество заявок, которые могут реально находиться в системе. Тем самым мы ограничиваем ДИП и систему уравнений для системы М/Е<sub>r</sub>/1/∞. Теперь эту систему можно решить и найти вероятности состояний  $P_i$ . Для дальнейшего анализа характеристик эффективности системы нам потребуются вероятности  $P_i^*$  нахождения в системе  $i$  заявок, которые можно рассчитать следующим образом:

$$P_i^* = \sum_{j=ik}^{k(i-1)+1} P_j \text{ .}$$

При большом числе состояний удобно использовать для решения системы

алгебраических уравнений метод итераций, равномерно распределив для первой итерации значения вероятностей состояний системы  $M/M/1/\infty$  между состояниями этапов системы  $M/E_r/1/\infty$ .

**2. Система  $E_k/M/1$ .** Входной поток – поток Эрланга порядка  $k$  с интенсивностью  $\Lambda = \frac{\lambda}{k}$  ( $\lambda$  – интенсивность порождающего простейшего потока), поток обслуживания – простейший с интенсивностью  $\mu$ . Очередь не ограничена. Будем предполагать, что прежде, чем попасть в систему, заявка должна пройти  $k$  этапов поступления в приемном устройстве. Время прохождения одного этапа имеет показательное распределение с интенсивностью  $k\lambda$ . Структура этой системы представлена на рис. 2.23.

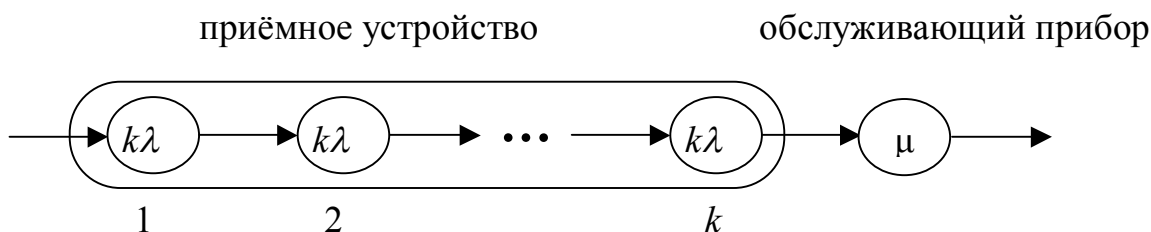


Рис 2.23. Система  $E_r / M / 1/\infty$

Будем описывать состояние системы в определённый момент времени общим числом этапов поступления, через которые прошли все находящиеся в данный момент в системе заявки плюс количество этапов, пройденных вновь поступающей заявкой.

Если в системе  $n$  заявок, а очередная поступающая находится на  $i$ -ом этапе, то общее число этапов поступления для всех заявок, то есть текущее состояние:

$$j = nk + i - 1.$$

На вход приёмного устройства поступает новая заявка, как только предыдущая появляется на его выходе. Уход из системы обслуженной заявки уменьшает суммарное число этапов поступления на  $k$ .

ДИП для этой системы представлена на рис. 2.24.

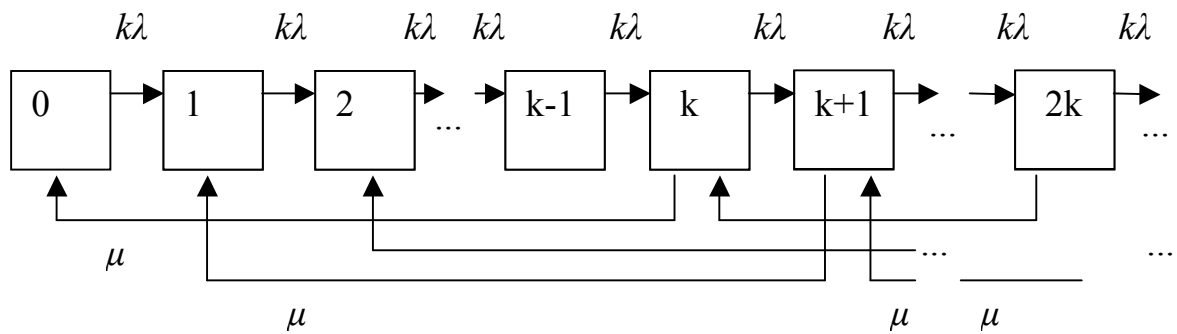


Рис. 2.24. ДИП для системы  $E_k / M / 1/\infty$

Используя закон сохранения потоков вероятностей (сумма входящих потоков для состояния равна сумме исходящих потоков), построим систему алгебраических уравнений для предельных вероятностей состояний:

$$\begin{cases} k\lambda P_0 = \mu P_k; \\ k\lambda P_j = k\lambda P_{j-1} + \mu P_{j+k}, & 1 \leq j \leq k-1; \\ (k\lambda + \mu)P_j = k\lambda P_{j-1} + \mu P_{j+k}, & j \geq k. \end{cases}$$

Для решения этой системы можно использовать приемы, описанные выше для системы  $M/E_k/1/\infty$ . Вероятности  $P_i^*$  нахождения в системе  $i$  заявок, рассчитываются следующим образом:

$$P_i^* = \sum_{j=ik}^{i(k+1)-1} P_j.$$

### 2.7.8. Немарковские СМО

До сих пор мы занимались только простейшими СМО, для которых все потоки событий, переводящие их из состояния в состояние, были простейшими. Если это условие не выполняется, то система становится немарковской.

Следует отметить, что для немарковских СМО существуют только отдельные результаты, позволяющие выразить в явном, аналитическом виде характеристики СМО через заданные условия задачи — число каналов, характер потока заявок, вид распределения времени обслуживания. Приведем некоторые из этих результатов.

**1)  $n$ -канальная СМО с отказами, с простейшим потоком заявок и произвольным распределением времени обслуживания ( $M/G/n$ ).** Ранее

мы вывели формулы Эрланга для финальных вероятностей состояний СМО с отказами (М/М/п). Б. А. Севастьянов доказал, что эти формулы справедливы не только при показательном, но и при произвольном распределении времени обслуживания.

**2) Одноканальная СМО с неограниченной очередью, простейшим потоком заявок и произвольным распределением времени обслуживания (М/G/1/∞).** Если на одноканальную СМО с неограниченной очередью поступает простейший поток заявок с интенсивностью  $\lambda$ , а время обслуживания имеет произвольное распределение с математическим ожиданием  $t_{об} = 1/\mu$  и коэффициентом вариации  $v_\mu$  то среднее число заявок в очереди равно

$$L_{оч} = \frac{\omega^2(1+v_\mu^2)}{2(1-\omega)},$$

а среднее число заявок в системе

$$L_{сист} = \frac{\omega^2(1+v_\mu^2)}{2(1-\omega)} + \omega,$$

где, как и ранее,  $\omega = \lambda/\mu$ , а  $v_\mu$  — коэффициент вариации, т.е. отношение среднего квадратического отклонения времени обслуживания к его математическому ожиданию. Эти формулы носят название формул Полячека — Хинчина.

Деля  $L_{оч}$  и  $L_{сист}$  на  $\lambda$ , получим, согласно формуле Литтла, среднее время пребывания заявки в очереди и среднее время пребывания в системе:

$$W_{оч} = \frac{\omega^2(1+v_\mu^2)}{2\lambda(1-\omega)},$$

$$W_{сист} = \frac{\omega^2(1+v_\mu^2)}{2\lambda(1-\omega)} + \frac{1}{\mu}.$$

Заметим, что в частном случае, когда время обслуживания — показательное,  $v_\mu = 1$ , то формулы Полячека — Хинчина превращаются в уже знакомые нам формулы для простейшей одноканальной СМО. Возьмем другой частный случай — когда время обслуживания вообще не случайно (поток обслуживаний регулярный) и  $v_\mu = 0$ . Тогда среднее число заявок в очереди уменьшается вдвое по сравнению с простейшим случаем. Это и естественно: если обслуживание заявки протекает более организованно, «регулярно», то СМО работает лучше, чем при плохо организованном, беспорядочном обслуживании.

**Пример.** Одноканальная СМО — ЭВМ, на которую поступают заявки (требования на расчеты). Поток заявок — простейший со средним интервалом

между заявками  $\bar{t} = 10$  мин. Время обслуживания  $T_{обсл}$  распределено по закону Эрланга 3-го порядка с математическим ожиданием  $\bar{t}_{обсл} = 8$  мин. Определить среднее число  $L_c$  заявок в СМО и среднее число  $L_{оч}$  заявок в очереди, а также средние времена пребывания заявки в системе  $W_c$  и в очереди  $W_{оч}$ .

**Решение.** Характеристики СМО могут быть найдены по формуле Полячека—Хинчина.

Имеем:  $\lambda = 0,1$  заявки/мин;  $\mu = 0,125$  заявки/мин;  $\omega = \lambda/\mu = 0,8$ .

Коэффициент вариации времени обслуживания для закона Эрланга 3-го порядка равен  $\frac{1}{\sqrt{3}}$ . Получаем:

$$L_{оч} = \frac{0,8^2(1 + \frac{1}{3})}{2(1 - 0,8)} \approx 2,13, \quad L_c = L_{оч} + 0,8 \approx 2,93.$$

По формуле Литтла:

$$W_c \approx 29,3 \text{ мин.}, \quad W_{оч} \approx 21,3 \text{ мин.}$$

**3. Одноканальная СМО с произвольным потоком заявок и произвольным распределением времени обслуживания (G/G/1/∞).** Рассматривается одноканальная СМО с неограниченной очередью, на которую поступает произвольный рекуррентный поток заявок с интенсивностью  $\lambda$  и коэффициентом вариации  $v_\lambda$  интервалов между заявками, заключенным между нулем и единицей:  $0 < v_\lambda < 1$ . Время обслуживания также имеет произвольное распределение со средним значением  $t_{об} = 1/\mu$  и коэффициентом вариации  $v_\mu$ , тоже заключенным между нулем и единицей. Для этого случая точных аналитических формул получить не удастся; можно только приближенно оценить среднюю длину очереди, ограничить ее сверху и снизу.

Доказано, что в этом случае

$$\frac{\omega^2(v_\lambda^2 + v_\mu^2)}{2(1-\omega)} - \frac{\omega(1-v_\lambda^2)}{2(1-\omega)} \leq L_{оч} \leq \frac{\omega^2(v_\lambda^2 + v_\mu^2)}{2(1-\omega)} - \frac{(1-\omega)(1-v_\lambda^2)}{2(1-\omega)}.$$

Если входящий поток — простейший, то обе оценки — верхняя и нижняя совпадают, и получается формула Полячека — Хинчина. Для грубо приближенной оценки средней длины очереди используют формулу:

$$L_{оч} = \frac{\omega^2(v_\lambda^2 + v_\mu^2)}{2(1-\omega)}.$$

Среднее число заявок в системе получается из  $L_{оч}$  простым прибавлением  $\omega$  — среднего числа обслуживаемых заявок:

$$L_{сист} = L_{оч} + \omega.$$

Что касается средних времен пребывания заявки в очереди и в системе, то они вычисляются через  $L_{оч}$  и  $L_{сист}$  по формуле Литтла делением на  $\lambda$ .

Таким образом, характеристики одноканальных СМО с неограниченной очередью могут быть (если не точно, то приближенно) найдены и в случаях, когда потоки заявок и обслуживании не являются простейшими.

**4. Многоканальная СМО с произвольным потоком заявок и произвольным распределением времени обслуживания ( $G/G/n/\infty$ ).** Точных аналитических методов для таких систем не существует. Единственное, что мы всегда можем найти, это среднее число занятых каналов  $\bar{k} = \omega$ .

Правда, если каналов действительно много (4 – 5 или больше), то непоказательное время обслуживания не страшно: был бы входной поток простейшим. Действительно, общий поток «освобождений» каналов складывается из потоков освобождений отдельных каналов, а в результате такого наложения («суперпозиции») получается, как мы знаем, поток, близкий к простейшему. Так что в этом случае замена непоказательного распределения времени обслуживания показательным приводит к сравнительно малым ошибкам. К счастью, входной поток заявок во многих задачах практики близок к простейшему.

Хуже обстоит дело, когда входной поток заведомо не простейший. В этом случае используют следующий прием. Подбирают две одноканальные СМО, из которых одна по своей эффективности заведомо «лучше» данной многоканальной, а другая — заведомо «хуже» (очередь больше, время ожидания больше). А для одноканальной СМО найти характеристики можно в любом случае.

Как же подобрать такие одноканальные СМО — «лучшую» и «худшую»? Это можно сделать по-разному. Оказывается, заведомо худший вариант можно получить, если расчленив данную  $n$ -канальную СМО на  $n$  одноканальных, а общий поступающий на них простейший поток распределять между этими одноканальными СМО в порядке очереди: первую заявку — в первую СМО, вторую — во вторую и т. д. Мы знаем, что при этом на каждую СМО будет поступать поток Эрланга  $n$ -го порядка, с коэффициентом вариации, равным  $1/\sqrt{n}$ . Что касается коэффициента вариации времени обслуживания, то он остается прежним. Для такой одноканальной СМО мы уже умеем вычислять время пребывания заявки в системе  $W_{\text{сист}}$ , оно будет заведомо больше, чем для исходной  $n$ -канальной СМО. Зная это время, можно дать «пессимистическую» оценку и для среднего числа заявок в очереди, пользуясь формулой Литтла и умножая среднее время на интенсивность  $\lambda$  общего потока заявок. «Оптимистическую» оценку можно получить, заменяя  $n$ -канальную СМО одной одноканальной, но с интенсивностью потока обслуживания в  $n$  раз большей, чем у данной, равной  $\mu n$ . Естественно, при этом параметр  $\omega$  тоже должен быть изменен, уменьшен в  $n$  раз. Для такой СМО время пребывания заявки в системе  $W_{\text{сист}}$  уменьшается за счет того, что обслуживание продолжается в  $n$  раз меньше времени. Пользуясь измененным значением  $\omega^* = \omega / n$ , коэффициентом вариации входящего потока  $v_\lambda$  и времени обслуживания  $v_\mu$ , мы можем приближенно вычислить среднее число заявок в системе  $L_{\text{сист}}$ . Вычитая из него



первоначальное (не измененное) значение  $\omega$ , мы получим среднее число заявок в очереди  $L_{\text{оч}} = L_{\text{сист}} - \omega$ . Обе характеристики будут меньше, чем для исходной  $n$ -канальной СМО (будут представлять собой «оптимистические» оценки). От них, деля на  $\lambda$ , можно перейти к «оптимистическим» оценкам для времени пребывания в СМО и в очереди.

### 3. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

#### 3.1. Условия применения имитационного моделирования

Множество математических моделей можно разбить на три подмножества: аналитические, имитационные и комбинированные (аналитико-имитационные) модели. Приведем сравнительный анализ двух первых видов моделей (комбинированные модели соединяют в себе черты моделей первых двух видов).

*Аналитической моделью* (АМ) называется совокупность функциональных соотношений или логических условий, описывающих связи между параметрами, переменными и показателями эффективности системы  $S$ .

Область применения АМ:

- 1) сравнительно простые системы;
- 2) системы, получаемые в результате упрощения (абстрагирования) реальных систем с целью изучения некоторых свойств системы.

**Достоинства АМ:**

- 1) универсальность, высокая степень общности и значимости результатов;

**Недостатки АМ:**

- 1) чувствительность к степени сложности системы;
- 2) ряд допущений, приводящих к неадекватности модели реальной системе:
  - замена реальных потоков событий при моделировании систем простейшими;
  - использование в качестве дисциплины обслуживания заявок из очереди дисциплины FIFO.

3) при моделировании сетей создание АМ становится практически невозможным при числе узлов в сети больше трех;

4) при использовании АМ невозможно исследование поведения системы при изменении со временем характера и скорости протекания процессов.

*Имитационной моделью* (ИМ) системы называются машинные программы (или алгоритмы), позволяющие имитировать на ЭВМ поведение отдельных элементов системы и связей между ними в течение заданного времени моделирования. Иначе говоря, — это формальное (то есть

выполненное на некотором формальном языке) описание логики функционирования исследуемой системы и взаимодействия отдельных ее элементов во времени, учитывающее наиболее существенные причинно-следственные связи, присущие системе, и обеспечивающее проведение статистических экспериментов.

Эксперименты на ЭВМ с имитационной моделью называются *имитационными (вычислительными) экспериментами*.

**Отличительные особенности ИМ:**

— при создании ИМ законы функционирования всей системы в целом могут быть неизвестны (достаточно знания алгоритмов, описывающих поведение отдельных элементов системы и связей между ними);

— в имитационной модели связи между параметрами и характеристиками системы выявляются, а значения исследуемых характеристик определяются в ходе имитационного эксперимента на ЭВМ.

Применение имитационного моделирования целесообразно в следующих случаях:

1) если не существует законченной постановки задачи на исследование и идет процесс познания объекта моделирования;

2) если характер протекающих в системе процессов не позволяет описать эти процессы в аналитической форме;

3) если необходимо наблюдать за поведением системы (или отдельных ее компонентов) в течение определенного периода, в том числе с изменением скорости протекания процессов;

4) при изучении новых ситуаций в системе либо при оценке функционирования ее в новых условиях;

5) если исследуемая система является элементом более сложной системы, другие элементы которой имеют реальное воплощение;

6) когда необходимо исследовать поведение системы при введении в нее новых компонентов;

7) при подготовке специалистов и освоении новой техники (в качестве тренажеров).

**Область применения ИМ:**

— широкий класс систем практически любой сложности.

**Достоинства ИМ:**

— часто единственно возможный метод исследования систем;

— возможность исследования системы на различных уровнях ее детализации, определяемых целью исследования;

— возможность исследования динамики взаимодействия элементов системы во времени и пространстве параметров системы;

— возможность оценивания характеристик системы в определенные моменты времени.

**Недостатки ИМ:**

- дороговизна: разработка хорошей ИМ часто обходится дороже создания АМ и требует больших временных затрат;
- результаты имитационного моделирования обладают меньшей степенью общности по сравнению с АМ и не позволяют выявить общие закономерности функционирования классов систем;
- не существует надежных методов оценки адекватности ИМ.

На этапе исследования и проектирования систем при построении и реализации машинных моделей (аналитических и имитационных) широко используется метод статистических испытаний (Монте-Карло), который базируется на использовании случайных чисел, т. е. возможных значений некоторой случайной величины с заданным распределением вероятностей. Статистическое моделирование представляет собой метод получения с помощью ЭВМ статистических данных о процессах, происходящих в моделируемой системе. Для получения представляющих интерес оценок характеристик моделируемой системы с учетом воздействий внешней среды статистические данные обрабатываются и классифицируются с использованием методов математической статистики.

Таким образом, сущность метода статистического (имитационного) моделирования сводится к построению для процесса функционирования исследуемой системы некоторого моделирующего алгоритма, имитирующего поведение и взаимодействие элементов системы с учетом случайных входных воздействий и воздействий внешней среды, и реализации этого алгоритма с использованием программно-технических средств ЭВМ.

В результате статистического моделирования системы получается серия частных значений искомых величин или функций, статистическая обработка которых позволяет получить сведения о поведении реального объекта или процесса в произвольные моменты времени. Если количество реализаций  $N$  достаточно велико, то полученные результаты моделирования системы приобретают статистическую устойчивость и с достаточной точностью могут быть приняты в качестве оценок искомых характеристик процесса функционирования системы.

Теоретической основой метода статистического моделирования систем на ЭВМ являются предельные теоремы теории вероятностей. Множества случайных явлений (событий, величин) подчиняются определенным закономерностям, позволяющим не только прогнозировать их поведение, но и количественно оценивать некоторые средние их характеристики, проявляющие определенную устойчивость. Характерные закономерности наблюдаются также в распределениях случайных величин, которые образуются при сложении множества воздействий. Выражением этих закономерностей и устойчивости средних показателей являются так называемые предельные теоремы теории вероятностей, часть из которых приводится ниже в пригодной для практического использования при

статистическом моделировании формулировке. Принципиальное значение предельных теорем состоит в том, что они гарантируют высокое качество статистических оценок при весьма большом числе испытаний (реализаций)  $N$ . Практически приемлемые при статистическом моделировании количественные, оценки характеристик систем часто могут быть получены уже при сравнительно небольших (при использовании ЭВМ)  $N$ .

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

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

**Теорема Пуассона.** Если проводится  $N$  независимых испытаний и вероятность осуществления события  $A$  в  $i$ -м испытании равна  $p_i$ , то относительная частота появления события  $m/N$  при  $N \rightarrow \infty$  сходится по вероятности к среднему из вероятностей  $p_i$ , т. е. при любом  $\varepsilon > 0$

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

**Теорема Чебышева.** Если в  $N$  независимых испытаниях наблюдаются значения  $x_1, x_2, \dots, x_N$  случайной величины  $\xi$ , то при  $N \rightarrow \infty$  среднее арифметическое значение случайной величины сходится по вероятности к ее математическому ожиданию  $m_\xi$ , т. е. при любом  $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\left\{ \left| (1/N) \sum_{i=1}^N x_i - m_\xi \right| \geq \varepsilon \right\} = 0.$$

**Центральная предельная теорема.** Если  $x_1, x_2, \dots, x_N$  — независимые одинаково распределенные случайные величины, то при  $N \rightarrow \infty$  закон распределения суммы  $\sum_{i=1}^N x_i$  неограниченно приближается к нормальному.

### 3.2. Этапы имитационного моделирования

Процесс имитационного моделирования можно условно разделить на три следующих последовательно выполняемых этапа:

1. Построение математической (концептуальной) модели;
2. Разработка моделирующего алгоритма и построение имитационной модели;
3. Исследование системы с помощью концептуальной модели (проведение имитационных экспериментов, обработка и интерпретация результатов).

Процесс имитационного моделирования не является строго поступательным. Между этапами существует и обратная связь, обеспечивающая уточнение, корректировку и учет дополнительной информации при разработке и использовании имитационной модели.

Дадим краткую характеристику каждому из этапов.

#### *Этап 1.*

На основании изучения содержательного описания системы, составленного в терминах предметной области, осуществляется переход к математической (концептуальной) модели.

Концептуальная модель сложной системы представляет собой упрощенное математическое или алгоритмическое описание сложной системы.

Построение концептуальной модели включает 5 взаимосвязанных этапов.

- 1.1. Постановка задачи и формулировка целей исследования.
- 1.2. Анализ системы: разбиение (декомпозиция) системы на элементы, допускающие удобное математическое или алгоритмическое описание; определение связей между элементами.
- 1.3. Определение параметров, переменных и возможных состояний системы, установление областей изменения для них.
- 1.4. Выбор показателей эффективности функционирования системы.
- 1.5. Описание концептуальной модели системы и проверка ее адекватности: использование математических моделей, проверка гипотез, предложений и математических соотношений.

#### *Этап 2.*

На данном этапе осуществляется переход от концептуальной модели к моделирующему алгоритму и имитационной модели. Этот переход осуществляется в 5 основных этапов.

- 2.1. Выбор способа имитации, а также вычислительных и программных средств реализации имитационной модели.
- 2.2. Построение логической схемы моделирующего алгоритма.
- 2.3. Алгоритмизация математических моделей, описывающих поведение элементов системы и связей между ними в рамках выбранного способа имитации.
- 2.4. Разработка имитационной модели, т.е. программирование моделирующего алгоритма.

2.5.Отладка, тестирование и проверка адекватности имитационной модели.

*Этап 3.*

Использование ИМ осуществляется в три этапа.

3.1.Планирование имитационных экспериментов .

3.2.Проведение имитационных экспериментов.

3.3.Обработка, анализ и интерпретация результатов моделирования.

Имитационные модели могут использоваться для решения различных задач исследования систем.

*Задача 1.* Оценивание значений показателей эффективности функционирования системы.

*Задача 2.* Исследование относительного влияния различных факторов на значения выходных характеристик системы.

*Задача 3* Оценивание функциональной зависимости выходных характеристик системы от входных факторов.

*Задача 4.* Сравнение эффективности функционирования системы для различных значений параметров .

*Задача 5.* Оптимизация системы, т.е. нахождение таких параметров системы, при которых показатели эффективности функционирования системы принимают максимальные или минимальные значения (экстремальный эксперимент).

Вид эксперимента влияет не только на выбор схемы формализации модели, но также на построение плана эксперимента и выбор метода обработки его результатов.

С точки зрения организации взаимодействия исследователя с моделью в ходе эксперимента ИМ делятся на автоматические и диалоговые.

Автоматическими называются ИМ, взаимодействие пользователя с которыми сводится только к вводу исходной информации и управлению началом и окончанием работы моделей.

Диалоговыми называются ИМ, позволяющие исследователю активно управлять ходом моделирования.

### **3.3. Способы моделирования случайных величин**

При создании имитационной модели возникает необходимость моделирования различных случайных факторов, к которым относятся *случайные величины, случайные события и случайные процессы*. Формирование на ЭВМ реализаций случайных объектов любой природы сводится к выработке и преобразованию случайных чисел.

На практике используются три основных способа генерации случайных чисел: аппаратный (физический), файловый (табличный) и алгоритмический (программный). Рассмотрим их, отметив достоинства и недостатки каждого.

**Аппаратный способ.** При этом способе генерации случайные числа вырабатываются специальной электронной приставкой — генератором (датчиком) случайных чисел, служащей в качестве одного из внешних устройств ЭВМ. Таким образом, реализация этого способа генерации не требует дополнительных вычислительных операций ЭВМ по выработке случайных чисел, а необходима только операция обращения к внешнему устройству (датчику). В качестве физического эффекта, лежащего в основе таких генераторов чисел, чаще всего используются шумы в электронных и полупроводниковых приборах, явления распада радиоактивных элементов и т. д.

*Достоинства:*

- реализация этого способа генерации не требует дополнительных вычислительных операций ЭВМ по выработке случайных чисел, а необходима только операция обращения к внешнему устройству;
- не занимает место в памяти машины для хранения больших массивов чисел.

*Недостатки:*

- требуется периодическая проверка статистических характеристик последовательностей;
- нельзя повторно воспроизводить одни и те же последовательности;
- используется специальное устройство и средства его сопряжения с ЭВМ;
- необходимы меры по обеспечению стабильности работы генератора.

**Табличный способ.** При использовании этого способа случайные числа, оформленные в виде таблицы, помещаются во внешнюю или оперативную память ЭВМ, предварительно сформировав из них соответствующий файл (массив чисел). Однако этот способ получения случайных чисел при моделировании систем на ЭВМ обычно рационально использовать при сравнительно небольшом объеме таблицы и соответственно файла чисел, когда для хранения можно применять оперативную память. Хранение файла во внешней памяти при частом обращении в процессе статистического моделирования не рационально, так как вызывает существенное увеличение затрат машинного времени при моделировании из-за необходимости обращения к внешнему накопителю (на магнитных дисках, лентах и т. д.). Возможны промежуточные способы организации файла, когда он переписывается в оперативную память периодически по частям. Это уменьшает время на обращение к внешней памяти, но сокращает объем

оперативной памяти, который можно использовать для моделирования процесса функционирования системы.

*Достоинства:*

- требуется однократная проверка статистических характеристик;
- можно повторно воспроизводить последовательности.

*Недостатки:*

- запас чисел ограничен;
- занимает много места в оперативной памяти или необходимо время на обращение к внешней памяти;
- невозможно при проведении эксперимента поменять значения статистических характеристик.

**Алгоритмический способ.** Способ получения последовательностей случайных чисел основан на формировании случайных чисел в ЭВМ с помощью специальных алгоритмов и реализующих их программ. Каждое случайное число вычисляется с помощью соответствующей программы по мере возникновения потребностей при моделировании системы на ЭВМ.

*Достоинства:*

- требуется однократная проверка статистических характеристик;
- можно многократно воспроизводить последовательности чисел;
- занимает мало места в памяти машины;
- не используются внешние устройства.

*Недостатки:*

- псевдослучайность чисел;
- Запас чисел последовательности ограничен ее периодом;
- Существенные затраты машинного времени.

Сравнение достоинств и недостатков трех перечисленных способов получения случайных чисел показывает, что алгоритмический способ получения случайных чисел наиболее рационален на практике при моделировании систем на ЭВМ.

### **3.4. Равномерно-распределённые случайные числа (РРСЧ).**

При исследовании систем методом имитационного моделирования существенное количество операций, а, следовательно, и времени расходуется на действия со случайными числами. Поэтому наличие простых и экономных способов программного формирования последовательностей случайных чисел во многом определяет возможность практического использования этого метода.

В качестве исходной (базовой) последовательности случайных чисел для имитации случайных факторов различной природы необходимо выбирать такую последовательность, которая может быть получена с наименьшими, по



возможности затратами машинного времени и, кроме того, обеспечивает простоту и удобство дальнейших преобразований.

Практика показывает, что в наибольшей степени этим требованиям удовлетворяет последовательность случайных чисел с равномерным распределением в интервале  $(0, 1)$ .

Непрерывная случайная величина  $\eta$  имеет равномерное распределение в интервале  $(a, b)$ , если ее функции плотности (рис.3.1) и распределения (рис.3.2) имеют следующий вид

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b); \\ 0 & x \notin (a, b). \end{cases}$$

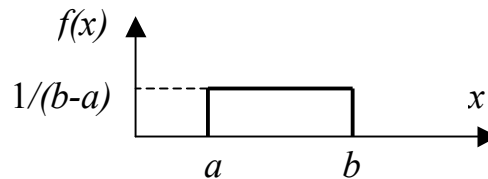


Рис.3.1. Функция плотности равномерного распределения

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

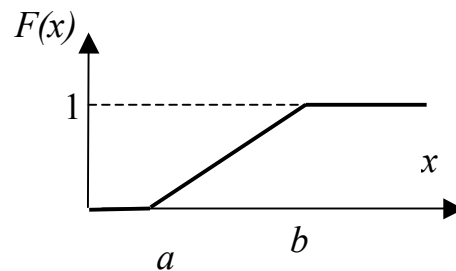


Рис. 3.2. Функция вероятностей равномерного распределения

Числовые характеристики этого распределения : математическое ожидание, дисперсия и среднее квадратическое отклонение равны соответственно

$$m_{\eta} = \frac{a+b}{2}, \quad D_{\eta} = \frac{(b-a)^2}{12}, \quad \sigma_{\eta} = \frac{b-a}{2\sqrt{3}}.$$

При  $a = 0, b = 1$  имеем

$$m_{\eta} = \frac{1}{2}; \quad \sigma_{\eta} = \frac{1}{2\sqrt{3}}; \quad D_{\eta} = \frac{1}{12}.$$

Это распределение требуется получить на ЭВМ. Но при этом следует учитывать то, машина оперирует с  $n$ -разрядными числами. Поэтому вместо непрерывной совокупности равномерных случайных чисел интервала  $(0, 1)$  используется дискретная последовательность  $2^n$  случайных чисел того же

интервала. Закон распределения такой дискретной последовательности называют *квазиравномерным распределением*.

Рассмотрим характеристики этого распределения.

При использовании ЭВМ целые числа представляются в двоичном виде как  $\xi^* = z_1 z_2 \dots z_n$ , где  $z_i$  принимает значение 0 или 1, а  $n$  – длина разрядной сетки. Если выполняется условие  $P(z_i = 0) = P(z_i = 1) = 0,5$ , то  $\xi^*$  – квазиравномерно распределённые числа. Для получения интересующей нас последовательности чисел  $\xi$  из интервала  $(0, 1)$ , надо числа  $\xi^*$  пронормировать.

При  $n$ -разрядной двоичной сетке можно представить  $2^n$  различных значений :

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

Вероятность каждого значения равна  $2^{-n}$ .

Найдём математическое ожидание  $m_\xi$  и среднее квадратичное отклонение  $\sigma_\xi$  случайной величины  $\xi$ .

$$m_\xi = \sum_{i=0}^{2^n-1} P_i x_i = \sum_{i=0}^{2^n-1} \frac{i}{(2^n - 1)} \cdot \frac{1}{2^n} = \frac{1}{2^n (2^n - 1)} \sum_{i=0}^{2^n-1} i.$$

Известно, что:

$$\sum_{i=0}^k i = \frac{k(k+1)}{2};$$

Отсюда:

$$m_\xi = \frac{1}{2}.$$

Дисперсию можно найти, используя начальный момент :

$$D_\xi = \sum_{i=0}^{2^n-1} P_i x_i^2 - m_\xi^2 = \sum_{i=0}^{2^n-1} \frac{1}{2^n} \cdot \frac{i^2}{(2^n - 1)^2} - \frac{1}{4} = \frac{1}{2^n (2^n - 1)^2} \sum_{i=0}^{2^n-1} i^2 - \frac{1}{4}.$$

Учитывая, что  $\sum_{i=0}^k i^2 = \frac{k(k+1)(2k+1)}{6}$ , получим:

$$D_\xi = \frac{1}{12} \cdot \frac{2^k + 1}{2^k - 1}, \quad \sigma_\xi = \frac{1}{2\sqrt{3}} \sqrt{\frac{2^k + 1}{2^k - 1}}.$$

Таким образом, математическое ожидание квазиравномерной случайной величины совпадает с математическим ожиданием равномерной случайной последовательности интервала  $(0, 1)$ , а дисперсия отличается только множителем

$$\frac{2^k + 1}{2^k - 1},$$

который при достаточно больших  $n$  близок к единице.

### 3.4.1. Методы формирования РРСЧ.

Наибольшее применение для генерации РРСЧ на ЭВМ получили алгоритмы вида

$$x_{i+1} = \Psi(x_i),$$

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

Для получения качественной последовательности РРСЧ важен вид  $\Psi(x_i)$ . Если рассматривать пары чисел  $(x_1, x_2), (x_3, x_4), \dots, (x_i, x_{i+1}) \dots$  как координаты точек, то, при равномерном распределении значений  $x_i$  интервале от 0 до 1, эти точки должны равномерно заполнить единичный квадрат. Следовательно, хорошую последовательность случайных чисел может породить только такая функция  $\Psi(x_i)$ , график которой достаточно плотно заполняет единичный квадрат. Примером такой функции может служить  $x_{i+1} = D(Ax_i)$  при больших целых положительных  $A$ , где  $D(Ax_i)$  – дробная часть числа  $Ax_i$  (рис. 3.3).

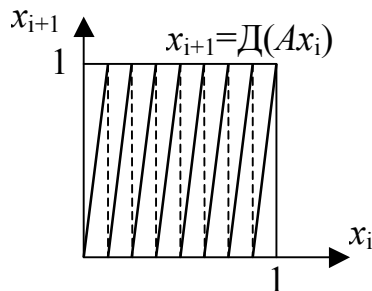


Рис. 3.3. Вид функции  $\Psi(x_i)$

Одной из исторически первых процедур получения псевдослучайных чисел была процедура, получившая название *метода серединных квадратов*. Пусть имеется  $2n$ -разрядное число, меньшее 1:

$$x_1 = 0, a_1 a_2 \dots a_{2n}.$$

Возведем его в квадрат:

$$x_1^2 = 0, b_1 b_2 \dots b_{4n},$$

а затем выделим средние  $2n$  разрядов полученного числа и будем использовать их в качестве очередного значения псевдослучайной последовательности:

$$x_{i+1} = 0, b_{n+1} b_{n+2} \dots b_{3n}.$$

В настоящее время почти все библиотеки стандартных программ ЭВМ для вычисления последовательностей равномерно распределенных случайных чисел основаны на *мультипликативном методе*.

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

$$X_{i+1} = A X_i \pmod{M},$$

то есть очередное значение  $X_{i+1}$  получается как остаток от деления  $A X_i$  на  $M$ .

Преобразование целых чисел  $\{X\}$  в дробные  $\{x\}$  из интервала от 0 до 1 осуществляется путем деления целых остатков на  $M$ :

$$x_i = X_i / M.$$

В силу детерминированности метода получаются воспроизводимые последовательности.

Смешанный метод позволяет вычислить последовательность неотрицательных чисел  $\{X_i\}$ , не превосходящих  $M$ , по формуле

$$X_{i+1} = (A X_i + C) \pmod{M},$$

т.е. в отличие от мультипликативного метода  $C \neq 0$ . С вычислительной точки зрения смешанный метод генерации сложнее мультипликативного на одну операцию сложения, но при этом возможность выбора дополнительного параметра позволяет уменьшить возможную корреляцию получаемых чисел.

Процедура эта чисто детерминированная. Если раскрыть рекуррентное соотношение то:

$$X_i = (a^i X_0 + \frac{(a^i - 1)C}{(a - 1)}) \pmod{M}.$$

Отсюда видно, что для любого  $i$   $X_i < m$ . При правильно выбранных  $X_0$ ,  $a$  и  $C$  получается квазиравномерная последовательность чисел:

$$x_i = \frac{X_i}{M}.$$

При машинной реализации метода для увеличения периода берут

$$M = 2^n,$$

где  $n$  – разрядность машины.

В качестве  $a$  берут число близкое к  $2^{n/2}$ ,  $x_0$  – любое нечётное число меньше  $M$ , а значение  $C$  подбирают экспериментально, оно влияет на корреляционные свойства последовательности.

### 3.4.2. Проверка качества последовательностей РРСЧ

Эффективность статистического моделирования систем на ЭВМ и достоверность получаемых результатов существенным образом зависят от качества исходных (базовых) последовательностей псевдослучайных чисел, которые являются основой для получения стохастических воздействий на элементы моделируемой системы. Поэтому, прежде чем приступить к реализации моделирующих алгоритмов на ЭВМ, необходимо убедиться в том, что исходная последовательность псевдослучайных чисел удовлетворяет предъявляемым к ней требованиям, так как в противном случае даже при наличии абсолютно правильного алгоритма моделирования процесса функционирования моделируемой системы по результатам моделирования нельзя будет достоверно судить о характеристиках системы. Поэтому все применяемые генераторы случайных чисел должны перед моделированием системы пройти тщательное предварительное тестирование, которое представляет собой комплекс проверок по различным статистическим критериям, включая в качестве основных проверки (тесты) на *равномерность*, *стохастичность* и *независимость*. Кроме того, очень важными показателями качества базовой последовательности являются *длина периода* и *длина отрезка аperiodичности*. Рассмотрим возможные методы проведения таких проверок, наиболее часто используемые в практике статистического моделирования.

#### 1) Проверка равномерности.

Проверка равномерности последовательностей псевдослучайных квазиравномерно распределенных чисел  $\{x_r\}$  может быть выполнена по гистограмме или с использованием косвенных признаков.

##### а) Проверка по гистограмме (рис.3.4).

Суть проверки по гистограмме сводится к следующему. Выдвигается гипотеза о равномерности распределения чисел в интервале  $(0, 1)$ . Затем интервал  $(0, 1)$  разбивается на  $m$  равных частей. При генерации последовательности РРСЧ подсчитывается количество попаданий  $N_k$  в каждый из  $m$  подинтервалов. Вычисляется относительная частота попадания случайных чисел последовательности  $\{x_r\}$  в каждый из подинтервалов

$$C_k = N_k / N,$$

где  $N = \sum_{k=1}^m N_k$  – общее количество чисел в последовательности  $\{x_r\}$ .

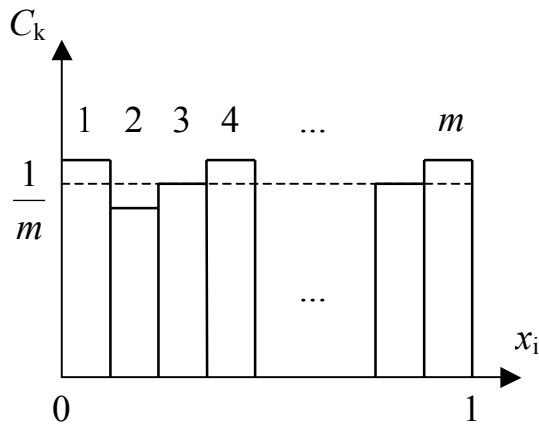


Рис.3.4. Проверка равномерности по гистограмме

Очевидно, что при равномерности последовательности чисел, частоты должны быть близкими при достаточно больших  $N$  к теоретической вероятности попадания в подинтервалы, равной  $1/m$ .

Оценка степени приближения, т. е. равномерности последовательности  $\{x_r\}$ , может быть проведена с использованием критериев согласия. На практике обычно принимается  $m = 20 \div 50$ ,  $N = (10^2 \div 10^3)m$ .

#### б) Проверка по косвенным признакам.

Суть проверки равномерности по косвенным признакам сводится к следующему. Вся последовательность  $\{x_r\}$  разбивается на пары чисел:

$$(x_1, x_2), (x_3, x_4), \dots, (x_{2i-1}, x_{2i}), \dots, (x_{N-1}, x_N).$$

Затем подсчитывают число пар  $K$ , для которых выполняется условие:

$$x_{2i-1}^2 + x_{2i}^2 < 1.$$

Геометрически это означает, что точка с координатами  $(x_{2i-1}, x_{2i})$  расположена внутри четверти круга радиуса  $R=1$ , вписанного в единичный квадрат (рис. 3.5).

В общем случае точка  $(x_{2i-1}, x_{2i})$  всегда попадет внутрь единичного квадрата. Тогда теоретическая вероятность попадания этой точки в четверть круга равна отношению площади четверти круга к площади единичного квадрата:

$$P = S_{1/4 \text{ круга}} / S_{\text{квадрата}} = (\pi R^2 / 4) / (1 \cdot 1) = \pi / 4.$$

Если числа последовательности  $\{x_r\}$  равномерны, то в силу закона больших чисел теории вероятностей при больших  $N$  относительная частота попадания точки в единичный квадрат, равная отношению числа  $K$  пар  $(x_{2i-1}, x_{2i})$ , для которых проверочное условие выполнилось к общему числу  $N/2$  пар последовательности должна сходиться к  $P$ :

$$\frac{2K}{N} \rightarrow \frac{\pi}{4}.$$

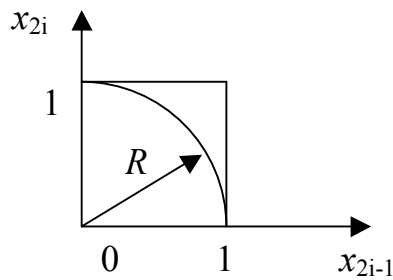


Рис. 3.5. Проверка равномерности по косвенным признакам

## 2) Проверка стохастичности

Это исследование последовательностей псевдослучайных чисел  $\{x_r\}$  наиболее часто проводится методами комбинаций и серий.

### а) метод комбинаций

Сущность метода комбинаций сводится к определению закона распределения закона распределения (появления) числа единиц (нулей) в  $n$ -разрядном двоичном числе  $x_r$ . На практике длину последовательности  $N$  берут достаточно большой и проверяют все  $n$  разрядов или только  $l$  старших разрядов числа  $x_r$ .

Теоретически закон появления  $j$  единиц в  $l$  разрядах двоичного числа описывается исходя из независимости отдельных разрядов биномиальным законом распределения:

$$P(j, l) = C_l^j P^j(1) [1 - P(1)]^{l-j} = C_l^j P^l(1),$$

где  $P(j, l)$  – вероятность появления  $j$  единиц в  $l$  разрядах числа  $x_r$ ;  $P(1)=P(0)=0,5$  – вероятность появления единицы (нуля) в любом разряде числа  $x_r$ ;  $C_l^j = l!/[j!(l-j)!]$ .

Тогда при фиксированной длине выборки  $N$  теоретически ожидаемое число появления случайных чисел с  $j$  единицами в проверяемых  $l$  разрядах будет равно

$$n_j = N C_l^j P^l(1).$$

После нахождения теоретических и экспериментальных вероятностей  $P(j, l)$  или чисел  $n_j$  при различных значениях  $l \leq n$  гипотеза о стохастичности проверяется с использованием критериев согласия.

### б) метод серий

В этом случае вся последовательность чисел  $\{x_r\}$  разбивается на элементы 1-го и 2-го рода по следующему правилу:

$$x_i = \begin{cases} a, & \text{если } x_i < p; \\ b & \text{если } x_i \geq p, \end{cases}$$

где  $0 < p < 1$ .

*Серией* называется любой отрезок последовательности  $\{x_r\}$ , состоящий из следующих друг за другом элементов одного и того же рода. Причем число элементов в отрезке ( $a$  или  $b$ ) называется *длиной серии*.

После разбиения последовательности  $\{x_r\}$  на серии первого и второго рода будем иметь, например, последовательность вида

...*aabbbbbaaabaabbbbab*...

Так как случайные числа  $a$  и  $b$  в данной последовательности независимы и принадлежат последовательности  $\{x_r\}$ , равномерно распределенной на интервале  $(0, 1)$ , то теоретическая вероятность появления серии длиной  $j$  в последовательности длиной  $l$  в  $N$  опытах (под опытом здесь понимается генерация числа  $x_i$  и проверка условия  $x_i < p$ ) определится формулой Бернулли

$$P(j, l) = c_l^j p^j (1-p)^{l-j}, \quad j = \overline{0, l}, \quad l = \overline{1, n}.$$

В случае экспериментальной проверки оцениваются частоты появления серий длиной  $j$ . В результате получаются теоретическая и экспериментальная зависимости  $P(j, l)$ , сходимость которых проверяется по известным критериям согласия, причем проверку целесообразно проводить при различных значениях  $p(0 < p < 1)$  и  $l$ .

### 3) Проверка независимости

Случайные величины  $\xi$  и  $\eta$  называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая.

Проверка независимости проводится на основе вычисления корреляционного момента. В общем случае корреляционный момент случайных величин  $\xi$  и  $\eta$  с возможными значениями  $x_i$  и  $y_j$  определяется по формуле

$$K_{\xi\eta} = \sum_i \sum_j (x_i - M[\xi])(y_j - M[\eta]) P_{ij}$$

где  $P_{ij}$  – вероятность того, что  $(\xi, \eta)$  принимает значение  $(x_i, y_j)$ , а  $M[\xi]$ ,  $M[\eta]$  – математические ожидания случайных величин.

Если случайные числа независимы, то  $K_{\xi\eta} = 0$ .

Независимость элементов последовательности  $\{x_r\}$  может быть проверена путем введения в рассмотрение последовательности  $\{y_r\}$  такой, что  $\{y_r\} = \{x_{r+\tau}\}$ , где  $\tau$  – величина сдвига последовательностей.

Иногда вместо корреляционного момента удобнее использовать коэффициент корреляции

$$\rho_{\xi\eta} = \frac{K_{\xi\eta}}{\sigma_{\xi} \sigma_{\eta}},$$



где  $\sigma_\xi$  и  $\sigma_\eta$  – среднеквадратические отклонения величин  $\xi$  и  $\eta$ .

Возможные значения коэффициента корреляции лежат в пределах от 0 (полная независимость) до 1 (жесткая функциональная связь).

При любом  $\tau \neq 0$  для достаточно больших  $N$  с доверительной вероятностью  $\beta$  справедливо соотношение

$$|\rho_{\xi\eta}| \leq \beta \sqrt{\frac{1}{N}}.$$

Если вычисленное экспериментальным путём  $\rho$  лежит в этих пределах, то вероятностью  $\beta$  можно утверждать, что последовательность корреляционно независима.

При проведении оценок коэффициента корреляции на ЭВМ удобно для вычисления использовать следующее выражение:

$$\rho_{\xi\eta} = \frac{(\frac{1}{N-\tau} \sum_{i=1}^{N-\tau} x_i x_{i+\tau} - \frac{1}{(N-\tau)^2} \sum_{i=1}^{N-\tau} x_i \sum_{i=1}^{N-\tau} x_{i+\tau})}{\sqrt{D[x_i]D[x_{i+\tau}]}} ,$$

где

$$D[x_i] = [\frac{1}{(N-\tau)} \sum_{i=1}^{N-\tau} (x_i^2 - \frac{1}{(N-\tau)^2} (\sum_{i=1}^{N-\tau} x_i)^2)],$$

$$D[x_{i+\tau}] = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} (x_{i+\tau}^2 - \frac{1}{(N-\tau)^2} (\sum_{i=1}^{N-\tau} x_{i+\tau})^2).$$

#### 4) Определение длины периода и длины отрезка аперiodичности

При статистическом моделировании с использованием программных генераторов псевдослучайных квазиравномерных последовательностей важными характеристиками качества генератора является *длина периода*  $P$  и *длина отрезка аперiodичности*  $L$ . Длина отрезка аперiodичности  $L$  псевдослучайной последовательности  $\{x_r\}$ , заданной уравнением

$$X_{i+1} = (A X_i + C) \bmod M, \quad x_i = X_i / M,$$

есть наибольшее целое число, такое, что при  $0 \leq k < i \leq L$  событие  $P(x_i = x_k)$  не имеет места. Это означает, что все числа  $x_i$  в пределах отрезка аперiodичности не повторяются.

Очевидно, что использование при моделировании систем последовательности чисел  $\{x_r\}$ , длина которой больше отрезка аперiodичности  $L$ , может привести к повторению испытаний в тех же условиях, что и раньше, т. е. увеличение числа реализаций не дает новых статистических результатов.

Способ экспериментального определения длины периода  $P$  и длины отрезка аперiodичности  $L$  сводится к следующему.

1) Запускается программа генерации последовательности чисел  $\{x_r\}$  с начальным значением  $x_0$  на  $V$  значений, фиксируется  $x_v$  (обычно полагают  $V = (1 \div 5)10^6$ );

2) Запуск программы генерации с  $x_0$  и фиксируется  $i_1$  и  $i_2$ , такие, что в первый и во второй раз выполняется условие  $x_{i_1}=x_v$  и  $x_{i_2}=x_v$ . Вычисляется длина периода последовательности  $P=i_2-i_1$ .

3) Запуск программы генерации с начальными значениями  $x_0$  и  $x_p$  и фиксируется минимальный номер  $i_3$ , для которого справедливо  $x_{i_3}=x_{i_3+p}$ . Вычисляется длина отрезка аperiodичности  $L=i_3+p$ .

Теоретически при использовании длина периода не может быть больше чем  $2^n$ , где  $n$  – разрядность ЭВМ. Для увеличения длины периода прибегают к специальным приемам. Рассмотрим некоторые из них.

1) Использование рекуррентных формул порядка  $r$

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

Естественно, это ведёт к увеличению затрат машинного времени на получение чисел и ограничивает возможности его применения на практике.

2) Метод возмущений

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

т.е. если  $i$  не кратно  $M$ , то вычисление значения последовательности осуществляется с помощью функции  $\Phi$ , если кратно, то используется другая функция  $\psi$ .

3) Метод Макларена – Марсальи

Метод основан на комбинировании двух датчиков РРСЧ, построенных с использованием мультипликативного метода. Пусть  $\{a_i\}$  и  $\{b_i\}$  – последовательности, порождаемые с помощью этих датчиков. Выходную последовательность обозначим как  $\{c_i\}$ . Будем использовать вспомогательный массив  $V=\{V_1, V_2, \dots, V_k\}$ .

Вначале массив заполняется элементами последовательности от первого датчика  $\{a_i\}$ :

$$V_i = a_i, i=1 \div k.$$

Затем с помощью числа от второго датчика разыгрывается случайный выбор номера  $s$  элемента в массиве  $V$ :

$$s=[b_j k], \text{ где скобки означают «целая часть»}.$$

Выбранный элемент массива пересылается в выходную последовательность, а на его место помещается очередное значение из последовательности  $\{a_i\}$ :

$$c_j = V_s, V_s = a_{k+j} (j=1, 2, \dots).$$

Этот метод позволяет ослабить зависимость между членами последовательности  $\{c_i\}$  и получить чрезвычайно большие периоды, если периоды последовательностей  $\{a_i\}$  и  $\{b_i\}$  – взаимно простые числа.

### 3.5. Формирование случайных величин с заданным законом распределения.

Для формирования чисел с заданным законом распределения исходным материалом служат случайные величины, имеющие равномерное распределение. Другими словами, возможные значения  $y_i$  случайной величины  $\eta$ , имеющей равномерное распределение в интервале  $(0, 1)$ , могут быть преобразованы в возможные значения  $x_i$  случайной величины  $\xi$ , закон распределения которой задан.

Существуют два основных пути такого преобразования случайных чисел. Первый из них состоит в реализации некоторой операции над числом  $y_i$ , формирующей число  $x_i$ , имеющее (точно или приближенно) заданный закон распределения. Второй основывается на моделировании условий соответствующей предельной теории вероятностей. Рассмотрим некоторые из существующих методов.

#### 3.5.1. Метод обратной функции.

Идея этого метода базируется на следующем утверждении.

Если случайная величина  $\xi$  имеет функцию распределения  $F_\xi(x)$ , то распределение случайной величины  $\eta = F_\xi(\xi)$  равномерно в интервале от 0 до 1 (рис. 3.6.а).

Иными словами, если бы у нас имелась реализация  $\{x_i\}$  случайной величины  $\xi$ , то, преобразовав ее с помощью функции  $F_\xi(x)$ , мы получим реализацию  $\{y_i\}$  случайной величины  $\eta$ , равномерно распределенной в интервале от 0 до 1:

$$y_i = F_\xi(x_i). \quad (3.1)$$

Но наша задача заключается в том, что бы, имея выборку  $\{y_i\}$  случайной величины  $\eta$ , равномерно распределенной в интервале от 0 до 1, получить числа  $\{x_i\}$  с функцией распределения  $F_\xi(x)$ . Отсюда вытекает идея метода – разрешить относительно  $x_i$  уравнение (3.1), то есть получить выражение для обратной относительно  $F_\xi(x)$  функции (рис. 3.6.б):

$$x_i = F_\xi^{-1}(y_i).$$

Докажем, что случайная величина, значения которой получены в соответствии с (3.1), будет иметь равномерное распределение.

Пусть  $\eta$  – равномерно распределена от 0 до 1. Рассмотрим случайное число  $\xi = \varphi(\eta)$ , где функция  $\varphi$  – монотонна и непрерывна.

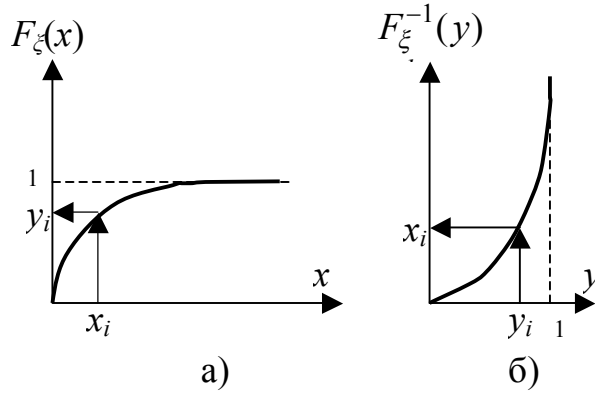


Рис. 3.6. Метод обратной функции

Тогда  $\eta = \varphi^{-1}(\xi)$ .

Найдем функцию распределения для величины  $\xi$ :

$$F_{\xi}(x) = P(\xi < x) = P(\varphi(\eta) < x).$$

Ввиду монотонности функции  $\varphi$  справедливо

$$p(\eta < \varphi^{-1}(x)) = F_{\eta}(\varphi^{-1}(x)) = \int_0^{\varphi^{-1}(x)} \underbrace{f_{\eta}(\varphi^{-1}(x))}_{1} dx = \varphi^{-1}(x).$$

Здесь для получения конечного выражения учтено то, что функция  $f_{\eta}(\varphi^{-1}(x))$  представляет собой функцию плотности для равномерного в интервале  $(0, 1)$  распределения и равна 1 в этом интервале и 0 вне его.

Окончательно получим:

$$F_{\xi}(x) = \varphi^{-1}(x).$$

Отсюда:

$$\varphi(\eta) = F_{\xi}^{-1}(\eta) = \xi.$$

Таким образом, функция  $\varphi$ , связывающая равномерно распределённую величину  $\eta$  с величиной  $\xi$ , имеющую функцию распределения  $F_{\xi}$  есть функция  $F_{\xi}^{-1}$ , т.е. обратная по отношению к  $F_{\xi}$ .

Рассмотрим пример использования метода обратной функции.

Надо получить число с показательным законом распределения. Для него функция плотности имеет вид:

$$f_{\xi}(x) = \lambda e^{-\lambda x}.$$

Получим функцию распределения для показательного закона:

$$F_{\xi}(x) = \int_0^x f(x)dx = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}.$$

В соответствии с тем, что было рассмотрено выше, если в это выражение подставить величину  $x_i$ , имеющую показательное распределение, то в результате получится число  $y_i$ , равномерно распределенное от 0 до 1:

$$y_i = 1 - e^{-\lambda x_i},$$

или 
$$1 - y_i = e^{-\lambda x_i}.$$

Прологарифмируем обе части уравнения:

$$\ln(1 - y_i) = -\lambda x_i.$$

Отсюда

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

Если величина  $y_i$  равномерно распределена от 0 до 1, то и  $1 - y_i$  будет обладать этим же свойством, поэтому окончательно:

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

Итак, если в нашем распоряжении имеются числа  $\{y_i\}$ , равномерно распределенные от 0 до 1, то, воспользовавшись полученным выражением, можно вычислить последовательность чисел  $\{x_i\}$ , имеющих показательное законом распределение.

### 3.5.2. Универсальный метод

Для использования метода обратной функции необходимо иметь в явном виде аналитическое выражение для функции распределения вероятностей  $F_{\xi}(x)$ . Как правило, законы распределения задаются функцией плотности  $f_{\xi}(x)$  а функцию распределения получают интегрируя  $f_{\xi}(x)$ :

$$F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(x)dx.$$

Для многих распределений использование метода обратной функции в аналитическом виде оказывается затруднительным или невозможным по ряду причин:

- а) интеграл от  $f_{\xi}(x)$  не берётся, либо после интегрирования получается выражение, требующее больших затрат машинного времени;
- б) описание вида распределения получено экспериментально в виде гистограммы.

В этих случаях используют универсальный способ получения случайных чисел, основанный на кусочной аппроксимации функции плотности.

Пусть требуется получить последовательность случайных чисел  $\{x_i\}$  с функцией плотности распределения  $f_\xi(x)$ .

Если область определения случайной величины  $\xi$  не ограничена, перейдем к усеченному распределению на интервале  $(c, d)$ . Разобьем интервал  $(c, d)$  на  $n$  подинтервалов (рис. 3.7)

$$(a_0, a_1), (a_1, a_2), \dots, (a_{n-1}, a_n), \quad a_0=c, a_n=d.$$

Границы интервалов выбираются так, чтобы вероятность попадания в любой из подинтервалов  $(a_k, a_{k+1})$  была постоянной, то есть не зависела от  $k$ . Для вычисления границы  $a_k$  воспользуемся следующим соотношением:

$$P_k = \int_{a_{k-1}}^{a_k} f(x) dx = \frac{1}{n}.$$

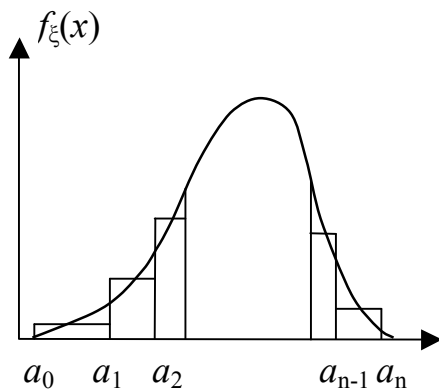


Рис. 3.7. Кусочная аппроксимация функции плотности

В простейшем случае будем представлять  $f_\xi(x)$  в виде кусочно-постоянной функции, то есть считаем значение  $f_\xi(x)$  на каждом подинтервале постоянной. Тогда значение случайной величины  $\xi$ , попадающей в подинтервал  $(a_k, a_{k+1})$  можно представить как

$$x_k = a_k + \eta_k^*,$$

где:

$a_k$  — левая граница подинтервала;

$\eta_k^*$  — РРСЧ, значение которого распределены от 0 до  $(a_{k+1} - a_k)$ .

Таким образом, вероятность всех чисел из любого подинтервала  $(a_k, a_{k+1})$  одинакова.

Машинный алгоритм этого способа:

1) Генерируется РРСЧ  $\eta_1$  из диапазона  $(0, 1)$  и выбирается номер  $k$  подинтервала из условия:

$$k \leq n \cdot \eta_1 < k+1, \quad k=0, \dots, n-1;$$

2) Генерируется РРСЧ  $\eta_2$  (от 0 до 1) и формируется случайная величина выходной последовательности

$$x_k = a_k + (a_{k+1} - a_k) \eta_2.$$

Универсальный метод при кусочно-постоянной аппроксимации  $f_{\xi}(x)$  весьма прост в реализации и требует малого количества операций ЭВМ. Когда требуется обеспечить особо высокую точность преобразования случайных чисел могут оказаться полезными и другие виды аппроксимации, например, линейно-кусочная. Правда, выражение для получения  $x_k$  при этом существенно усложняется.

В случае, когда описание вида распределения задано в виде гистограммы, применяют модификацию универсального метода.

Пусть требуется произвести формирование случайной последовательности  $\{x_i\}$ , причем закон распределения для нее представлен в виде равноинтервальной гистограммы (рис. 3.8), содержащей  $n$  интервалов с границами  $a_0, a_1, a_2, \dots, a_{n-1}, a_n$ .

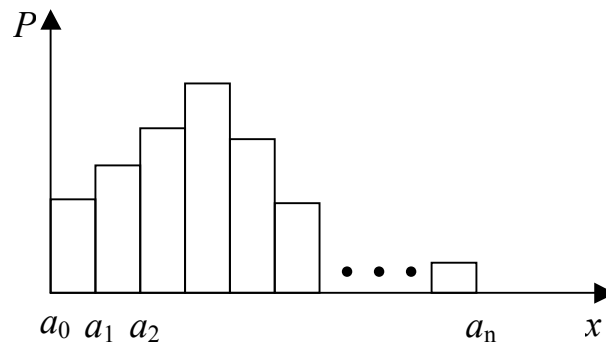


Рис. 3.8. Задание закона распределения гистограммой

Построим вспомогательную шкалу:

$$0, c_1, c_2, c_3, \dots, c_n, 1,$$

где  $c_i = \sum_{j=1}^{i-1} P_j$ ;  $P_i$  – относительная частота попадания в  $i$ -й интервал гистограммы.

Машинный алгоритм реализации способа:

1) Генерируется РРСЧ  $\eta_1$  из диапазона  $(0, 1)$  и выбирается номер  $k$  интервала вспомогательной шкалы из условия:

$$c_k \leq \eta_1 < c_{k+1}, \quad k=0, \dots, n-1;$$

2) Генерируется РРСЧ  $\eta_2$  (от 0 до 1) и формируется случайная величина выходной последовательности

$$x_k = a_k + (a_{k+1} - a_k) \eta_2.$$

Этот метод можно применить и в том случае, если задана  $f(x)$  в аналитическом виде. Для этого предварительно рассчитывают значения  $P_i$  как

$$P_i = \int_{a_{i-1}}^{a_i} f(x) dx$$

или

$$P_i = \frac{a_n - a_0}{n} \cdot \frac{f(a_i) + f(a_{i+1})}{2},$$

а далее по описанному выше алгоритму.

### 3.5.3. Метод исключения (отбраковки, режекции, Дж. Неймана)

Основой метода является следующая теорема:

*Если множество точек  $(x,y)$  является реализацией случайного вектора  $\|\xi, \eta\|$ , равномерно распределенного в области, ограниченной осью  $OX$  и кривой  $f(x)$ , такой, что*

$$\int_{-\infty}^{\infty} f(x) dx = \bar{G} < \infty,$$

*то одномерная плотность распределения величины  $\xi$  равна*

$$\varphi_{\xi}(x) = f(x) / \bar{G}.$$

Отсюда следует, что, если  $f(x)$  – искомая функция плотности распределения (т.е.  $\bar{G}=1$ ), то, реализовав процедуру построения множества случайных векторов, равномерно распределенных под графиком  $f(x)$ , мы, тем самым, получим множество значений абсцисс  $\{x_i\}$ , распределённых в соответствии с законом, определяемым функцией  $f(x)$ .

Однако, процедура генерирования случайных векторов может оказаться достаточно сложной в вычислительном плане (надо использовать условные функции плотности вероятностей). Для упрощения процедуры формирования случайных чисел воспользуемся следующим достаточно очевидным утверждением.

*Если вектор  $\|\xi, \eta\|$  равномерно распределён под кривой  $g_1(x)$ , такой, что*

$$g_1(x) \geq f(x) \geq 0$$

*для всей области определения  $f(x)$ , то его реализации, ограниченные областью между  $OX$  и  $f(x)$  будут равномерно распределены в этой области.*

Функция  $g_1(x)$  называется мажорирующей функцией по отношению к  $f(x)$  (рис. 3.9) .

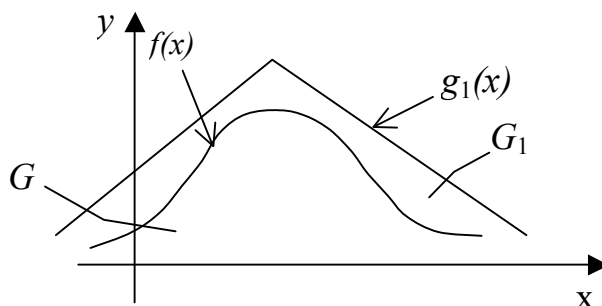




Рис. 3.9. Мажорирующая функция

Можно выбрать функцию  $g_1(x)$  такой, что вектора, равномерно распределенные в области  $G_1$ , будет просто генерировать. Удобно выбирать функцию, имеющую постоянное значение на всей области определения  $f(x)$ :

$$g_1(x) = \text{const},$$

как правило, равное максимальному значению  $f_{\max}$  функции  $f(x)$ .

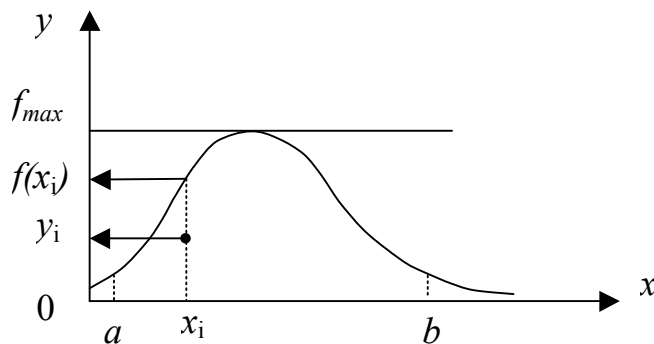


Рис. 3.10. Метод исключения

Отсюда вытекает процедура метода (рис. 3.10):

а) если функция  $f(x)$  не ограничена, то, учитывая допустимую погрешность, область ее определения ограничивают интервалом  $(a, b)$ ;

б) имитируется реализация  $(x_i, y_i)$  вектора  $\|\xi, \eta\|$ , равномерно распределенного в области  $G_1$  (между  $OX$  и  $g_1(x)$ ):

– генерируется РРСЧ  $x_i$  из диапазона  $(a, b)$ ;

– генерируется РРСЧ  $y_i$  из диапазона  $(0, f_{\max}(x))$ ;

в) если,  $y_i < f(x_i)$  то  $(x_i, y_i)$  является реализацией вектора, распределённого в области  $G$  (между осью  $OX$  и  $f(x)$ ) и  $x_i$  – искомая реализация  $\xi$ . Иначе процедура повторяется, начиная с пункта б).

Так как в качестве базовой последовательности обычно берут числа, равномерно распределенные в диапазоне от 0 до 1, то от случайной величины  $\xi$  с помощью функции распределения  $f_\xi$  переходят к случайной величине

$$\xi^* = \frac{\xi - a}{b - a}.$$

Для нее область возможных значений  $(0, 1)$ , а функция плотности вероятностей имеет вид

$$f_{\xi^*}(z) = (b-a)f_\xi[a+(b-a)z]$$

(сжатие по оси абсцисс).

Произведем сжатие по оси ординат:

$$f_{\xi}^{**}(z) = f_{\xi}^{*}(z)/f_{\max},$$

где  $f_{\max}$  максимальное значение  $f_{\xi}^{*}(z)$ .

В результате произведенных действий функция  $f_{\xi}^{**}(z)$  расположилась в единичном квадрате в начале системы координат.

Теперь процедура получения последовательности чисел с функцией плотности  $f_{\xi}(x)$  сводится к следующему.

- 1) генерируется пара РРСЧ из диапазона  $(0, 1) - \eta_i, \eta_{i+1}$ ;
- 2) проверяется выполнение условия

$$\eta_{i+1} \leq f_{\xi}^{**}(\eta_i);$$

- 3) если это условие выполняется, то очередное число, включаемое в выходную последовательность получаем как

$$x_j = a + (b-a) \eta_i,$$

иначе процедура повторяется, начиная с пункта 1.

Эффективность метода исключения характеризуется коэффициентом использования  $k_n$  – отношением количества отобранных реализаций к общему числу реализаций. Очевидно, что это отношение при равномерном распределении реализаций равно отношению площадей под графиками функций  $f(x)$  и  $g_1(x)$ :

$$k_n = \frac{G}{G_1} \leq 1,$$

где  $G_1 = \int_{-\infty}^{\infty} g_1(x) dx$ , а  $G = \int_{-\infty}^{\infty} f(x) dx$ .

Чем ближе значение  $k_n$  к 1, тем эффективнее использование метода. Можно, например, использовать в качестве мажорирующей функции  $g_1(x)$  ступенчатую функцию (рис. 3.11). При этом значение коэффициента  $k_n$  увеличивается, но следует иметь в виду, что при этом одновременно увеличивается число вычислений.

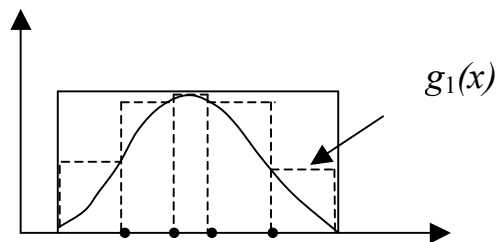


Рис. 3.11. Ступенчатая мажорирующая функция

### Пример. 3.1.

Рассмотрим использование метода для получения чисел с треугольным распределением:

$$f_{\xi}(x) = \begin{cases} \frac{2(x-a)}{(b-a)^2} & x \in [a, b]; \\ 0 & x \notin [a, b]. \end{cases}$$

Переходим от случайной величины  $\xi$  к величине  $\xi^*$ , реализацией которой является значение

$$z = \frac{x-a}{b-a}.$$

Значения  $z$  лежат в диапазоне от 0 до 1 (рис. б) и для нее функция плотности распределения

$$f_{\xi^*}(z) = (b-a)f_{\xi}[a+(b-a)z] = (b-a) \frac{2(a+(b-a)z-a)}{(b-a)^2} = 2z.$$

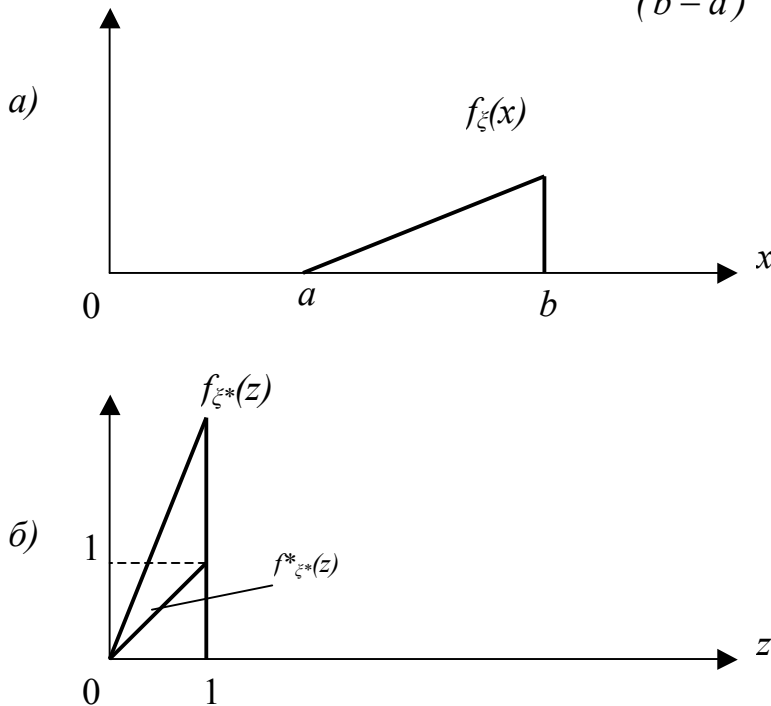


Рис. 3.12. Имитация треугольного распределения  
Максимальное значение эта функция принимает при  $z=1$ :

$$f_{\max}=2.$$

Произведем масштабирование по вертикальной оси:

$$f_{\xi^*}^*(z) = f_{\xi^*}(z)/f_{\max} = z.$$

Таким образом, мы пришли к функции, вписанной в единичный квадрат (рис. 3.12. б).

Теперь формирование чисел с треугольным распределением будем производить по следующему алгоритму:

- 1) Берём  $\eta_1$  и  $\eta_2$  – равномерно распределенные в диапазоне от 0 до 1 числа;
- 2) Если  $\eta_1 > \eta_2$ , то включаем в выходную последовательность очередное число  $x_j = a + (b-a) \eta_1$ , иначе переходим к пункту 1.

Для ускорения процедуры можно заполнять квадрат сразу парами точек, симметричными относительно диагонали с координатами  $(\eta_1, \eta_2)$  и  $(\eta_2, \eta_1)$ .

Тогда пункт 2 будет иметь вид:  $x_j = a + (b-a) \max((\eta_1, \eta_2))$ .

### 3.5.4. Метод композиции (суперпозиции).

Этот метод может быть применён, когда применение метода обратной функции или метода исключения становится затруднительным ввиду сложности функции.

Суть метода заключается в разбиении фигуры, образуемой графиком, функции  $f_\xi(x)$  на произвольное число непересекающихся областей, форма которых позволяет использовать один из ранее рассмотренных методов формирования случайных величин (рис. 3.13).

Другими словами, осуществляется аппроксимация функции  $f_\xi(x)$  композицией более простых функций в виде

$$f_\xi(x) = \sum_{j=1}^{\infty} P_j \varphi_j(x), \text{ где } \sum_{j=1}^{\infty} P_j = 1.$$

Значение  $P_j$  фактически представляет собой площадь фигуры  $q_j$ .

При практической реализации метода число компонент в композиции конечно:

$$f_\xi(x) = \sum_{j=1}^N P_j \varphi_j(x).$$

Здесь  $\varphi_j(x)$  – условные плотности вероятностей, соответствующие форме  $q_j$ . Их ординаты получаются делением на  $P_j$  (нормированием) отрезков вертикальных прямых, лежащих в области  $q_j$ .

Имитация реализаций  $\xi$  сводится в этом случае к реализации дискретной величины  $j$ , распределение которой задано рядом вероятностей

$$P_1, P_2, \dots, P_N,$$

т.е. к выбору одной из  $\varphi_j$  и к имитации величины с плотностью  $\varphi_j(x)$  одним из известных способов.

Основной принцип разбиения заключается в том, что частям  $q_i$ , имеющим наибольшую площадь (т.е. наибольшее значение  $P_i$ ) должны соответствовать наиболее просто и быстро имитируемые плотности  $\varphi_i(x)$ .

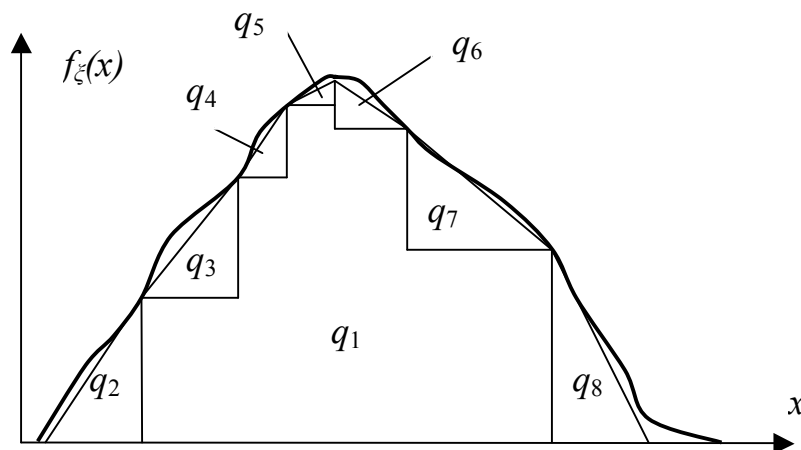


Рис. 3.13. Метод композиции

Например, разбиению на рис. соответствуют:

- ступенчатая плотность  $\varphi_1$  (хорошо применим универсальный метод);
- треугольные плотности  $\varphi_2 - \varphi_7$  (наиболее удобен метод исключений);
- остаточная плотность:

$$\varphi_8 = \frac{f(x) - \sum_{i=1}^7 P_i \varphi_i}{1 - \sum_{i=1}^7 P_i}.$$

Таким образом, моделирующий алгоритм включает следующие этапы:

— разбиение области под графиком  $f_\xi(x)$  на  $N$  непересекающихся областей  $q_j$  достаточно простой формы и определение их площадей ;

— для каждой  $q_j$  строится условная плотность  $\varphi_j(x)$ , ординаты которой получаются делением на  $P_j$  отрезков вертикальных прямых  $x=\text{const}$ , лежащих в области  $q_j$ ;

— строится шкала  $0, a_1, a_2, \dots, a_N=1$ , где  $a_k = \sum_{i=1}^k P_i$  ;

— генерируется равномерно распределенное число  $\eta$  из диапазона от 0 до 1 и с его помощью определяется номер  $k$  плотности  $\varphi_k(x)$  по условию:

$$a_{k-1} < \eta < a_k;$$

— с использованием одного из рассмотренных ранее методов формируется случайное число  $x_i$  с функцией плотности распределения  $\varphi_k(x)$ , которое включается в выходную последовательность.

Если чисел нужно больше, то последние два пункта повторяются необходимое число раз.

### 3.6. Формирование случайных векторов с заданными вероятностными характеристиками

Случайный вектор можно задать проекциями на оси координат. Эти проекции являются случайными величинами, описываемыми совместным законом распределения.

В том случае, когда значения координат  $N$ -мерного вектора представляют собой независимые случайные величины, формирование векторов сводится к раздельному формированию  $N$  случайных чисел с использованием рассмотренных ранее способов. При наличии корреляции между значениями координат задача усложняется.

В простейшем случае вектор двумерный и может быть задан совместным законом распределения его проекций  $\xi$  и  $\eta$  на оси  $OX$  и  $OY$ .

Рассмотрим способы формирования случайных векторов для моделирования дискретных и непрерывных случайных процессов на примере двумерных векторов. При большей размерности содержание рассматриваемых ниже процедур сохраняется, лишь удлиняется цепочка действий.

#### 1. Дискретный случайный процесс.

Пусть требуется смоделировать случайный вектор  $(\xi, \eta)$ . Составляющая  $\xi$  может принимать значения  $x_1, x_2, \dots, x_m$ , а составляющая  $\eta$  – значения  $y_1, y_2, \dots, y_n$  и каждой паре  $(x_i, y_j)$  соответствует вероятность  $P_{ij}$ . Совместную функцию распределения вероятностей зададим в виде матрицы

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & \dots & P_{mn} \end{bmatrix}, \quad \sum_{i=1}^m \sum_{j=1}^n P_{i,j} = 1.$$

Тогда каждому возможному значению  $x_i$  случайной величины  $\xi$  будут соответствовать вероятность

$$P_i = \sum_{j=1}^n P_{ij},$$

представляющая собой сумму вероятностей  $i$ -й строки матрицы.

Набор значений  $P_1, P_2, \dots, P_m$  представляет собой одномерный закон распределения вероятностей для первой координаты. Используя его, можно разыграть значение координаты  $\xi = x_i$ , генерируя равномерно распределенное в диапазоне от 0 до 1 число  $\beta_1$  и определив значение  $i$  по правилу:

$$\alpha_{i-1} < \beta_1 < \alpha_i,$$

где  $\alpha_0=0$ ,  $\alpha_i = \sum_{k=1}^i P_k$ .

Из всех значений матрицы  $P$  выбираем последовательность  $P_{i1}, P_{i2}, \dots, P_{in}$  ( $i$ -ю строку) и преобразуем её:

$$P_{ij}^* = \frac{P_{ij}}{P_i} \quad (j=1,2, \dots, n).$$

Эта последовательность описывает условное распределение величины  $\eta$  при условии, что  $\xi = x_i$ .

Используя тот же прием, что и при розыгрыше значения первой координаты, определяем с помощью равномерно распределенного в диапазоне от 0 до 1 числа  $\beta_2$  конкретное значение  $y_j$  случайной величины  $\eta$ . Получена первая реализация  $(x_i, y_j)$  вектора. Если нужно большее количество реализаций, процедуру повторяем циклически.

## 2. Непрерывный случайный процесс.

В этом случае двумерная случайная величина  $(\xi, \eta)$  описывается совместной функцией плотности распределения  $f_{\xi\eta}(x, y)$ .

Зная закон распределения системы двух случайных величин, можно определить одномерный закон распределения одной из них

$$f_{\xi}(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Имея этот закон и, используя один из рассмотренных ранее методов (например, метод обратной функции), можно сформировать случайное число  $x_i$ , а затем при условии, что  $\xi = x_i$  определить условное распределение случайной величины  $\eta$ :

$$f_{\eta}(y|\xi = x_i) = f(x_i, y) / f_{\xi}(x_i).$$

В соответствии с этой плотностью, можно определить случайное число  $y_j$  и получить реализацию вектора  $(x_i, y_j)$ .

Аналогичным образом можно моделировать случайные вектора и большей размерности. Например, если вектор трехмерный и задан совместной функцией плотности  $f_{\xi\eta\zeta}(x, y, z)$ , то значения случайных чисел  $x_i, y_j, z_k$  выбираются в соответствии с функциями плотности:

$$f_{\xi}(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) dy dz,$$

$$f_{\eta}(y|\xi = x_i) = \int_{-\infty}^{\infty} (f(x_i, y, z) / f_{\xi}(x_i)) dz,$$

$$f_{\zeta}(z|\xi = x_i, \eta = y_i) = f(x_i, y_i, z) / f_{\xi}(x_i) f_{\eta}(y_i | \xi = x_i).$$

## 3.7. Моделирование случайных событий

Простейшими случайными объектами при статистическом моделировании систем являются *случайные события*. Рассмотрим особенности их моделирования. Будем при этом предполагать, что в нашем распоряжении

имеется последовательность  $\{\eta_i\}$  псевдослучайных чисел, равномерно распределенных в интервале  $(0, 1)$ .

### 1) Простое событие.

Пусть надо реализовать случайное событие  $A$ , наступающее с вероятностью  $P$ . Будем определять его как выполнение условия

$$\eta_i \leq P.$$

Противоположное событие  $\bar{A}$  состоит в том, что  $\eta_i$  удовлетворяет неравенству

$$\eta_i > P.$$

Процедура моделирования в этом случае состоит в выборе значений  $\eta_i$  и сравнении их с  $P$ , по результатам которого дается вывод о том, какое событие произошло –  $A$  или  $\bar{A}$ .

### 2) Полная группа несовместных событий

Если моделируется полная группа несовместных событий  $A_1, A_2, \dots, A_s$ , наступающих с вероятностями  $P_1, P_2, \dots, P_s$  ( $\sum_{i=1}^s P_i = 1$ , события не могут произойти одновременно), то наступление события  $A_m$  будем определять как то, что выбранное значение  $\eta_i$  удовлетворяет условию

$$q_{m-1} < \eta_i \leq q_m,$$

где  $q_m = \sum_{i=1}^m P_i$ ,  $q_0 = 0$ .

Процедура моделирования испытаний в этом случае состоит в последовательном сравнении случайных чисел  $\eta_i$  со значениями  $q_m$ . Исходом испытания оказывается событие  $A_m$ , если выполняется данное условие. Эту процедуру называют *определением исхода испытания по жребию* в соответствии с вероятностями  $P_1, P_2, \dots, P_s$ .

Очевидно, что эта же процедура может быть использована для реализации дискретной случайной величины  $\xi$ , принимающей конечное число возможных значений  $x_1, x_2, \dots, x_s$  с вероятностями  $P_1, P_2, \dots, P_s$ . Для дискретной случайной величины, принимающей бесконечное (счетное) число возможных значений, этот путь позволяет получить приближенное решение задачи.

Пусть дискретная случайная величина  $\xi$  принимает счетное множество возможных значений  $x_k$  ( $k=1, 2, \dots$ ), а  $P_k$  задается соотношением

$$P_k = P(x_k), \quad x_1 < x_2 < \dots < x_n < \dots,$$

где  $\sum_{i=1}^{\infty} P_k = 1$ .



Возьмем очередное РРСЧ  $\eta_i$  из базовой последовательности и будем формировать сумму  $\sum_{i=1}^r P_i$  до тех пор, пока не станет справедливым неравенство:

$$\eta_i \leq \sum_{i=1}^r P_i.$$

Тогда считаем, что очередным значением случайной величины  $\xi$  будет  $x_r$ .

### 3) Сложные события

Сложное событие – это событие, зависящее от двух или более простых событий.

Пусть в основе сложного события лежат два простых независимых события  $A$  и  $B$  с вероятностями  $P_A$  и  $P_B$ . исходами совместных испытаний в этом случае будут события

$$AB, \quad A\bar{B}, \quad \bar{A}B, \quad \bar{A}\bar{B}$$

с вероятностями

$$P_AP_B, \quad P_A(1-P_B), \quad (1-P_A)P_B, \quad (1-P_A)(1-P_B).$$

Для моделирования совместных испытаний можно использовать два варианта процедуры:

а) последовательную проверку условия наступления простых событий  $A$  и  $B$  и принятие решения о исходе сложного события по результатам этих проверок;

б) определение одного из исходов, считая, что они представляют собой полную группу несовместных событий с соответствующими вероятностями.

Первый вариант требует двух чисел  $\eta_i$  и сравнений. При втором варианте можно обойтись одним числом  $\eta_i$ , но сравнений может потребоваться больше. С точки зрения удобства построения моделирующего алгоритма и экономии количества операций и ячеек памяти ЭВМ более предпочтителен первый вариант.

Рассмотрим теперь случай, когда события  $A$  и  $B$  являются зависимыми и наступают с вероятностями  $P_A$  и  $P_B$ . Обозначим через  $P(B|A)$  условную вероятность наступления события  $B$  при условии, что событие  $A$  произошло. При этом считаем, что условная вероятность  $P(B|A)$  задана.

Рассмотрим один из вариантов построения модели. Из базовой последовательности случайных чисел извлекается очередное число  $\eta_i$  и проверяется справедливость неравенства  $\eta_i < P_A$ . Если это неравенство справедливо, то наступило событие  $A$ . Для испытания, связанного с событием  $B$ , используется вероятность  $P(B|A)$ . Из совокупности чисел  $\{\eta_i\}$  берется очередное число  $\eta_{i+1}$  и

проверяется условие  $\eta_i < P(B|A)$ . В зависимости от того, выполняется или нет это неравенство, исходом испытания являются  $AB$  или  $A\bar{B}$ .

Если неравенство  $\eta_i < P_A$  не выполняется, то наступило событие  $A$ . Поэтому для испытания, связанного с событием  $B$ , необходимо определить вероятность  $P(B|\bar{A})$ . Это можно сделать, используя формулу полной вероятности

$$P_B = P_A P(B|A) + (1 - P_A) P(B|\bar{A}),$$

откуда

$$P(B|\bar{A}) = \frac{P(B) - P(A)P(B|A)}{1 - P(A)}.$$

### 3. 8. Сетевые модели

Этапу программной реализации модели (т. е. ее описанию на одном из языков программирования) должен предшествовать так называемый этап алгоритмизации. Другими словами, прежде чем превратить имитационную модель в работающую компьютерную программу, ее создатель должен воспользоваться каким-то менее формальным и более наглядным средством описания логики работы будущей программы. Разумеется, это требование не является обязательным для всех разработчиков и для всех создаваемых моделей: при наличии достаточного опыта программа не очень сложной модели может быть написана сразу. Для описания логики работы модели могут быть использованы различные средства: либо русский язык (устный или письменный), либо традиционные схемы алгоритмов, либо какие-то другие «подручные» средства. Первые два варианта являются, как правило, наиболее знакомыми и наиболее часто используемыми. Однако, если попробовать описать в виде схемы алгоритма модель достаточно сложной системы, то это окажется напрасной тратой времени и сил. Прежде всего потому, что такие схемы совершенно не приспособлены для описания параллельных процессов.

В этом случае удобно использовать весьма распространенное средство описания параллельных процессов — сети Петри, начало которым положил немецкий ученый Карл Петри, предложивший в 1962 г. этот класс сетей в качестве модели информационных потоков для изучения конечных автоматов.

#### 3.8.1. Сети Петри

Одно из основных достоинств аппарата сетей Петри заключается в том, что они могут быть представлены как в графической форме (что обеспечивает наглядность), так и в аналитической (что позволяет автоматизировать процесс их анализа).

При графической интерпретации сеть Петри представляет собой граф особого вида, состоящий из вершин двух типов — *позиций* и *переходов*, соединенных ориентированными дугами, причем каждая дуга может связывать лишь разнотипные вершины (позицию с переходом или переход с позицией). Вершины-позиции обозначаются кружками, вершины-переходы — черточками. С содержательной точки зрения, переходы соответствуют событиям, присущим исследуемой системе, а позиции — условиям их возникновения. Таким образом, совокупность переходов, позиций и дуг позволяет описать причинно-следственные связи, присущие системе, но в статике. Чтобы сеть Петри «оживила», вводят еще один вид объектов сети — так называемые *фишки*, или *метки* позиций. Переход считается активным (событие может произойти), если в каждой его входной позиции есть хотя бы одна фишка.

Расположение фишек в позициях сети называется *разметкой сети*.

В аналитической форме сеть Петри может быть представлена совокупностью следующих элементов:

$$P = (B, D, I, O, M),$$

где  $B = \{b_i\}$  — конечное непустое множество позиций;

$D = \{d_i\}$  — конечное непустое множество переходов;

$I: B \times D \rightarrow 0, 1$  — входная функция (прямая функция инцидентности), которая для каждого перехода задает множество его входных позиций;

$O: D \times B \rightarrow 0, 1$  — выходная функция (обратная функция инцидентности), которая для каждого перехода задает множество его выходных позиций;

$M$  — функция разметки сети,  $M: B \rightarrow 0, 1, 2, \dots$  — ставит каждой позиции сети в соответствие неотрицательное целое число.

С учетом введенных обозначений необходимое условие срабатывания перехода  $d_j$  может быть записано следующим образом:

$$\forall b_i \in I(d_j) \{ M(b_i) \geq 1 \} \text{ (для всех входных позиций разметка должна быть } \geq 1 \text{)}.$$

Срабатывание перехода  $d_i$  изменяет разметку сети  $M(B)$  на разметку по следующему правилу:

$$M_1(B) = M(B) - I(d_i) + O(d_i),$$

то есть переход  $d_j$  изымает по одной метке из каждой своей входной позиции и добавляет по одной метке в каждую из выходных позиций (рис. 3.14).

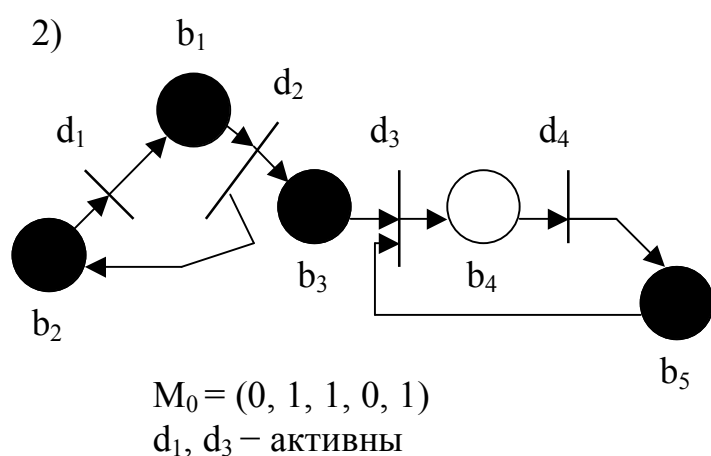
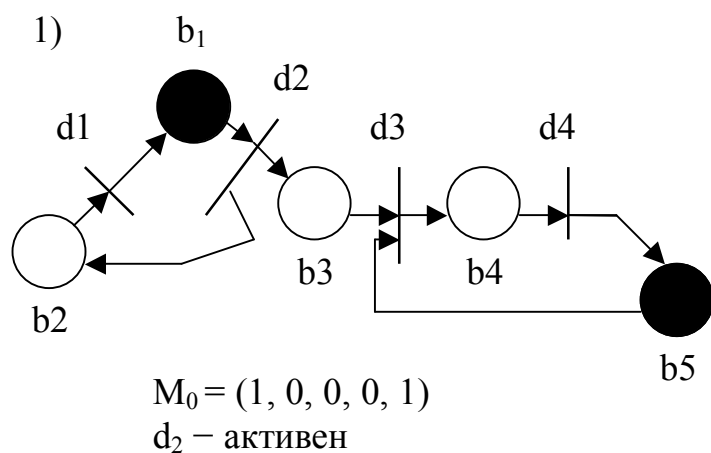


Рис.3.14. Пример изменения разметки сети Петри при срабатывании переходов

Входная и выходная функции сети Петри ( $I$  и  $O$ ) позволяют описать любую сеть с помощью двух матриц размера  $m \times n$  (матриц входных и выходных позиций), имеющих следующую структуру (таб. 3.1):

Основные направления анализа сети Петри следующие:

1. Проблема достижимости: в сети Петри с начальной разметкой  $M_0$  требуется определить, достижима ли принципиально некоторая разметка  $M_i$  из  $M_0$ .

С точки зрения исследования моделируемой системы, эта проблема интерпретируется как проблема достижимости (реализуемости) некоторого состояния системы.

Таблица 3.1.

	$d_1$	$d_2$	$d_j$	$d_n$
$b_1$	0	1	0	0

$b_2$	1	1	0	1
			...	
$b_i$	0	1	0	1
			...	
$b_m$	1	0	1	0

2. *Свойство живости.* Под живостью перехода понимают возможность его срабатывания в данной сети при начальной разметке  $M_0$ .

Анализ модели на свойство живости позволяет выявить невозможные состояния в моделируемой системе (например, неисполняемые ветви в программе).

3. *Безопасность сети.* Безопасной является такая сеть Петри, в которой ни при каких условиях не может появиться более одной метки в каждой из позиций.

Для исследуемой системы это означает возможность функционирования ее в стационарном режиме. На основе анализа данного свойства могут быть определены требования к буферным накопителям в системе.

Итак, достоинства сетей Петри заключаются в том, что они:

- 1) позволяют моделировать параллельные процессы всех возможных типов с учетом вероятных конфликтов между ними;
- 2) обладают наглядностью и обеспечивают возможность автоматизированного анализа;
- 3) позволяют переходить от одного уровня детализации описания системы к другому (за счет раскрытия/закрытия переходов).

Вместе с тем, сети Петри имеют ряд недостатков, ограничивающих их возможности. Основной из них — время срабатывания перехода считается равным 0, что не позволяет исследовать с помощью сетей Петри временные характеристики моделируемых систем.

### 3.8.2. Е-сети

В результате развития аппарата сетей Петри был разработан ряд расширений. Одно из наиболее мощных — так называемые Е-сети (evaluation — «вычисления», «оценка») — «оценочные сети». Структурно Е-сеть представляет собой граф, состоящий, как и сеть Петри, из двух типов вершин: позиций и переходов, соединенных друг с другом ориентированными дугами, причем каждая дуга может связывать лишь позицию с переходом или, наоборот, переход с позицией. Следовательно, структура Е-сети, как и структура сети Петри, эквивалентна ориентированному двудольному графу, у которого одно множество вершин содержит только позиции, а другое множество вершин — переходы.

В отличие от сетей Петри, в Е-сетях:

- имеются несколько типов вершин-позиций: простые позиции, позиции-очереди, разрешающие позиции;
- фишки (метки) могут снабжаться набором признаков (атрибутов);
- с каждым переходом может быть связана ненулевая задержка и функция преобразования атрибутов фишек.
- введены дополнительные виды вершин-переходов;
- в любую позицию может входить не более одной дуги и выходить также не более одной.

В Е-сетях существуют следующие позиции: простые (одноместные), позиции-очереди (многоместные) и разрешающие. При графическом представлении простая позиция изображается кружком, позиция-очередь — овалом и разрешающая позиция — шестиугольником или для упрощения вычерчивания квадратиком. Переход в Е-сетях изображается, как и в сетях Петри, отрезком прямой линии.

Структуру любой конечной Е-сети формально можно представить той же пятеркой, что и сеть Петри:

$$P = (B, D, I, O, M),$$

Назовем любую дугу, соединяющую позицию с переходом или переход с позицией, структурной связкой. Совокупность, состоящую из перехода  $H$  и структурных связок, в которых участвует переход  $H$ , назовем элементарной сетью  $E_H$ , отвечающей переходу  $H$ . При этом позиции, из которых дуги направлены на переход  $H$ , называются входными позициями, а те позиции, на которых заканчиваются дуги, начинающиеся на переходе  $H$ , — выходными позициями этого перехода. Некоторые позиции могут быть одновременно входными и выходными для одного и того же перехода. В элементарной сети, отвечающей некоторому переходу, его входные и выходные позиции являются соответственно входными и выходными позициями этой сети.

В каждой элементарной сети может быть не более одной разрешающей позиции, причем такая позиция должна быть входной. Разрешающая позиция, если она имеется, выполняет, как отмечалось, управляющую функцию при срабатывании перехода и характеризуется конечным множеством состояний  $\{0, 1, 2, \dots\}$ , зависящим от элементарной сети, и разрешающей процедурой. Разрешающая процедура предназначена для вычисления состояния разрешающей позиции. Состояние разрешающей позиции, равное нулю, считается неопределенным. Множество элементарных сетей можно разбить на классы.

На рис. 3.15 показан базовый набор классов, или типов, элементарных сетей, достаточный для моделирования вычислительных процессов в операционных системах. На этом рисунке представлены шесть типов элементарных сетей, обозначенные буквами Т, J, F, X, Y, I. Для каждого типа показано графическое представление соответствующей элементарной сети, причем буквой  $x$  (без индекса или с индексом) обозначена простая входная

позиция, буквой  $y$  (без индекса или с индексом) — простая выходная позиция, буквой  $r$  — разрешающая позиция, которая может быть только входной. При необходимости любая простая позиция может быть заменена позицией-очередью. Такая позиция, как уже упоминалось, изображается овалом.

Переход  $H$ , отвечающий элементарной сети  $E_H$ , описывается тройкой:

$$H = (D, L, Z),$$

где  $D$  — тип элементарной сети, которой принадлежит переход;  $L$  — процедура временной задержки;  $Z$  — процедура преобразования.

Тип  $D$  показывает, к какому классу в выбранном фиксированном наборе классов относится элементарная сеть, и определяет правила срабатывания перехода и перемещения фишек из входных позиций в выходные в результате срабатывания перехода.

Процедура  $L$  задает интервал времени, в течение которого длится фаза активности сработавшего перехода. И, наконец, процедура  $Z$  отражает действия, которые должны выполняться над атрибутами фишек, перемещающихся из входных позиций в выходные в конце фазы активности сработавшего перехода.

Соединяя друг с другом элементарные сети из базового набора типов, можно создавать Е-сети произвольной сложности. Точками соединения элементарных сетей служат простые позиции и. позиции-очереди. При связывании двух элементарных сетей выходная позиция одной элементарной сети сливается с выбранной входной позицией другой элементарной сети, в результате чего получается единственная позиция. Объединяемые позиции обеих элементарных сетей должны быть одного типа, т. е. простыми или позициями-очередями. Тип результирующей позиции совпадает с типами объединяемых позиций.

Любая простая позиция или позиция-очередь, содержащая фишку, называется маркированной. Фишка — это некоторый абстрактный объект, который может характеризоваться конечным упорядоченным набором признаков, или атрибутов. Смысл атрибутов фишки зависит от характера моделируемой системы и не влияет на поведение Е-сети.

Начальная маркировка позиций в Е-сети задается, как отмечалось в формальном определении структуры сети, функцией разметки  $M$ . При этом в отличие от сетей Петри никакая простая позиция не может содержать более одной фишки. Позиция-очередь может хранить произвольное число фишек, упорядоченных в соответствии с дисциплиной, являющейся характеристикой этой позиции. Функция разметки определена и для разрешающих позиций, но в этом случае ее значение трактуется как состояние позиции, а не как наличие фишки в этой позиции.

Динамика Е-сети определяется срабатываниями переходов и перемещениями фишек из входных позиций переходов в выходные позиции в результате срабатывания. Одновременно могут срабатывать переходы и,

следовательно, перемещаться фишки во многих элементарных сетях данной Е-сети. Тем самым обеспечивается возможность представления параллельных процессов в моделируемой системе, так как с каждой перемещающейся фишкой может быть ассоциирован определенный процесс. При этом перемещение фишки из входной позиции некоторого перехода в выходную позицию сопровождается в общем случае модификацией атрибутов фишки с помощью процедуры преобразования  $Z$  и соответствует изменению состояния соответствующего процесса. Длительность интервала изменения состояния определяется процедурой временной задержки  $L$ .

Для срабатывания перехода, отвечающего некоторой элементарной сети, необходимо, чтобы маркировка его входных и выходных позиций удовлетворяла определенному условию, зависящему от типа элементарной сети.

Для элементарных сетей типов  $X$  и  $Y$  требуется, кроме того, чтобы в результате вычисления разрешающей процедуры получилось ненулевое состояние разрешающей позиции данной элементарной сети. И только вслед за этим может сработать переход, вступив в фазу активности. Длительность этой фазы, как уже отмечалось, диктуется процедурой задержки  $L$ . В конце фазы активности перехода осуществляется перемещение фишек из входных позиций в выходные, а затем над атрибутами перемещенных фишек выполняется процедура преобразования.

Каждая простая позиция имеет память для хранения атрибутов фишки, которую можно представить вектором ячеек памяти:

$$A = (a_1, a_2, \dots, a_n),$$

где  $a_i$ , — содержимое  $i$ -й ячейки,  $i = 1, 2, \dots, n$ .

Если позиция хранит фишку, т. е. маркирована, то  $a_i$ , есть значение  $i$ -го атрибута фишки. При отсутствии фишки значения компонентов вектора  $A$  считаются неопределенными. Назовем число компонентов  $n$  вектора  $A$  емкостью позиции.

В отличие от простой позиции позиция-очередь характеризуется теоретически неограниченной цепочкой векторов памяти. Каждый вектор может хранить одну фишку. Все векторы ячеек памяти позиции-очереди имеют одно и то же число компонентов. Это число представляет собой емкость позиции-очереди (для одной фишки).

Емкости позиций задаются на этапе детальной спецификации Е-сети и могут быть различными у разных позиций. В общем случае емкость позиции и число атрибутов фишки, направляемой в эту позицию, не совпадают. Если число атрибутов фишки больше емкости позиции, в которую поступает фишка, то лишние (справа) атрибуты фишки отбрасываются. Если же, наоборот, число атрибутов фишки меньше емкости позиции, то в избыточные (справа) ячейки памяти этой позиции при поступлении фишки стандартно заносятся нулевые значения.



Рассмотрим правила срабатывания переходов в элементарных сетях базового набора типов.

*Элементарная сеть типа Т.* Условием срабатывания перехода в элементарной сети типа Т служит наличие фишки во входной позиции и отсутствие фишки в выходной позиции. Стандартной процедурой преобразования является копирование атрибутов фишки из входной позиции в выходную по принципу «урезания» лишних атрибутов или добавления новых с нулевыми значениями. Этот принцип используется в стандартных процедурах преобразования и всех остальных типов элементарных сетей.

*Элементарная сеть типа J.* В элементарной сети этого типа переход срабатывает, если есть фишка в каждой входной позиции и свободна выходная позиция. В конце фазы активности сработавшего перехода в выходной позиции появляется фишка, а все входные позиции освобождаются. Стандартной процедурой преобразования является копирование атрибутов фишки из единственной входной позиции  $x_1$  (см. рис. 3.15) в выходную позицию. Таким образом, в стандартной процедуре преобразования значения атрибутов фишек, находящихся во всех входных позициях, кроме позиции  $x_1$ , не влияют на атрибуты фишки, попадающей в выходную позицию.

*Элементарная сеть типа F.* Переход срабатывает, если есть фишка во входной позиции и свободны все выходные позиции. В конце фазы активности сработавшего перехода фишка перемещается из входной позиции сразу во все выходные позиции, т. е. размножается. Стандартной процедурой преобразования при этом является копирование атрибутов фишки из входной позиции во все выходные позиции. Так как емкости выходных позиций могут быть различны, то различаются в общем случае и результаты копирования в разных выходных позициях.

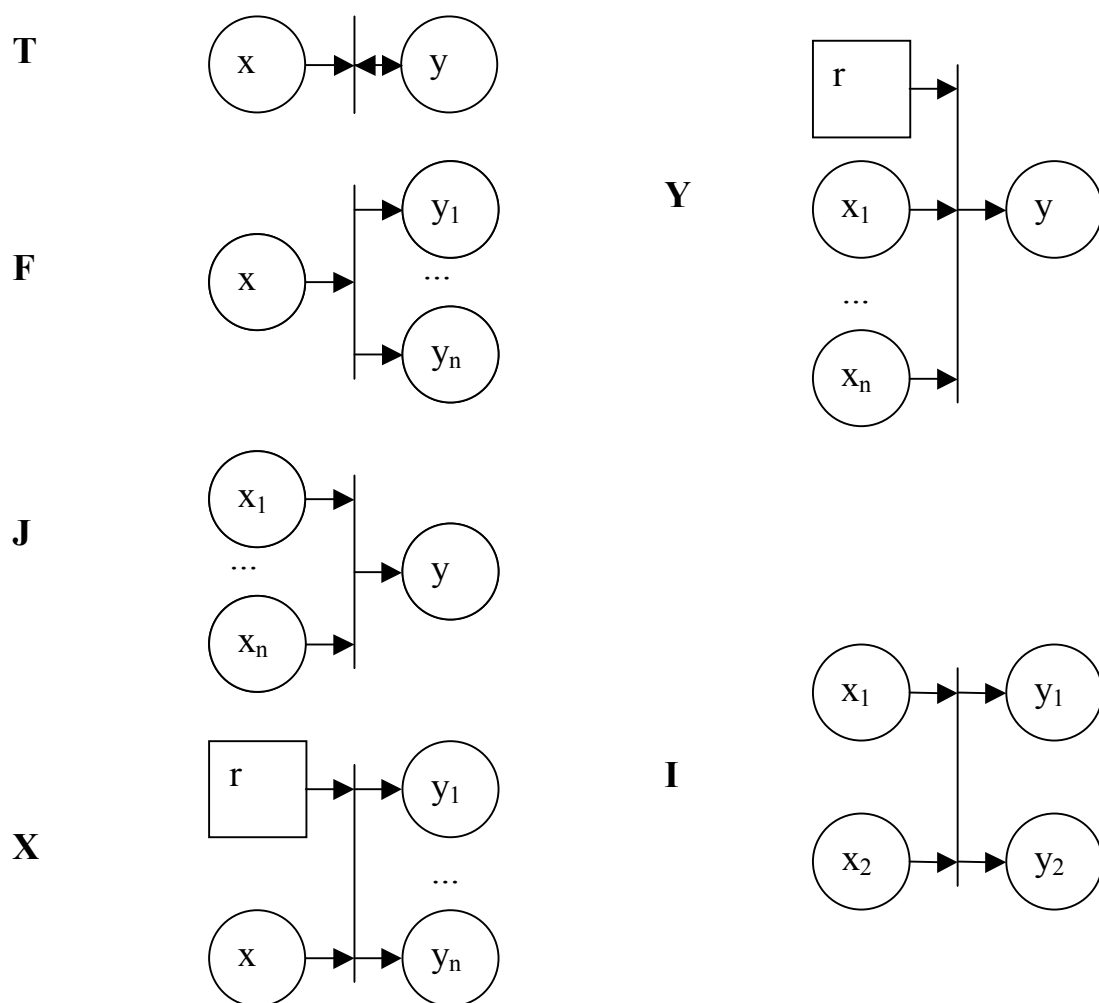


Рис. 3.15. Базовый набор типов элементарных сетей

*Элементарная сеть типа X.* Пусть, как показано на рис. 3.15., выходные позиции перехода в элементарной сети данного типа связаны с переходом с помощью дуг, пронумерованных в порядке 1, 2, ...,  $n$ , начиная сверху. Необходимым (но не достаточным) условием срабатывания перехода в сети данного типа является наличие фишки во входной позиции  $x_1$ . При этом условии должна быть вычислена разрешающая процедура, ассоциированная с разрешающей позицией  $r$ . Пусть в результате этого вычисления будет определено состояние разрешающей позиции, равное целому числу  $k$  в диапазоне от 1 до  $n$ . Если при этом выходная позиция  $y_k$  пуста, то переход срабатывает. В конце фазы его активности фишка перемещается из входной позиции  $x_1$  в выходную  $y_k$ .

Стандартной процедурой преобразования является копирование атрибутов фишки из позиции  $x_1$  в позицию  $y_k$ . Содержимое всех остальных выходных позиций не изменяется в результате срабатывания перехода.

Если вычисленное состояние разрешающей позиции окажется неопределенным (при этом  $k = 0$ ), то переход не срабатывает. При работе в составе некоторой Е-сети такой переход помещается в особый список, а разрешающая процедура, ассоциированная с позицией  $r$  этого перехода, вычисляется заново всякий раз после срабатывания каждого другого перехода данной Е-сети. Как только результат очередного вычисления разрешающей процедуры примет целое значение в диапазоне от 1 до  $n$ , переход срабатывает, как описано выше.

Если окажется, что после вычисления разрешающей процедуры позиция  $y_k$  занята фишкой, то переход не срабатывает. И в этом случае при работе в составе Е-сети разрешающая процедура вычисляется заново всякий раз после срабатывания каждого другого перехода Е-сети, пока позиция  $y_k$  не освободится.

Изложенные правила срабатывания перехода в элементарной сети типа Х соответствуют нестандартной разрешающей процедуре. Предусмотрена также стандартная разрешающая процедура, которая задает состояние позиции  $r$ , численно равное номеру первой свободной простой позиции или первой позиции-очереди в списке выходных позиций перехода. Если в этом списке все простые позиции заняты, а позиций-очереди нет, то стандартная разрешающая процедура фиксирует неопределенное состояние позиции  $r$ . Переход в этом случае не срабатывает.

*Элементарная сеть типа Y.* Для срабатывания перехода в сети данного типа необходимо (но не достаточно) наличие фишки хотя бы в одной из входных позиций  $x_1, x_2, \dots, x_n$ , и отсутствие фишки в выходной позиции. При этом условии должна быть вычислена разрешающая процедура, ассоциированная с позицией  $g$ . Пусть в результате этого вычисления будет определено состояние позиции  $g$ , равное целому числу  $k$  в диапазоне от 1 до  $n$ , где  $1, 2, \dots, n$  — последовательные номера дуг, соединяющих входные позиции  $x_1, x_2, \dots, x_n$  с переходом. Если при этом входная позиция  $x_k$  содержит фишку, то переход срабатывает и в конце фазы его активности фишка из позиции  $x_k$  перемещается в выходную позицию.

Если вычисленное состояние позиции  $g$  является неопределенным ( $k=0$ ) или входная позиция  $x_k$  пуста, то переход не срабатывает. При работе в составе Е-сети разрешающая процедура заново вычисляется всякий раз после срабатывания каждого другого перехода Е-сети до тех пор, пока не станет определенным состояние позиции  $g$  или не появится фишка в позиции  $x_k$ .

Стандартная процедура преобразования в сети данного типа осуществляет копирование атрибутов фишки из выбранной входной позиции в выходную.

Предусмотрена также стандартная (т. е. не задаваемая явно) разрешающая процедура. Если выходная позиция пуста и имеется фишка хотя бы в одной из входных позиций  $x_1, x_2, \dots, x_n$ , то позиция  $g$  принимает состояние  $k$ , равное номеру дуги, соединяющей с переходом первую же занятую фишкой входную

позицию (в порядке просмотра сверху вниз). После этого переход срабатывает, как описано выше.

*Элементарная сеть типа I.* Сеть данного типа обеспечивает возможность прерывания фазы активности сработавшего перехода.

Фишка, попадающая во входную позицию  $x_1$ , называется основной, а в позицию  $x_2$  — прерывающей. Переход срабатывает при одном из следующих двух условий:

1. Входная позиция  $x_1$  содержит фишку, а все прочие позиции свободны. Переход срабатывает, начинается фаза его активности. Если в этой фазе не появится фишка в позиции  $x_2$ , то в конце фазы активности фишка перемещается из позиции  $x_1$  в позицию  $y_1$ . Стандартной процедурой преобразования при этом является копирование атрибутов фишки из позиции  $x_1$  в позицию  $y_1$ . Этот случай соответствует срабатыванию перехода без прерывания.

Если же в фазе активности сработавшего перехода появится фишка и в позиции  $x_2$ , то фаза активности перехода немедленно прерывается. Вслед за этим фишка из позиции  $x_1$  перемещается в позицию  $y_1$ , а фишка из позиции  $x_2$  — в позицию  $y_2$ . Стандартная процедура преобразования обеспечивает копирование атрибутов фишки из позиции  $x_1$  в позицию  $y_1$ , а из позиции  $x_2$  в позицию  $y_2$ . Этот случай соответствует срабатыванию перехода с последующим прерыванием.

2. Входная позиция  $x_2$  содержит фишку, а позиции  $x_1$ ,  $y_2$  свободны. Независимо от занятости позиции  $y_1$  переход срабатывает, и немедленно вслед за этим (т. е. с нулевой задержкой времени) фишка из позиции  $x_2$  перемещается в позицию  $y_2$ . Стандартная процедура преобразования в этой ситуации копирует атрибуты фишки из позиции  $x_2$  в позицию  $y_2$ .

В Е-сети все переходы обладают свойством безопасности. Это означает, что в выходных позициях (которые в свою очередь могут быть входными для следующего перехода) никогда не может быть более одной метки.

Вместе с тем, в Е-сетях существуют понятия макроперехода и макропозиции, которые позволяют отображать в модели процессы накопления обслуживаемых транзактов в тех или иных узлах системы.

Рассмотрим некоторые из них (рис. 3.16.)

Макропозиция *очередь* представляет собой линейную композицию вершин-позиций и Т-переходов; количество вершин-позиций определяет «емкость» очереди. Макропозиция *генератор* позволяет представлять в сети источник меток (транзактов). Если необходимо задать закон формирования меток, то «генератор» может быть дополнен разрешающей позицией.

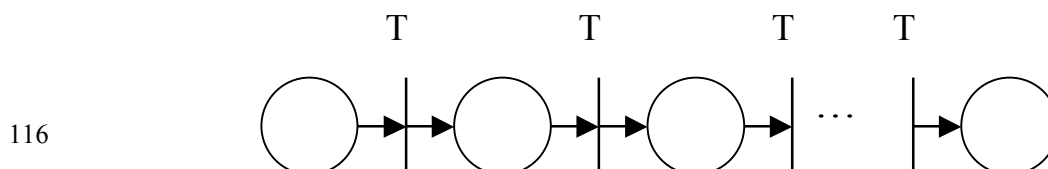


Рис. 3.16. Макропозиции

Поскольку в Е-сети нельзя «накапливать» метки, то вводится макропозиция *поглощения (аккумулятор)*

Аналогичным образом путем композиции  $N$  однотипных переходов могут быть получены макропереходы других типов.

Рассмотренные особенности Е-сетей существенно расширяют их возможности для моделирования дискретных систем вообще и параллельных процессов в частности. Ниже приведен пример описания в виде Е-сети мультипрограммной вычислительной системы (рис. 3.17). Обработка поступающих заданий организована в ней по принципу квантования времени: каждому заданию выделяется равный отрезок (квант) процессорного времени; если задание выполнено, то оно покидает систему, если же времени оказалось

недостаточно, то задание встает в очередь и ждет повторного выделения кванта времени.

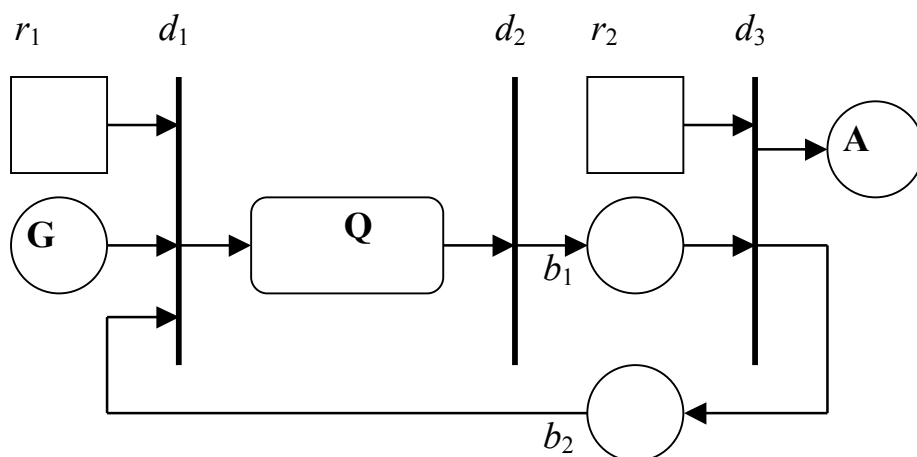


Рис 3.17. Пример описания вычислительной системы в виде Е-сети

На рисунке переходы, соответствующие определенным событиям в системе ( $d_i$ ), имеют следующие обозначения.:

- $d_1$  — постановка задания в очередь;
- $d_2$  — выполнение задания в течение одного кванта времени;
- $d_3$  — анализ степени завершенности задания.

Разрешающие позиции  $r_1$  и  $r_2$  служат для задания закона формирования случайных интервалов времени между поступающими заданиями и интервалов времени, необходимых для полного обслуживания каждого из них. задать закон формирования

Отметим, что технология моделирования систем в виде Е-сетей весьма эффективно реализуется с помощью инструмента SIMULINK, входящего в состав пакета MATLAB.

### 3.8.3. Сетевая модель взаимодействующих параллельных процессов в операционной системе.

Рассмотрим пример применения Е-сетей для имитационного моделирования совокупности взаимодействующих параллельных процессов. Эта совокупность «погружена» в некоторую распределенную операционную систему и обеспечивает формирование потока байтов данных, их последовательную передачу по линии связи удаленному абоненту, прием байтов этим абонентом (с выдачей сигнала уведомления передающей стороне о завершении приема каждого байта) и отображение каждого принятого байта на терминале. Соответственно перечисленным функциям в моделируемой системе можно выделить следующие четыре взаимодействующих процесса:

1. Генерация потока байтов данных;
2. Передача байтов и сигналов уведомления по двум линиям связи;

3. Прием байтов (с формированием сигнала уведомления);
4. Отображение принятых байтов на терминале.

Посылка сигналов уведомления удаленным абонентом передающей стороне на каждый принятый байт не является необычной и предусмотрена аппаратно в ряде существующих микропроцессорных контроллеров, предназначенных для асинхронной передачи данных по последовательной линии связи со скоростью в диапазоне от 50 до 19 200 бод.

Опишем более детально функции и протокол взаимодействия названных четырех процессов. Процесс генерации формирует за определенное время очередной байт данных и подготавливает его к передаче удаленному абоненту. Процесс передачи начинается только после получения сигнала уведомления от удаленного абонента о завершении им приема предыдущего байта. Одновременно с началом передачи байта может начаться генерация следующего байта, которая занимает гораздо меньше времени, чем передача байта по линии связи.

Байт данных, посланный удаленному абоненту, после прохождения по линии связи временно запоминается в буферном регистре абонента. Прием байта завершается процессом приема в момент извлечения байта из буферного регистра, а это происходит после окончания вывода ранее принятого байта на терминал с помощью процесса отображения. Вслед за тем процесс приема формирует сигнал уведомления и направляет его по другой линии связи передающей стороне.

С этого момента принятый байт обрабатывается с помощью процесса отображения, который проверяет сначала готовность терминала к восприятию нового байта и в случае готовности обеспечивает вывод байта на терминал. Затем процесс отображения сообщает процессу приема о том, что может быть принят следующий байт. Получив это сообщение, процесс приема может извлечь из буферного регистра очередной байт, если он уже был передан. Далее описанная совокупность действий повторяется.

На рис. 3.18 приведена структура сетевой модели описанной системы взаимодействующих процессов. Для облегчения понимания модель условно разделена пунктирными линиями на четыре поименованные секции, каждая из которых соответствует одному из рассмотренных процессов. На этом рисунке переходы обозначены целыми числами от 1 до 17, простые позиции — идентификаторами  $S_1, S_2, \dots, S_{22}$ , единственная позиция-очередь — идентификатором  $Q_1$  и разрешающие позиции — идентификаторами  $R_1$  и  $R_2$ . В идентификаторах позиций буква характеризует тип позиции ( $S$  — простая позиция,  $Q$  — позиция-очередь и  $R$  — разрешающая позиция), а цифровая часть — уникальный номер позиции данного типа. Позиции  $Q_1, S_2, S_{11}, S_{19}, S_{22}$  в исходном состоянии модели маркированы (содержат по одной фишке каждая, что обозначено точками внутри символов позиций).

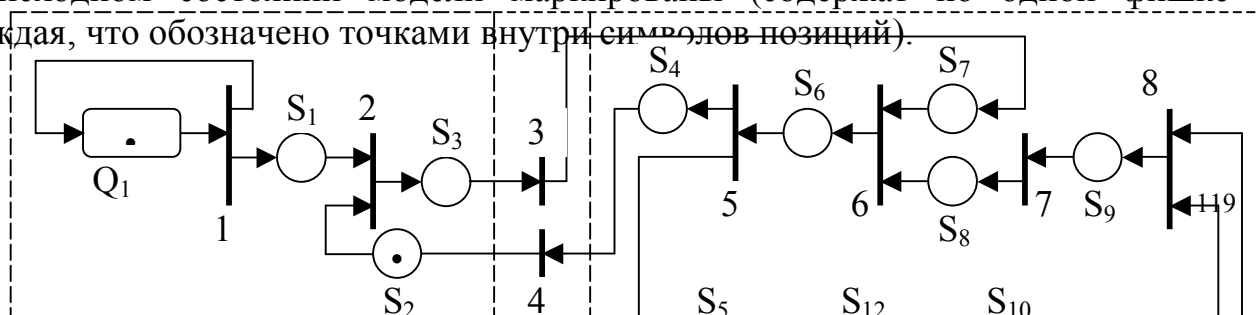


Рис 3.18. Структура сетевой модели системы взаимодействующих процессов

Процесс генерации потока байтов моделируется переходами 1 и 2 с соответствующими входными и выходными позициями. Позиция  $Q_1$ , совместно с переходом 1 представляет собой источник байтов данных. Фишка в позиции  $S_2$  свидетельствует о том, что линия связи для передачи байта удаленному абоненту свободна, и поэтому сразу при появлении фишки в позиции  $S_1$  (а это случится после генерации нового байта) срабатывает переход 2, так как к этому времени позиция  $S_3$  пуста.

Передача байта по линии связи моделируется переходом 3, который срабатывает при появлении фишки в позиции  $S_3$ , в результате чего фишка переместится в позицию  $S_7$ . Это событие является признаком попадания переданного байта в буферный регистр удаленного абонента.

Появление фишки в позиции  $S_{10}$  служит признаком того, что процесс приема получил сообщение от процесса отображения о выдаче предыдущего байта на терминал. Так как позиция  $S_{11}$  занята, то срабатывает сначала переход 8, а потом и переход 7, вследствие чего фишка появится теперь в позиции  $S_8$ , свидетельствуя о готовности процесса



приема извлечь байт из буферного регистра. Эта операция моделируется срабатыванием перехода 6. Затем немедленно сработает и переход 5, а в позициях  $S_4$ ,  $S_5$  появится по одной фишке. Наличие фишки в позиции  $S_4$  есть признак возникновения сигнала уведомления, который по другой линии связи, моделируемой переходом 4, поступает на передающую сторону и появляется там в форме фишки в позиции  $S_2$ . Таким образом, для процесса генерации возникает исходное состояние, с которого было начато описание работы сетевой модели.

Из позиции  $S_5$  фишка после срабатывания переходов 9 и 10, будучи размножена, переместится в позиции  $S_{11}$  и  $S_{13}$ . Появление фишки в позиции  $S_{11}$  означает, что процесс приема вернулся в исходное состояние, являющееся состоянием ожидания сообщения от процесса отображения.

Если в рассматриваемый момент времени позиция  $S_{15}$  содержит фишку (в результате предшествующего срабатывания перехода 17), то сработает переход 11, фишка возникнет в позиции  $S_{14}$ , а из нее после срабатывания перехода 12 с использованием стандартной разрешающей процедуры, ассоциированной с позицией  $R_2$ , переместится в позицию  $S_{16}$ . Траектория дальнейшего движения фишки зависит от состояния позиции  $R_2$ . Это состояние равно 1 или 2 в зависимости от того, готов или не готов терминал к восприятию очередного байта.

О готовности терминала свидетельствует наличие фишки в позиции  $S_{19}$ . Если позиция  $S_{19}$  пуста, то фишка из позиции  $S_{16}$  после срабатывания перехода 13, а затем и перехода 12 вернется в позицию  $S_{16}$ . Этот цикл перемещения моделирует опрос состояния терминала процессором в реальной системе.

Если при очередном попадании фишки в позицию  $S_{16}$  позиция  $S_{19}$  окажется занятой, то из позиции  $S_{16}$  фишка поступит в позицию  $S_{18}$  и сработает переход 14, моделирующий вывод байта на терминал. Затем фишка переместится в позицию  $S_{20}$  и после срабатывания перехода 15 появится в позициях  $S_{22}$  и  $S_{21}$ . Следовательно, для позиции  $S_{22}$  установится исходное состояние.

Из позиции  $S_{21}$  фишка через интервал «восстановления» терминала, моделируемый переходом 16, придет в позицию  $S_{19}$ , установив тем самым признак готовности терминала к восприятию очередного байта.

Описанный характер функционирования сетевой модели не отражается во всех деталях на ее структурной схеме, приведенной на рис. 3.18. В частности, схема не содержит информации об атрибутах фишек, нестандартных процедурах преобразования, процедурах временной задержки, а также о нестандартных разрешающих процедурах. Полностью специфицировать сетевую модель можно, представив ее на языке, в терминах которого были бы описаны все ее структурные, процедурные и информационные аспекты.

### 3.9. Управление модельным временем

Имитационный эксперимент представляет собой наблюдение за поведением системы в течение некоторого промежутка времени. Конечно, далеко не во всех статистических испытаниях фактор времени играет ведущую роль, а в некоторых и вообще может не рассматриваться. Но значительно больше таких задач, в которых оценка эффективности моделируемой системы напрямую связана с временными характеристиками ее функционирования. К ним относятся задачи по оценке производительности, некоторые задачи по оценке надежности, качества распределения ресурсов, а также все задачи, связанные с исследованием эффективности процессов обслуживания. Характерной особенностью большинства практических задач является то, что скорость протекания рассматриваемых в них процессов значительно ниже скорости реализации модельного эксперимента. Например, если моделируется работа авторемонтной мастерской в течение недели, вряд ли кому-то придет в голову воспроизводить этот процесс в модели в таком же масштабе времени.

С другой стороны, даже те имитационные эксперименты, в которых временные параметры работы системы не учитываются, требуют для своей реализации определенных затрат времени работы компьютера.

В связи с этим при разработке практически любой имитационной модели и планировании проведения модельных экспериментов необходимо соотносить между собой три представления времени:

- реальное время, в котором происходит функционирование имитируемой системы;
- модельное (или, как его еще называют, системное) время, в масштабе которого организуется работа модели;
- машинное время, отражающее затраты времени ЭВМ на проведение имитации.

С помощью механизма модельного времени решаются следующие задачи:

- 1) отображается переход моделируемой системы из одного состояния в другое;
- 2) производится синхронизация работы компонент модели;
- 3) изменяется масштаб времени «жизни» (функционирования) исследуемой системы;
- 4) производится управление ходом модельного эксперимента;
- 5) моделируется квазипараллельная реализация событий в модели.

Приставка «квази» в данном случае отражает последовательный характер обработки событий (процессов) в имитационной модели, которые в реальной системе возникают (протекают) одновременно.

Необходимость решения последней задачи связана с тем, что в распоряжении исследователя находится, как правило, однопросессорная вычислительная система, а модель может содержать значительно большее

число одновременно работающих подсистем. Поэтому действительно параллельная (одновременная) реализация всех компонент модели невозможна. Даже если используется так называемая распределенная модель, реализуемая на нескольких узлах вычислительной сети, совсем необязательно число узлов будет совпадать с числом одновременно работающих компонент модели. Следует отметить, что реализация квазипараллельной работы компонент модели является достаточно сложной технической задачей.

Наиболее часто при создании имитационной модели применяют два метода реализации механизма модельного времени – с постоянным шагом (метод  $\Delta t$ ) и по особым состояниям.

Выбор метода реализации механизма модельного времени зависит от назначения модели, ее сложности, характера исследуемых процессов, требуемой точности результатов и т. д.

При использовании метода постоянного шага отсчет системного времени ведется через фиксированные, выбранные исследователем интервалы времени. События в модели считаются наступившими в момент окончания этого интервала (рис.3.18). Погрешность в измерении временных характеристик системы в этом случае зависит от величины шага моделирования  $\Delta t$ .

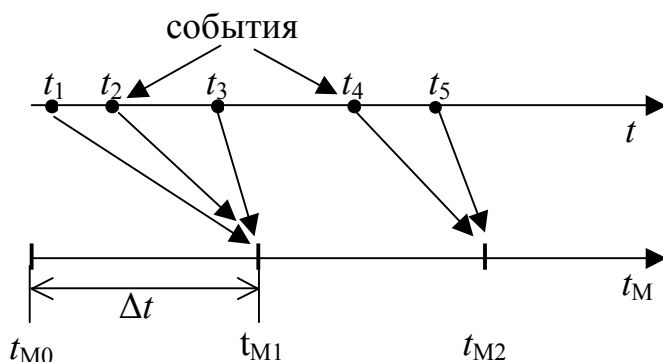


Рис. 3.18. Пример привязки событий к оси модельного времени

Метод постоянного шага предпочтительнее, если:

- события появляются регулярно, их распределение во времени достаточно равномерно;
- число событий велико и моменты их появления близки;
- невозможно заранее определить моменты появления событий.

Принцип  $\Delta t$  является наиболее универсальным принципом построения моделирующих алгоритмов, охватывающих весьма широкий класс реальных сложных систем и их элементов дискретного и непрерывного характера. Вместе с тем, принцип  $\Delta t$  оказывается весьма неэкономичным с точки зрения расхода машинного времени, если можно так выразиться, самым расточительным из

всех известных в настоящее время принципов построения моделирующих алгоритмов для сложных систем. В самом деле, при малых  $\Delta t$  будет затрачено очень много машинного времени на бесполезное определение моментов возникновения событий при их отсутствии на достаточно продолжительных интервалах времени. Если  $\Delta t$  сделать недостаточно малым, появится опасность восприятия некоторых событий как одновременных, в то время, когда в реальных процессах они не совпадают во времени., что вообще исключает возможность получения правильных результатов при моделировании. Правда, в некоторых случаях для предотвращения пропуска событий могут быть приняты специальные меры, например, разработаны приемы обнаружения пропуска и возврата к предыдущему моменту времени для повторного прохождения данного интервала с малым  $\Delta t$ .

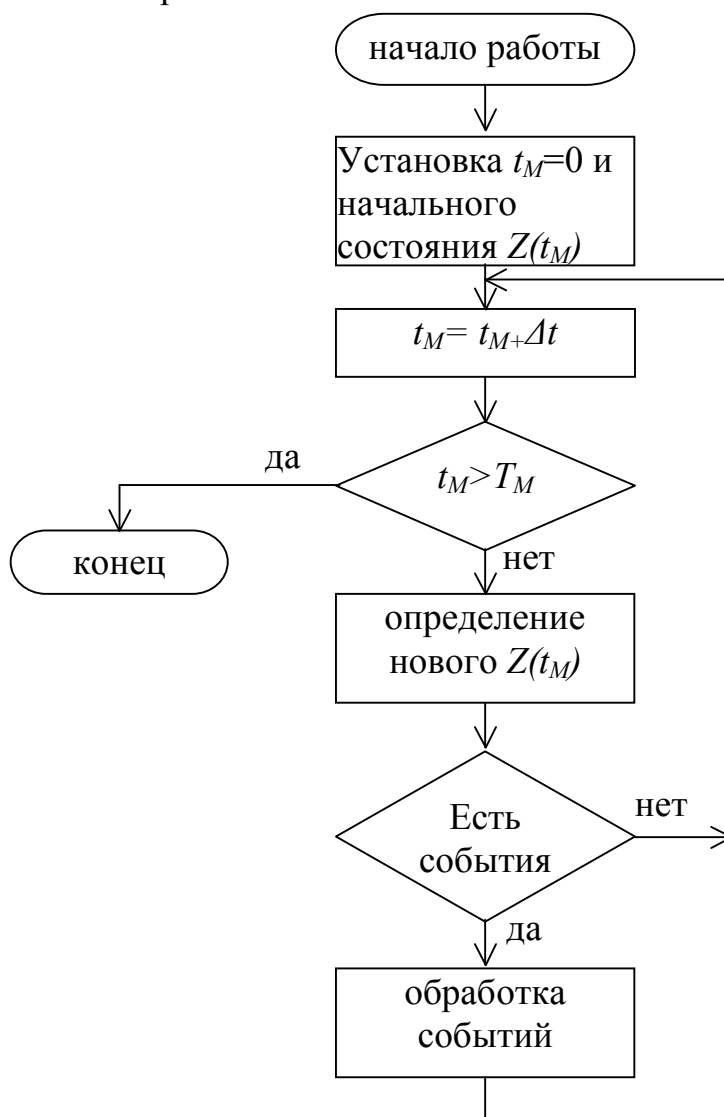


Рис. 3.19. Алгоритм моделирования с постоянным шагом

В общем виде алгоритм моделирования с постоянным шагом представлен на рис. 3.19 ( $t_M$  – текущее значение модельного времени,  $T_M$  – интервал моделирования).

Выбор величины шага моделирования является нелегким и очень важным делом. Универсальной методики решения этой проблемы не существует, но во многих случаях можно использовать один из следующих подходов:

- принимать величину шага, равной средней интенсивности возникновения событий различных типов;
- выбирать величину  $\Delta t$ , равной среднему интервалу между наиболее частыми (или наиболее важными) событиями.

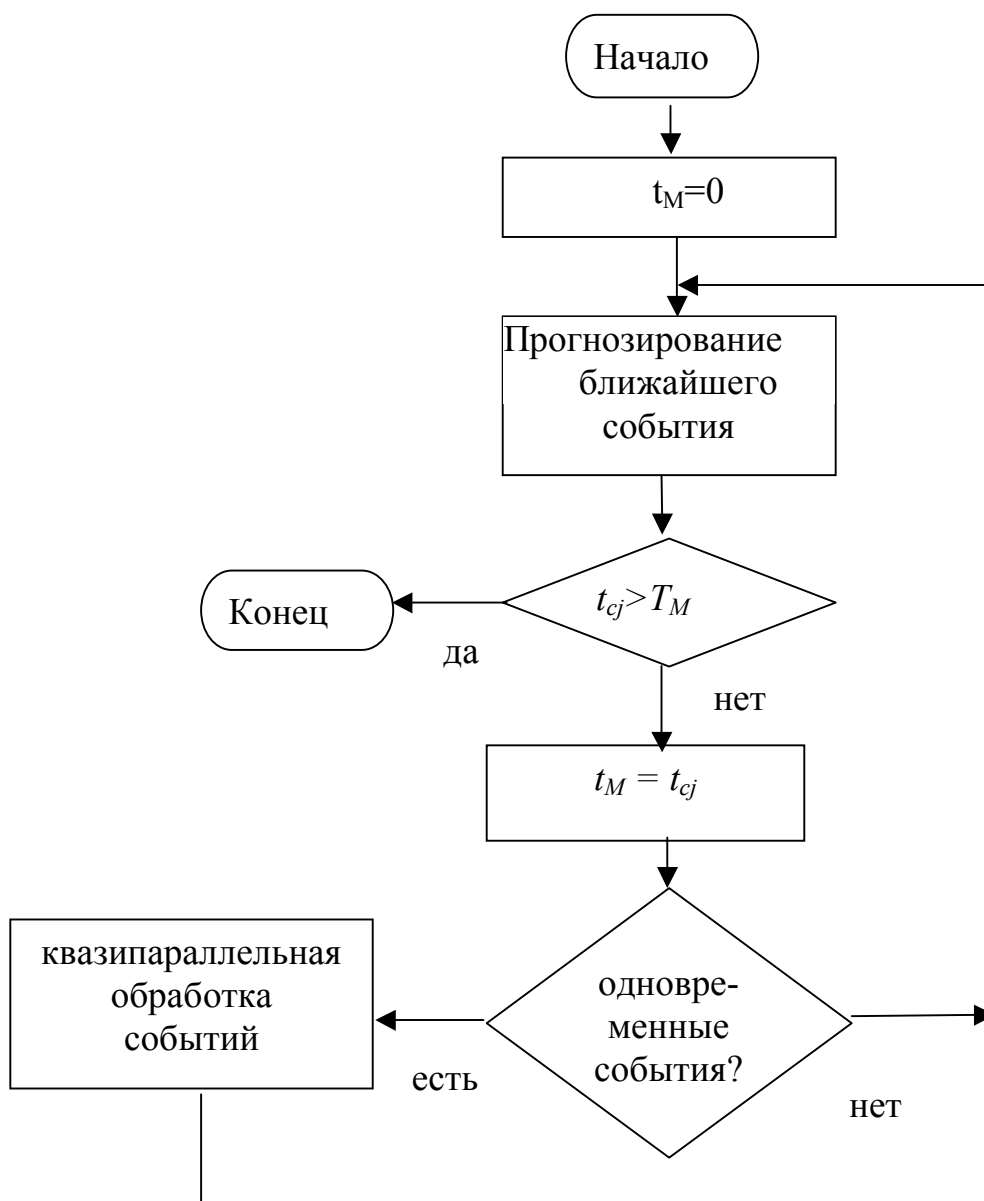


Рис. 3.20. Алгоритм моделирования по особым состояниям

При моделировании по особым состояниям системное время каждый раз изменяется на величину, строго соответствующую интервалу времени до момента наступления очередного события. В этом случае события обрабатываются в порядке их наступления, а одновременно наступившими считаются только те, которые являются одновременными в действительности.

Когда речь идёт об особых состояниях, подразумеваются моменты времени возникновения событий, изменяющих состояния моделируемого процесса. Например, если речь идёт о моделировании СМО, в которых состояние определяется числом заявок, находящихся в системе в данный момент, то это моменты прихода заявок в систему или моменты окончания их обслуживания.

Метод моделирования по особым состояниям сложнее в реализации, так как для него требуется разработка специальной процедуры планирования событий (так называемого календаря событий).

Моделирование по особым состояниям целесообразно использовать, если:

- события распределяются во времени неравномерно или интервалы между ними велики;
- предъявляются повышенные требования к точности определения взаиморасположения событий во времени;
- необходимо реализовать квазипараллельную обработку одновременных событий.

Дополнительное достоинство метода заключается в том, что он позволяет экономить машинное время, особенно при моделировании систем периодического действия, в которых события длительное время могут не наступать.

Обобщенная схема алгоритма моделирования по особым состояниям представлена на рис. 3.20 ( $t_{\text{сobj}}$  – прогнозируемый момент наступления  $j$ -го события).

При моделировании процессов обработки заявок в системах массового обслуживания иногда удобно строить моделирующие алгоритмы по принципу, идея которого состоит в последовательном воспроизведении истории отдельных заявок в порядке поступления их в систему: алгоритм обращается к сведениям о других заявках лишь в том случае, если это необходимо для решения вопроса о дальнейшем порядке обслуживания данной заявки. Такого рода моделирующие алгоритмы весьма экономны, не требуют специальных мер для учета особых состояний системы, однако они имеют весьма сложную логическую структуру и не всегда доступны для построения человеку, не обладающему достаточным опытом решения задач методом статистического моделирования. Этот принцип называется иногда «принципом последовательной проводки заявок».

Ниже приведен пример алгоритма модели, в которой реализован этот принцип.

**Одноканальная СМО с неограниченной очередью и “нетерпеливыми” заявками.** В систему поступают заявки со случайными интервалами времени  $\tau_{\Pi}$  между ними. Время обслуживания  $\tau_o$  тоже является случайной величиной. Если поступившая заявка застаёт канал занятым, то она ожидает освобождения канала, но не более, чем  $\tau_{ж}$ , после чего получает отказ. Величины  $\tau_{ж}$  случайны и для различных заявок независимы.

В результате моделирования надо получить значение доли заявок, получивших отказ (вероятность отказа) и среднее время ожидания в очереди обслуженных заявок.

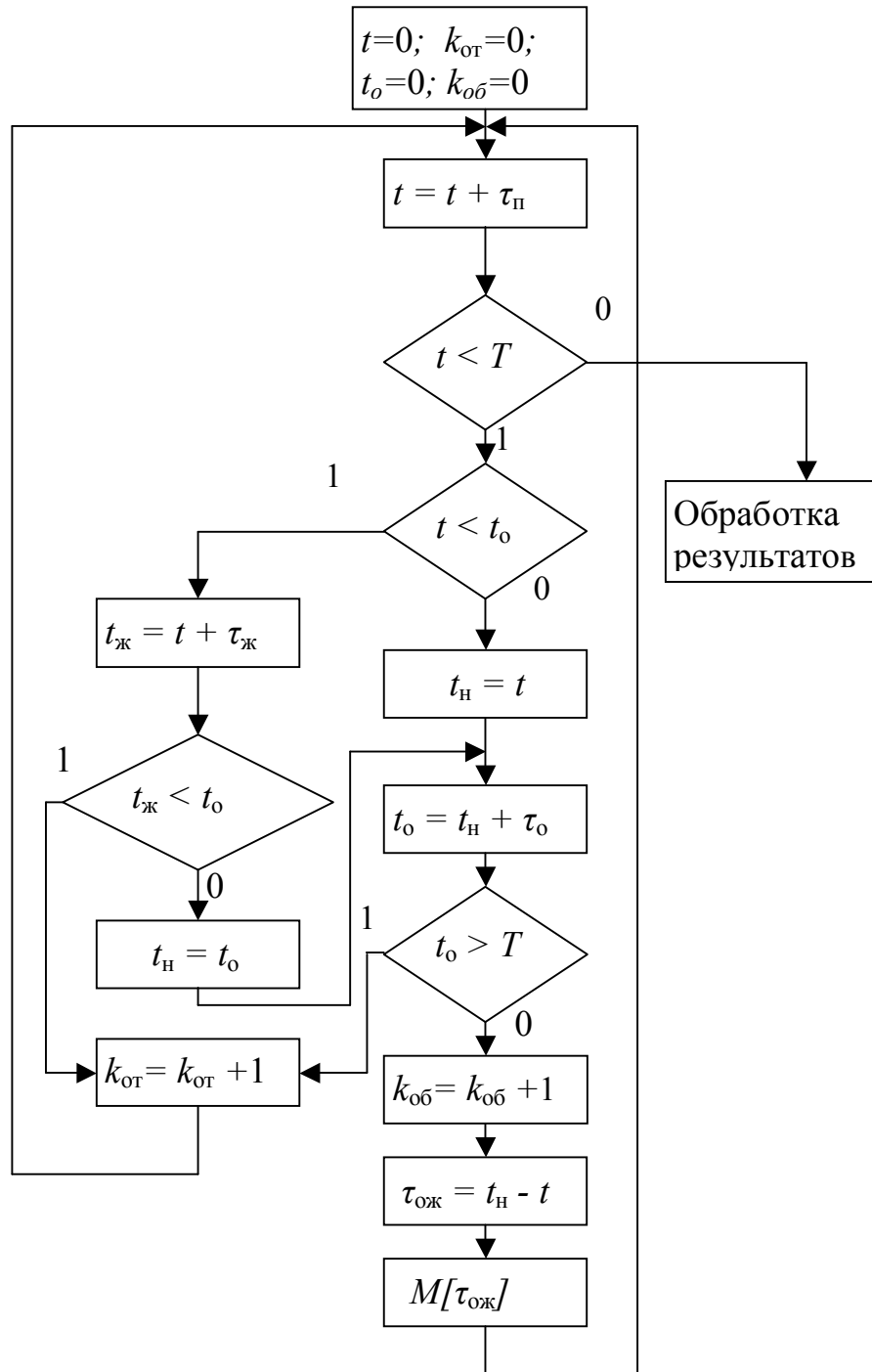


Рис.3.21. Алгоритм моделирования с последовательной проводкой заявок

Процесс функционирования СМО будем рассматривать на интервале времени  $[0, T]$ . На рис. 3.21 приведена блок-схема имитационной модели. В ней использованы следующие обозначения:  $t$  – время поступления очередной заявки;

$t_o$  – время освобождения канала (окончания обслуживания заявки);

$t_n$  – время начала обслуживания заявки каналом;

$t_{ж}$  – верхняя граница времени ожидания;

$k_{от}$  – счетчик количества заявок, получивших отказ;

$k_{об}$  – счетчик количества обслуженных заявок.

Присутствие в записи  $\tau_n$ ,  $\tau_o$  или  $\tau_{ж}$  означает обращение к соответствующему датчику случайных чисел.  $M[\tau_{ож}]$  – экспресс анализ (уточнение в очередном цикле) среднего значения времени ожидания в очереди.

На практике не всегда строго выдерживается один из упомянутых здесь принципов построения моделирующих алгоритмов. Иногда моделирующие алгоритмы строятся на нескольких принципах одновременно. Например, общая структура моделирующего алгоритма базируется на принципе особых состояний, а между особыми состояниями используется принцип последовательной проводки заявок и, кроме того, в моменты особых состояний реализуются жребии для определения случайных значений параметров и характеристик системы. Легко привести и другие примеры комбинирования различных принципов построения моделирующих алгоритмов.

### 3.10. Планирование машинных экспериментов

Машинный эксперимент с моделью системы при ее исследовании и проектировании проводится с целью получения информации о характеристиках процесса функционирования рассматриваемого объекта. Эта информация может быть получена как для анализа характеристик, так и для их оптимизации при заданных ограничениях, т. е. для синтеза структуры, алгоритмов и параметров моделируемой системы. В зависимости от поставленных целей моделирования имеются различные подходы к организации имитационного эксперимента с машинной моделью. Основная задача планирования машинных экспериментов – получение необходимой информации об исследуемой системе при ограничениях на ресурсы (затраты машинного времени, памяти и т. п.). К числу частных задач, решаемых при планировании машинных экспериментов, относятся задачи уменьшения затрат машинного времени на моделирование, увеличения точности и достоверности результатов моделирования, проверки адекватности модели и т. д.

Эффективность машинных экспериментов с имитационными моделями систем, существенно зависит от выбора плана эксперимента, так как именно план определяет объем и порядок проведения вычислений на ЭВМ, приемы накопления и статистической обработки результатов моделирования системы и



в целом влияет на эффективность использования ресурсов ЭВМ при моделировании.

Планирование модельных экспериментов преследует две основные цели:

1) сокращение общего объема испытаний при соблюдении требований к достоверности и точности их результатов;

2) повышение информативности каждого из экспериментов в отдельности.

Математические методы планирования экспериментов основаны на кибернетическом представлении процесса проведения эксперимента, наиболее подходящей моделью которого является абстрактная схема (рис. 3.22) типа "черного ящика" вида

$$Y = \varphi(X),$$

где

$X = (x_1, x_2, \dots, x_n)$  – множество векторов входных независимых переменных, называемых факторами (для машинного эксперимента они являются экзогенными или управляемыми);

$Y = (y_1, y_2, \dots, y_m)$  – множество векторов зависимых выходных переменных, называемых реакциями (для машинного эксперимента эти переменные являются эндогенными).

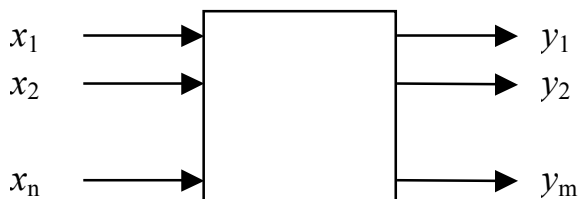


Рис. 3.22. Представление объекта моделирования в виде «черного ящика»

Функция  $\varphi$ , связывающая реакцию с факторами, называется функцией реакции.

Исследователю заранее не известен вид зависимостей  $\varphi(X)$ , поэтому используют приближенные соотношения:

$$Y = f(X).$$

Зависимости  $f(X)$  находятся по данным эксперимента. Последний необходимо поставить так, чтобы при минимальных затратах ресурсов (например, минимальном числе испытаний), варьируя по специально сформулированным правилам значения входных переменных, построить математическую модель системы и оценить ее характеристики.

При планировании экспериментов необходимо определить основные свойства факторов. Факторы при проведении экспериментов могут быть управляемыми и неуправляемыми, наблюдаемыми и ненаблюдаемыми, изучаемыми и неизучаемыми, количественными и качественными, фиксированными и случайными.

Фактор называется управляемым, если его уровни целенаправленно выбираются исследователем в процессе эксперимента. При машинной реализации модели исследователь принимает решения, управляя изменением в допустимых пределах различных факторов.

Фактор называется наблюдаемым, если его значения наблюдаются и регистрируются. Обычно в машинном эксперименте с моделью наблюдаемые факторы совпадают с управляемыми, так как нерационально управлять фактором, не наблюдая его. Но неуправляемый фактор также можно наблюдать. Например, на этапе проектирования конкретной системы нельзя управлять заданными воздействиями внешней среды, но можно наблюдать их в машинном эксперименте. Наблюдаемые неуправляемые факторы получили название сопутствующих. Обычно при машинном эксперименте с моделью число сопутствующих факторов велико, поэтому рационально учитывать влияние лишь тех из них, которые наиболее существенно воздействуют на интересующую исследователя реакцию.

Фактор относится к изучаемым, если он включен в модель для изучения свойств системы, а не для вспомогательных целей, например для увеличения точности эксперимента.

Фактор будет количественным, если его значения – числовые величины, влияющие на реакцию, а в противном случае фактор называется качественным. Например, в модели системы, формализуемой в виде схемы массового обслуживания (Q-схемы), количественными факторами являются интенсивности входящих потоков заявок, интенсивности потоков обслуживания, емкости накопителей, количество обслуживающих каналов и т. д., а качественными факторами – дисциплины постановки в очередь, выбора из очереди, обслуживания заявок каналами и т. д. Качественным факторам в отличие от количественных не соответствует числовая шкала. Однако и для них можно построить условную порядковую шкалу, с помощью которой производится кодирование, устанавливая соответствие между условиями качественного "фактора и числами натурального ряда.

Фактор называется фиксированным, если в эксперименте исследуются все интересующие экспериментатора значения фактора, а если экспериментатор исследует только некоторую случайную выборку из совокупности интересующих значений факторов, то фактор называется случайным. На основании случайных факторов могут быть сделаны вероятностные выводы и о тех значениях факторов, которые в эксперименте не исследовались.

Основными требованиями, предъявляемыми к факторам, являются требование управляемости фактора и требование непосредственного воздействия на объект. Под управляемостью фактора понимается возможность установки и поддержания выбранного нужного уровня фактора постоянным в течение всего испытания или изменяющимся в

соответствии с заданной программой. Требование непосредственного воздействия на объект имеет большое значение в связи с тем, что трудно управлять фактором, если он является функцией других факторов.

Поскольку факторы могут носить как количественный, так и качественный характер (например, отражать некоторую стратегию управления), значения факторов обычно называют уровнями. Если при проведении эксперимента исследователь может изменять уровни факторов, эксперимент называется активным, в противном случае – пассивным.

При планировании эксперимента обычно одновременно изменяются несколько факторов. Определим требования, которые предъявляются к совокупности факторов. Основные из них – совместимость и независимость. Совместимость факторов означает, что все их комбинации осуществимы, а независимость соответствует возможности установления фактора на любом уровне независимо от уровней других.

Фиксированный набор уровней факторов определяет одно из возможных состояний рассматриваемой системы. Одновременно этот набор представляет собой условия проведения одного из возможных экспериментов.

Поиск плана эксперимента производится в так называемом факторном пространстве.

Факторное пространство – это множество внешних и внутренних параметров модели, значения которых исследователь может контролировать в ходе подготовки и проведения модельного эксперимента.

При проведении эксперимента должно быть установлено значение функции реакции

$$Y = f(X).$$

Этот эксперимент необходимо поставить так, чтобы при минимальных затратах ресурсов (например, минимальном числе испытаний), варьируя по специально сформулированным правилам значения входных переменных, построить математическую модель системы и оценить ее характеристики.

Под планом эксперимента мы будем понимать множество точек факторного пространства, в которых будет проводиться эксперимент. Иначе говоря, план представляет собой таблицу, каждая строка которой содержит набор значений (уровней) факторов, при которых будут проводиться прогоны программной модели (испытания).

В зависимости от решаемых задач различают *стратегическое* и *тактическое* планирование эксперимента.

При стратегическом планировании эксперимента должны быть решены три основные задачи:

- 1) идентификация факторов;
- 2) выбор вида функции реакции;

### 3) выбор уровней факторов.

Совокупность методов определения необходимого числа опытов при проведении каждого отдельного испытания для обеспечения требуемой точности и достоверности результатов относятся к тактическому планированию эксперимента. Эти вопросы будут рассмотрены в разделе, связанном с обработкой экспериментальных данных, а сейчас мы более подробно ознакомимся со стратегическим планированием имитационного эксперимента

Цель методов стратегического планирования имитационных экспериментов – получение максимального объема информации об исследуемой системе в каждом эксперименте (наблюдении). Другими словами, стратегическое планирование позволяет ответить на вопрос, при каком сочетании уровней внешних и внутренних факторов может быть получена наиболее полная и достоверная информация о поведении моделируемой системы.

Под идентификацией факторов понимается их ранжирование по степени влияния на значение наблюдаемой переменной (показателя эффективности).

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

Наиболее распространенными и полно отвечающими задачам статистического моделирования являются полиномиальные модели. Задача нахождения полиномиальной модели, описывающей систему или отдельные ее характеристики, состоит в оценке вида и параметров некоторой функции  $f(x_1, x_2, \dots, x_m)$ .

Если в модели учитывается влияние  $m$  количественных факторов  $x_i$ , ( $i = 1, 2, \dots, m$ ) на некоторую реакцию, то полином степени  $d$  от  $m$  переменных, в общем случае будет иметь вид

$$y = b_0 + \sum_{i=1}^m b_i x_i + \sum_{1 < i < j \leq m} b_{ij} x_i x_j + \dots + \sum_{i_1, i_2, \dots, i_m} b_{i_1, i_2, \dots, i_m} x_1^{i_1} x_2^{i_2} \dots x_m^{i_m},$$
$$\sum i_j = d.$$

Этот полином содержит  $C_{m+d}^d$  коэффициентов. Значения этих коэффициентов должны быть установлены по результатам проведения эксперимента с моделью. Поэтому план эксперимента  $D$  должен содержать, по крайней мере,  $N \geq C_{m+d}^d$  различных точек факторного пространства:

$$D = \begin{vmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1N} & x_{2N} & \dots & x_{mN} \end{vmatrix}$$

Реализовав испытания в  $N$  точках области факторного пространства, отведенной для экспериментирования, получим вектор наблюдений, имеющий следующий вид:

$$\vec{y} = \begin{vmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{vmatrix},$$

где  $y_k$  – реакция, соответствующая  $k$ -й точке плана  $\vec{x}_k = \|x_{1k}, x_{2k}, \dots, x_{mk}\|$ .

На практике при аппроксимации полиномами функции реакции чаще всего используют два частных случая:

1) линейная форма

$$f(x_1, x_2, \dots, x_m) = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m;$$

2) неполная квадратичная форма

$$f(x_1, x_2, \dots, x_m) = b_0 + \sum_{i=1}^m b_i x_i + \sum_{i < j} b_{ij} x_i x_j.$$

Линейная форма подразумевает отсутствие взаимосвязи между влиянием отдельных факторов на исследуемые характеристики моделируемой системы. Неполная квадратичная форма учитывает взаимодействие факторов.

Пусть имеет место эксперимент, учитывающий три фактора и функция реакции предполагается линейной:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3.$$

Функция содержит четыре коэффициента  $b_i$ , значения которых надо установить в результате проведения машинного эксперимента с моделью, поэтому план эксперимента должен содержать четыре строки

$$D = \begin{vmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \\ x_{14} & x_{24} & x_{34} \end{vmatrix}.$$

Проведя четыре эксперимента с моделью (четыре прогона программы), в каждом из которых будет использоваться для задания

значений факторов поочередно одна из строк плана, получим четыре значения выходной характеристики  $y$ .

Таким образом, мы приходим к системе из четырех уравнений, в которой неизвестными являются коэффициенты  $b_i$

$$y_1 = b_0 + b_1x_{11} + b_2x_{21} + b_3x_{31},$$

$$y_2 = b_0 + b_1x_{12} + b_2x_{22} + b_3x_{32},$$

$$y_3 = b_0 + b_1x_{13} + b_2x_{23} + b_3x_{33},$$

$$y_4 = b_0 + b_1x_{14} + b_2x_{24} + b_3x_{34}.$$

Решив эту систему, найдем значения коэффициентов, определив тем самым конкретный вид функции

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3.$$

Выбор уровней факторов производится с учетом двух противоречивых требований:

1) уровни фактора должны перекрывать (заполнять) весь возможный диапазон его изменения;

2) общее количество уровней по всем факторам не должно приводить к чрезмерному объему моделирования.

Отыскание компромиссного решения, удовлетворяющего этим требованиям, и является одной из задач стратегического планирования эксперимента.

Вначале следует выбрать границы  $x_{imin}$  и  $x_{imax}$  области определения факторов  $G$ , задаваемые исходя из свойств исследуемого объекта, т. е. на основе анализа априорной информации о моделируемой системе и внешней среде. Например, такая переменная, как температура, при термобарических экспериментах принципиально не может быть ниже абсолютного нуля и выше температуры плавления материала, из которого изготовлена термобарокамера.

После определения области  $G$  необходимо найти локальную подобласть для планирования эксперимента путем выбора основного (нулевого) уровня  $x_{i0}$  и интервалов варьирования  $\Delta x_i$ ,  $i=1, 2, \dots, m$ . В качестве исходной точки  $x_{i0}$  выбирают такую, которая соответствует наилучшим условиям, определенным на основе анализа априорной информации о моделируемой системе. Причем эта точка не должна лежать близко к границам области определения факторов  $x_{imin}$  и  $x_{imax}$ . На выбор интервала варьирования  $\Delta x_i$  накладываются естественные ограничения снизу (интервал не может быть меньше ошибки фиксирования уровня фактора, так как в противном случае верхний и нижний уровни окажутся неразличимыми) и сверху (верхний и нижний уровни не должны выходить за область определения  $G$ ).

В рамках выбранной модели планирования в виде алгебраических полиномов строится план эксперимента путем варьирования каждого из

факторов  $x_i$ ,  $i=1, 2, \dots, m$  на нескольких уровнях  $q$  относительно исходной точки  $x_{i0}$ , представляющей собой центр эксперимента.

Эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется *полным факторным экспериментом* (ПФЭ).

Общее число различных комбинаций уровней в ПФЭ для  $k$  факторов можно вычислить так:  $l_1 l_2 \dots l_k$ ,  
где  $l_i$  – число уровней  $i$ -го фактора.

Недостаток ПФЭ – большие временные затраты на подготовку и проведение эксперимента.

Например, если в модели отражены 4 фактора, влияющие на значение выбранного показателя эффективности, каждый из которых имеет 3 возможных уровня (значения), то план проведения ПФЭ будет включать 81 эксперимент. Если при этом каждый из них длится хотя бы одну минуту (с учетом времени на изменение значений факторов), то на однократную реализацию ПФЭ потребуется более часа.

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

Если такие взаимодействия считают отсутствующими или их эффектом пренебрегают, проводят частичный факторный эксперимент (ЧФЭ).

Построение плана эксперимента для ПФЭ не вызывает затруднений, так как при этом в план просто включаются все возможные комбинации уровней факторов. При построении плана ЧФЭ из всего множества возможных сочетаний надо отобрать только те, которые будут использованы в эксперименте. Известны и применяются на практике различные варианты построения планов ЧФЭ. Мы рассмотрим только некоторые из них.

**Рандомизированный план** – предполагает выбор сочетания уровней для каждого прогона случайным образом.

**Латинский план** («латинский квадрат») – используется в том случае, когда проводится эксперимент с одним первичным фактором и несколькими вторичными. Суть такого планирования состоит в следующем. Если первичный фактор  $A$  имеет  $l$  уровней, то для каждого вторичного фактора также выбирается  $l$  уровней. Выбор комбинации уровней факторов выполняется на основе специальной процедуры, которую мы рассмотрим на примере.

Пусть в эксперименте используется первичный фактор  $A$  и два вторичных фактора –  $B$  и  $C$ ; число уровней факторов  $l$  равно 4.

Соответствующий план можно представить в виде квадратной матрицы размером  $4 \times 4$  относительно уровней фактора  $A$ . При этом матрица строится таким образом, чтобы в каждой строке и в каждом столбце данный уровень фактора  $A$  встречался только один раз (табл.):

Таблица 3.1. Пример латинского плана

Значение фактора $B$	Значение фактора $C$			
	$C_1$	$C_2$	$C_3$	$C_4$
$B_1$	$A_1$	$A_2$	$A_3$	$A_4$
$B_2$	$A_2$	$A_3$	$A_4$	$A_1$
$B_3$	$A_3$	$A_4$	$A_1$	$A_2$
$B_4$	$A_4$	$A_1$	$A_2$	$A_3$

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

**Эксперимент с изменением факторов по одному** (однофакторный план).

Суть его состоит в том, что один из факторов «пробегает» все  $l$  уровней, а остальные  $m-1$  факторов поддерживаются постоянными. Такой план обеспечивает исследование эффектов каждого фактора в отдельности. Он требует всего  $N = l_1 + l_2 + \dots + l_m$  прогонов ( $l_i$  – число уровней  $i$ -го фактора).

Для рассмотренного выше примера (3 фактора, имеющие по 4 уровня)  $N = 4 + 4 + 4 = 12$ .

Еще раз подчеркнем, что такой план применим только при отсутствии взаимодействия между факторами.

**Дробный факторный эксперимент.**

При таком способе построения плана каждый фактор имеет два уровня – нижний и верхний, поэтому общее число вариантов эксперимента  $N = 2^m$ ,  $k$  – число факторов.

### 3.11. Обработка экспериментальных данных



### 3.11.1. Экспериментальные оценки

Чтобы найти закон распределения случайной величины, нужно располагать достаточно обширным статистическим материалом, порядка нескольких сотен опытов (наблюдений). Однако на практике часто приходится иметь дело со статистическим материалом весьма ограниченного объема – с двумя-тремя десятками наблюдений, часто даже меньше. Это обычно связано с дороговизной и сложностью постановки каждого опыта. Такого ограниченного материала явно недостаточно для того, чтобы найти заранее неизвестный закон распределения случайной величины, но все же этот материал может быть обработан и использован для получения некоторых сведений о случайной величине. Например, на основе ограниченного статистического материала можно определить—хотя бы ориентировочно – важнейшие числовые характеристики случайной величины: математическое ожидание или дисперсию. На практике часто бывает, что вид закона распределения известен заранее, а требуется найти только некоторые параметры, от которых он зависит. Например, если заранее известно, что закон распределения случайной величины нормальный, то задача обработки сводится к определению двух его параметров  $m$  и  $\sigma$ . Если заранее известно, что величина распределена по закону Пуассона, то подлежит определению только один его параметр: интенсивность  $\lambda$ . Наконец, в некоторых задачах вид закона распределения вообще несуществен, а требуется знать только его числовые характеристики.

Например, оценкой для математического ожидания может служить среднее арифметическое наблюдаемых значений случайной величины в  $n$  независимых опытах. При очень большом числе опытов среднее арифметическое будет с большой вероятностью весьма близко к математическому ожиданию. Если же число опытов невелико, то замена математического ожидания средним арифметическим приводит к какой-то ошибке. Эта ошибка в среднем тем больше, чем меньше число опытов. Так же будет обстоять дело и с оценками других неизвестных параметров. Любая из таких оценок случайна; при пользовании ею неизбежны ошибки. Желательно выбрать такую оценку; чтобы эти ошибки были по возможности минимальными.

Любое значение искомого параметра, вычисляемого на основе ограниченного числа опытов, всегда будет содержать элемент случайности. Такое приближённое, случайное значение называют *оценкой* параметра. При большом числе опытов оценка будет близка к значению параметра, при малом может иметь место достаточно большая ошибка. Желательно выбрать оценку так, чтобы эта ошибка была минимальна.

Пусть имеется случайная величина  $X$ , закон распределения которой содержит параметр  $a$ . Надо найти оценку для  $a$  по результатам  $n$  независимых опытов.

Пусть в результате проведенных опытов были получены значения случайной величины:

$$x_1, x_2, \dots, x_n.$$

Обозначим оценку для  $a$  как  $\tilde{a}$ . Она должна быть функцией от наблюдаемых величин

$$\tilde{a} = \tilde{a}(x_1, x_2, \dots, x_n).$$

Т.к.  $\tilde{a}$  есть функция от случайных величин, то она тоже случайная величина. Закон распределения  $\tilde{a}$  зависит, во-первых, от закона распределения величины  $X$  (и, в частности, от самого неизвестного параметра  $a$ ), во-вторых, от числа опытов  $n$ .

Предъявим к  $\tilde{a}$  ряд требований, которым она должна удовлетворять, чтобы быть в каком-то смысле «доброкачественной» оценкой:

- оценка *состоятельная*, если при увеличении  $n$  она приближается (сходится по вероятности) к  $a$ ;
- оценка *несмещённая*, если математическое ожидание оценки  $M[\tilde{a}] = a$ ;
- оценка, обладающая по сравнению с другими наименьшей дисперсией, называется *эффективной*.

На практике не всегда удаётся удовлетворить всем этим требованиям. Иногда в интересах простоты расчётов применяют незначительно смещённые оценки.

### 3.11.2. Оценки для математического ожидания и дисперсии

Пусть имеется случайная величина  $X$  с математическим ожиданием  $m$  и дисперсией  $D$ .

Произведено  $n$  опытов, давших результаты

$$x_1, x_2, \dots, x_n.$$

В качестве оценки для математического ожидания естественно предложить среднее арифметическое наблюдаемых значений  $x_i$ .

$$\tilde{m} = m^* = \frac{1}{n} \sum_{i=1}^n x_i.$$

Эта оценка является состоятельной (т.к. по теореме Чебышева частоты сходятся к математическому ожиданию при  $n \rightarrow \infty$ ) и несмещённой т.к.

$$M[\tilde{m}] = \frac{1}{n} \sum_{i=1}^n m = m.$$

Дисперсия этой оценки  $D[\tilde{m}] = \frac{1}{n} D$ ;  $\sigma_{\tilde{m}} = \sqrt{\frac{D}{n}}$ .

Рассмотрим оценку дисперсии. Можно использовать в качестве оценки статистическую дисперсию:

$$D^* = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m})^2 \quad \text{Можно показать, что } D^* \text{ сходится по}$$

вероятности к  $D$ , т.е. оценка  $D^*$  состоятельна.

Если подставить выражение для  $D^*$  в формулу для  $\tilde{m}$  и преобразовать сумму, то получим

$$M[D^*] = \frac{n-1}{n} D,$$

т.е. эта оценка не является несмещённой. Смещение можно ликвидировать, если умножить  $D^*$  на  $\frac{n}{n-1}$ . Получим

$$\tilde{D} = \frac{n}{n-1} D^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - \tilde{m})^2.$$

При  $n \rightarrow \infty$   $\tilde{D} \rightarrow D^*$  и, так как  $D^*$  состоятельна, то и  $\tilde{D}$  состоятельна ( $\frac{n}{n-1}$  — поправка Бесселя, при  $n > 50$  между  $D^*$  и  $\tilde{D}$  практически нет разницы).

На практике вместо этой формулы можно использовать равносильную ей, выразив дисперсию через 2-й центральный момент

$$\tilde{D} = \frac{n}{n-1} \left[ \frac{1}{n} \sum_{i=1}^n x_i^2 - \tilde{m}^2 \right]$$

При работе с имитационной моделью для получения оценки математического ожидания и дисперсии можно использовать следующие соотношения:

$$\tilde{m}_n = \tilde{m}_{n-1} \frac{n-1}{n} + \frac{x_n}{n} \quad \text{и}$$

$$\tilde{D}_n = \tilde{D}_{n-1} \frac{n-2}{n-1} + \frac{1}{n} (x_n - \tilde{m}_{n-1})^2.$$

С их помощью можно осуществлять экспресс-анализ значений математического ожидания и дисперсии в ходе моделирования без сохранения  $x_i$  в массиве.

### 3.11.2. Доверительные интервал и вероятность

Мы рассмотрели вопрос об оценке неизвестного параметра  $a$  одним числом  $\tilde{a}$  (это так называемая точечная оценка). В ряде задач требуется не

только найти для параметра  $a$  подходящее численное значение, но и оценить его точность и надежность

Надо знать, к каким ошибкам может привести замена  $a$  на  $\tilde{a}$  и с какой степенью уверенности можно ожидать, что эти ошибки не выйдут за заданные пределы. Это особенно актуально при малом числе наблюдений  $n$ . Чтобы дать представление о точности и надёжности  $\tilde{a}$  пользуются доверительными интервалами и доверительными вероятностями.

Пусть для  $a$  из опыта получена оценка  $\tilde{a}$ . Зададим достаточно большую вероятность  $\beta$  такую, что событие с такой вероятностью можно считать практически достоверным. Найдём такое значение  $\varepsilon$ , для которого:

$$P(|\tilde{a} - a| < \varepsilon) = \beta.$$

Перепишем выражение так:

$$P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \beta.$$

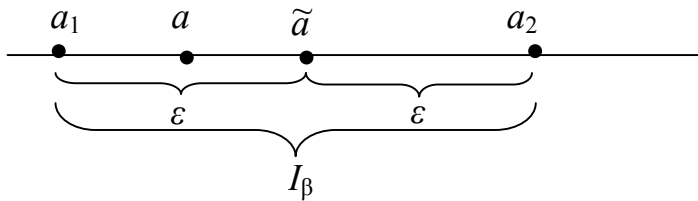


Рис. 3.23. Доверительный интервал

Вероятность  $\beta$  принято называть *доверительной вероятностью*, а интервал  $I_\beta$  — *доверительным интервалом*. Границы интервала  $a_1 = \tilde{a} - \varepsilon$  и  $a_2 = \tilde{a} + \varepsilon$  называются *доверительными границами* (рис. 3.23).

Отметим, что величина  $a$  не случайна, а положение интервала  $I_\beta$  на оси абсцисс, определяемое его центром  $\tilde{a}$  случайно. Поэтому величину  $\beta$  можно трактовать как вероятность того, что случайный интервал  $I_\beta$  накроет точку  $a$ .

Рассмотрим вопрос об определении доверительных границ  $a_1$  и  $a_2$ . Пусть для параметра  $a$  имеется несмещенная оценка  $\tilde{a}$ . Если бы нам был известен закон распределения величины  $\tilde{a}$ , задача нахождения доверительного интервала была бы весьма проста: достаточно было бы найти такое значение  $\varepsilon$ , для которого

$$P(|\tilde{a} - a| < \varepsilon) = \beta.$$

Затруднение состоит в том, что закон распределения оценки  $\tilde{a}$  зависит от закона распределения величины  $X$  и, следовательно, от его неизвестных параметров (в частности, и от самого параметра  $a$ ).

Чтобы обойти это затруднение, можно применить следующий грубо приближенный прием: заменить в выражении для  $\varepsilon$  неизвестные параметры их точечными оценками. При сравнительно большом числе опытов  $n$  (порядка  $20 \div 30$ ) этот прием обычно дает удовлетворительные по точности результаты.

В качестве примера рассмотрим задачу о доверительном интервале для математического ожидания.

Если произведено  $n$  независимых опытов над случайной величиной  $X$ , характеристики которой — математическое ожидание  $m$  и дисперсия  $D$  — неизвестны и для этих параметров получены оценки  $\tilde{m}$  и  $\tilde{D}$ , то воспользуется тем, что оценка  $\tilde{m}$  представляет собой сумму одинаково распределённых случайных величин  $x_i$  и, согласно центральной предельной теореме закон распределения  $\tilde{m}$  близок к нормальному с параметрами

$$M[\tilde{m}] = m, \quad D[\tilde{m}] = \frac{1}{n}D.$$

Между  $\varepsilon$  и доверительной вероятностью  $\beta$  существует связь и её можно выразить аналитически, используя функцию распределения.

Для нормального закона функция плотности вероятности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Функция распределения

$$F(x) = \int_{-\infty}^x f(x)dx.$$

Обозначим  $t = \frac{x-m}{\sigma}$  и сделаем замену переменной выражение для  $F(x)$ :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dt.$$

Этот интеграл не выражается через элементарные функции, но его можно вычислить через специальную функцию  $\Phi^*(x)$ , для которой составлены таблицы:

$$\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dx.$$

Нетрудно видеть, что эта функция представляет собой функцию распределения для нормально распределенной случайной величины с параметрами  $m=0$  и  $\sigma=1$ . Очевидно, что

$$F(x) = \Phi^*\left(\frac{x-m}{\sigma}\right).$$

Одним из свойств этой функции является то, что

$$\Phi^*(-x) = 1 - \Phi^*(x).$$

Теперь определим вероятность попадания случайной величины  $\tilde{m}$  в доверительный интервал, то есть решим вопрос о том, с какой вероятностью полученная оценка  $\tilde{m}$  отличается от действительного значения  $m$  не более, чем на  $\varepsilon$  (рис. ).

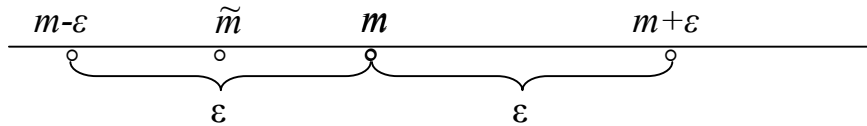


Рис. Попадание оценки в доверительный интервал.

Так как оценка  $\tilde{m}$  имеет нормальное распределение, то

$$P(|\tilde{m} - m| < \varepsilon) = F(m + \varepsilon) - F(m - \varepsilon) = \Phi^*\left(\frac{\varepsilon}{\sigma}\right) - \Phi^*\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi^*\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

Но это доверительная вероятность:

$$\beta = 2\Phi^*\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

Отсюда  $\varepsilon = \sigma \arg \Phi^*\left(\frac{1+\beta}{2}\right) = \sigma t_\beta,$

где  $t_\beta = \arg \Phi^*\left(\frac{1+\beta}{2}\right)$ ,  $\arg \Phi^*(x)$  – функция, обратная  $\Phi^*(x)$ , т.е. такое

значение аргумента, при котором нормальная функция  $\Phi^*(x)$  равна  $x$ . Для нормального закона распределения разработаны специальные таблицы, используя которые, можно для каждого  $\beta$  выбрать значение  $t_\beta$ .

Окончательно:

$$\varepsilon = t_\beta \sigma = t_\beta \sqrt{D}.$$

### 3.11.3. Точность. Определение числа реализаций

Выбор количества реализаций является задачей тактического планирования эксперимента и зависит от того, какие требования по точности предъявляются к результатам моделирования.

Пусть целью моделирования будет вычисление вероятности появления события  $A$ . В каждой из  $N$  реализаций событие  $A$  может наступить или не наступить. Иначе говоря, количество  $\xi$  наступления событий в данной

реализации является случайной величиной, принимающей значения  $x_1 = 1$  с вероятностью  $p$  и значение  $x_2 = 0$  с вероятностью  $1-p$ .

Математическое ожидание  $\xi$ :

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

дисперсия для  $\xi$ :

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

В качестве оценки для вероятности  $p$  принимается частота  $m/N$  наступления события  $A$  в  $N$  реализациях.

Частоту можно представить в виде:

$$\frac{m}{N} = \frac{1}{N} \sum_{i=1}^N \xi_i,$$

где  $\xi_i$  - количество наступления события  $A$  в  $i$ -й реализации.

В силу центральной предельной теоремы (так как  $m/N$  представляет собой сумму случайных величин) частота  $m/N$  при достаточно больших значениях  $N$  будет иметь распределение близкое к нормальному.

Найдем характеристики для частоты.

Можно представить частоту в следующем виде:

$$\frac{m}{N} = \sum_{i=1}^N \frac{\xi_i}{N}.$$

Обозначим:  $\frac{\xi_i}{N} = a_i$ . Тогда:

$$M[a_i] = \frac{1}{N} p + \frac{0}{N} (1 - p) = \frac{p}{N};$$

$$M\left[\frac{m}{N}\right] = M[a_i] \cdot N = p;$$

$$D[a_i] = \frac{1}{N^2} (1 - p)^2 p + \frac{1}{N^2} (0 - p)^2 (1 - p) = \frac{p(1 - p)}{N^2};$$

$$D\left[\frac{m}{N}\right] = ND[a_i] = \frac{p(1 - p)}{N}.$$

Для каждого значения достоверности  $\beta$  из таблиц нормального распределения можно выбрать такую величину  $t_\beta$ , что точность  $\varepsilon$  будет равна:

$$\varepsilon = t_\beta \sqrt{D\left[\frac{m}{N}\right]}.$$

Подставляя значение дисперсии, получим:

$$\varepsilon = t_\beta \sqrt{\frac{p(1 - p)}{N}}$$

Отсюда можно получить количество реализаций  $N$ , необходимое для получения оценки  $m/N$  с точностью  $\varepsilon$  и достоверностью  $\beta$ :

$$N = t_{\beta}^2 \frac{p(1-p)}{\varepsilon^2}.$$

Но  $p$  заранее не известно. Поэтому поступают так. Берут  $N_0 = 50 - 100$ , определяют по результатам  $N_0$  реализаций  $m/N_0$ , а затем, принимая  $p \approx \frac{m}{N_0}$ , и окончательно назначают значение  $N$  и проводят остальные  $N - N_0$  испытаний.

Для случая, когда надо получить оценку математического ожидания:

$$\varepsilon_{\beta} = \sigma_{\tilde{m}} t_{\beta} = t_{\beta} \sqrt{\frac{\tilde{D}}{N}} = \frac{t_{\beta} \cdot \tilde{\sigma}}{\sqrt{N}}.$$

Отсюда:

$$N \geq t_{\beta}^2 \frac{\tilde{D}}{\varepsilon_{\beta}^2}.$$

Для оценки дисперсии:

$$\varepsilon_{\beta} = t_{\beta} \cdot \sigma_{\tilde{D}} = t_{\beta} \sqrt{\frac{2}{N-1} \tilde{D}}$$

Отсюда:

$$N \geq 2 t_{\beta}^2 \frac{\tilde{D}^2}{\varepsilon^2}.$$



# ЛИТЕРАТУРА

## Основная литература

- 1.Советов Б.Я., Яковлев С.А. Моделирование систем: Учебник для вузов. – М.: Высш. школа, 2001.
- 2.Моделирование вычислительных систем .Н.Альянах. Л.: Машиностроение. Ленинградское отд.,1988.
- 3.Артамонов Г.Т., Брехов О.М. Аналитические вероятностные модели функционирования ЭВМ. – М.: Энергия, 1978.
- 4.Гультяев А.К. Имитационное моделирование в среде Windows: практическое пособие. –СПб.: КОРОНА принт, 1999.
- 5.Клейнрок Л. Теория массового обслуживания. –М.: Машиностроение, 1979.
- 6.Харин Ю.С. и др. Основы имитационного и статистического моделирования. Учебное пособие – Мн.: Дизайн ПРО, 1997.
7. Бусленко Н.П. Моделирование сложных систем. – М.: Наука, 1978.
8. Асатурян В.И. Теория планирования эксперимента. – М.: Радио и связь 1984.
- 9.Тихоненко О.М. Модели массового обслуживания в информационных системах. – Мн.: УП «Технопринт», 2003.

## Дополнительная литература

- 1.Питерсон Дж. Теория сетей Петри и моделирование систем. –М.:Мир, 1984.
- 2.Р.Шеннон. Имитационное моделирование систем - искусство и наука: Пер. с англ. – М.:Мир, 1978.
- 3.Полляк Ю.Г. Вероятностное моделирование на электронных вычислительных машинах. –М.: Советское радио, 1971.
- 4.Советов Б.Я., Яковлев С.А. Моделирование систем: Лабораторный практикум :Учебное пособие для вузов по спец. АСОИиУ. – М.: Высш. школа, 1989.