МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ РАСТИТЕЛЬНЫХ ПОЛИМЕРОВ»

Кафедра прикладной математики и информатики

Н.Л. ЛЕОНОВА

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Конспект лекций

Санкт-Петербург
2015

УДК 004.942

Имитационное моделирование: конспект лекций / Н.Л. Леонова; СПбГТУРП. – СПб., 2015. − 94 с.

В пособии излагаются материалы по курсу «Имитационное моделирование».

Учебное пособие предназначено для использования студентами, обучающимися на кафедре прикладной математики и информатики по направлениям подготовки бакалавров: 010400 (01.03.02) «Прикладная математика и информатика».

Рецензент:
профессор кафедры государственного и муниципального управления Санкт-Петербургского государственного института кино и телевидения, д-р техн. наук В.М. Пестриков.

Подготовлено и рекомендовано к печати кафедрой прикладной математики и информатики СПбГТУРП (протокол № 2 от 14.10.2014 г.).

Утверждено к изданию методической комиссией факультета АСУТП СПбГТУРП (протокол № 6 от 11.12.2014 г.).

Редактор и корректор В.А. Басова
Технический редактор Л.Я. Титова Темплан 2015, поз. 2

Подп. к печати 18.02.2015. Формат 60х84/16. Бумага тип № 1.
Печать офсетная. Объем 5,9 печ.л.; 5,9 уч. изд.л. Тираж 30 экз. Изд. № 2
Цена "С". Заказ

Ризограф Санкт-Петербургского государственного технологического университета растительных полимеров, СПб., 198095, ул. Ивана Черных,4.

ЛЕКЦИЯ 1
Предмет и содержание курса

Основные определения и типы моделей

В общем виде *модель* можно определить как условный образ (упрощенное изображение) реального объекта (процесса), который создается для более глубокого изучения действительности. *Модель есть материально или теоретически сконструированный объект, который заменяет (представляет) объект исследования в процессе познания, находится в отношении сходства с последним и более удобен для исследования.*

Метод исследования, базирующийся на разработке и использовании моделей, называется *моделированием*. Необходимость моделирования обусловлена сложностью, а порой и невозможностью прямого изучения реального объекта (процесса). Значительно доступнее создавать и изучать прообразы реальных объектов (процессов), т.е. модели. Можно сказать, что теоретическое знание о чем-либо, как правило, представляет собой совокупность различных моделей. Эти модели отражают существенные свойства реального объекта (процесса), хотя на самом деле действительность значительно содержательнее и богаче.

Подобие между моделируемым объектом и моделью может быть физическое, структурное, функциональное, динамическое, вероятностное и геометрическое. При физическом подобии объект и модель имеют одинаковую или сходную физическую природу. Структурное подобие предполагает наличие сходства между структурой объекта и структурой модели. При выполнении объектом и моделью под определенным воздействием сходных функций наблюдается функциональное подобие. При наблюдении за последовательно изменяющимися состояниями объекта и модели отмечается динамическое подобие, вероятностное подобие при наличии сходства между процессами вероятностного характера в объекте и модели, а геометрическое подобие при сходстве пространственных характеристик объекта и модели.

Важнейшая особенность модели состоит в возможности неограниченного накопления специализированных знаний без потери целостного взгляда на объект исследования. Моделирование процессов в обществе, природе и технических системах – это основная компонента системного подхода к познанию этих процессов и управлению ими.

Адекватность модели объекту исследований всегда ограничена и зависит от цели моделирования. Всякая модель не учитывает некоторые свойства оригинала и поэтому является его *абстракцией*. Смысл абстрагирования заключается в отвлечении от некоторых несущественных в данном контексте свойств предмета и одновременном выделении существенных свойств.

На сегодняшний день общепризнанной единой классификации моделей не существует. Однако из множества моделей можно выделить *словесные, графические, физические, экономико-математические и некоторые другие типы.*

Словесная, или монографическая, модель представляет собой словесное описание объекта, явления или процесса. Очень часто она выражается в виде определения, правила, теоремы, закона или их совокупности.

Графическая модель создается в виде рисунка, географической карты или чертежа. Например, зависимость между ценой и спросом может быть выражена в виде графика, на оси ординат которого отложен спрос (D), а на оси абсцисс цена (P). Кривая нам наглядно иллюстрирует, что с ростом цены спрос падает, и наоборот. Конечно, данную зависимость можно выразить и словесно, но графически она намного нагляднее (рис. 1).

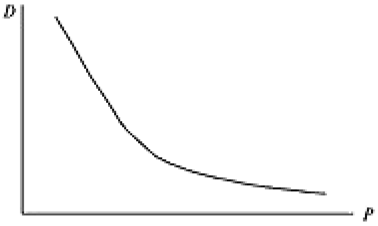


Рисунок 1 – Графическая модель, зависимость между спросом и ценой

Физические, или вещественные, модели создаются для конструирования пока еще несуществующих объектов. Создать модель самолета или ракеты для проверки ее аэродинамических свойств значительно проще и экономически целесообразнее, чем изучать эти свойства на реальных объектах.

Экономико-математические модели отражают наиболее существенные свойства реального объекта или процесса с помощью системы уравнений. Единой классификации экономико-математических моделей также не существует, хотя можно выделить наиболее значимые их группы в зависимости от признака классификации.

По степени агрегирования объектов моделирования различают модели:

- микроэкономические;
- одно-, двухсекторные (одно-, двухпродуктовые);
- многосекторные (многопродуктовые);
- макроэкономические;
- глобальные.

По учету фактора времени различают модели:

- статические;
- динамические.

В статических моделях система описана в статике, применительно к одному определенному моменту времени. Это как бы снимок, срез, фрагмент динамической системы в какой-то момент времени. Динамические модели описывают систему в развитии.

По цели создания и применения различают модели:

- балансовые;
- эконометрические;
- оптимизационные;
- сетевые;
- систем массового обслуживания;
- имитационные (экспертные).

В *балансовых моделях* отражается требование соответствия наличия ресурсов и их использования.

Параметры *эконометрических моделей* оцениваются с помощью методов математической статистики. Наиболее распространены эконометрические модели, представляющие собой системы регрессионных уравнений. В данных уравнениях отражается зависимость эндогенных (зависимых) переменных от экзогенных (независимых) переменных. Данная зависимость в основном выражается через тренд (длительную тенденцию) основных показателей моделируемой экономической системы. Эконометрические модели используются для анализа и прогнозирования конкретных экономических процессов с использованием реальной статистической информации.

Оптимизационные модели позволяют найти из множества возможных (альтернативных) вариантов наилучший вариант работы системы, производства, распределения или потребления. Ограниченные ресурсы при этом будут использованы наиболее эффективным образом для достижения поставленной цели.

Сетевые модели наиболее широко применяются в управлении проектами. Сетевая модель отображает комплекс работ (операций) и событий и их взаимосвязь во времени. Обычно сетевая модель предназначена для выполнения работ в такой последовательности, чтобы сроки выполнения проекта были минимальными. В этом случае ставится задача нахождения критического пути. Однако существуют и такие сетевые модели, которые ориентированы не на критерий времени, а, например, на минимизацию стоимости работ.

Модели систем массового обслуживания создаются для минимизации затрат времени на ожидание в очереди и времени простоев каналов обслуживания.

Имитационная модель наряду с машинными решениями содержит блоки, где решения принимаются человеком (экспертом). Вместо непосредственного участия человека в принятии решений может выступать база знаний. В этом случае ЭВМ, специализированное программное обеспечение, база данных и база знаний образуют экспертную систему. Экспертная система предназначена для решения одной или ряда задач методом имитации действий человека, эксперта в данной области.

По учету фактора неопределенности различают модели:

- детерминированные (с однозначно определенными результатами);
- стохастические (с различными вероятностными результатами).

По типу математического аппарата различают модели:

- линейного и нелинейного программирования;
- корреляционно-регрессионные;
- матричные;
- сетевые;
- теории игр;
- теории массового обслуживания и т.д.

ЛЕКЦИЯ 2
Основные понятия имитационного моделирования

Имитационное моделирование (ИМ) – распространённая разновидность аналогов моделирования, реализуемого с помощью набора математических инструментальных средств, специальных имитирующих программных средств и технологий программирования, позволяющих посредствам процессов аналогов провести целенаправленное исследование структуры и функций реального сложного процесса в памяти компьютера в режиме «имитации», выполнить оптимизацию некоторых его параметров.

Имитационной моделью (ИМ) называется специальный программный комплекс, позволяющий имитировать деятельность какого-либо сложного объекта. Он выполняет на компьютере параллельно взаимодействующие процессы, которые являются по своим временным параметрам (с точностью по масштабам времени и пространства) аналогами исследуемых процессов.

ИМ удобно для исследования практических задач: определение показателей эффективности, сравнение вариантов построения и алгоритмов функционирования систем, проверки устойчивости режимов системы при малых отклонениях входных переменных от расчётных значений. Полнота имитации может быть проверена путём построения серии последовательно уточняемых моделей. Если дальнейшая детализация свойств модели не влияет на конечные показатели, то усложнение модели можно прекратить. Как правило, моделируются те свойства процесса, которые могут влиять на выбранный показатель эффективности или критичны к наложенным ограничениям. Промежуточные результаты имитационного моделирования имеют четкий физический смысл и позволяют обнаружить ошибки программы.

Однако ИМ присущи и недостатки:

– большой расход машинного времени;
– малая точность вероятностных характеристик редких событий;
– трудность получения обобщающих выводов и рекомендаций;
– сложность оптимизации системы (многовариантность расчётов при наличии вероятностных помех);
– вероятностная оценка погрешности.

Таким образом применение ИМ становится целесообразным:

– для накопления первичных данных об изучаемом явлении, если эти данные нельзя получить в натурном эксперименте;
– для проверки планомерности допущений, сделанных разработчиком в целях перехода к аналитическим методам,
– для демонстрации конечных результатов исследования на достаточно полной модели реальной ситуации,
– при «безысходности», когда сложность ситуации намного превосходит возможности аналитических методов, известных разработчику.

Основные функции ИМ

Для создания ИМ необходима специальная система моделирования, имеющая набор языковых средств, сервисные подпрограммы, приёмы и технологии программирования. ИМ должна отражать большое число параметров, логику и закономерности поведения моделируемого объекта во времени (временная динамика), а для объектов экономики существует понятие финансовой динамики.

ИМ контролируемого объекта или процесса обеспечивается двумя видами деятельности, выполняемыми с помощью компьютера:

– работа по созданию или модификации ИМ;
– эксплуатация ИМ и интерпретация результатов.
– ИМ систем применяется в двух случаях:
– для управления сложным процессом, когда ИМ управляемого объекта используется в качестве инструментального средства в контуре адаптивной системы управления, создаваемой на основе имитационных технологий;
– при проведении экспериментов с дискретно-непрерывными моделями сложных объектов для получения и отслеживания их динамики в экстренных ситуациях, связанными с рисками, натурное моделирование которых нежелательно или невозможно.

Типовые задачи, решаемые средствами компьютерного моделирования

– моделирование процессов логистики для определения временных и стоимостных параметров;
– управление процессом реализации инвестиционного проекта на различных этапах его жизненного цикла с учётом возможных рисков и тактики выделения денежных средств;
– анализ процессов в работе сети кредитных организаций с учётом процессов взаимозачётов в условиях Российской Банковской Системы;
– прогнозирование финансовых результатов деятельности предприятия на конкретный период времени с учётом анализа динамики сальдо на счетах;
– бизнес-реинженеринг несостоятельного предприятия (изменяя его структуры ресурсов, прогноз финансовых результатов, выбор того или иного варианта реконструкции);
– анализ адаптивных, свойств живучести компьютерной региональной банковской информационной системы;
– оценка параметров надёжности и задержек в централизованной экономической информационной системе с возможностью коллективного доступа;
– анализ эксплуатационных параметров распределённой, многоуровневой, ведомственной информационной управляющей системы с учётом неоднородной структуры, пропускной способности каналов связи и не совершенства физической организации распределённой базы данных в региональных центрах;

– моделирование действий курьерской группы в регионе пострадавшем в результате природной катастрофы или промышленной аварии;

– анализ сетевой модели для проектов замены и наладки производственного оборудования с учётом возникновения неисправностей;

– анализ работы автотранспортного предприятия, занимающимся коммерческим перевозом грузов, с учётом спецификации товарных и денежных потоков в регионе;

– расчёт параметров надёжности и издержек обработки информации в банковской информационной системе.

Системы имитационного моделирования

1 период 1970-1980 гг. – впервые методы имитации для анализа экономических процессов применил Т. Нейлор. Однако применение имитационных методов для анализа носило эпизодический характер из-за сложности формализации экономических процессов, так как в математическом обеспечении ЭВМ отсутствовали формализованные языковые методы поддержки элементарных процессов и их функций, и отсутствовали формализуемые методы структурного системного анализа необходимые для иерархического разложения реального моделируемого процесса на элементарные составляющие.

Алгоритмические методы моделирования были достаточно трудоёмкими и при моделировании простых составляющих экономических процессов уступали математическим методам и решениям в аналитической форме.

Таким образом, ИМод применялось только в научной деятельности.

Однако для этого периода характерно появление первых технологичных средств ИМод, которые обладали свойствами инструментальных средств контролируемых процессов.

Одна из первых систем GPSS позволяла создавать модели контролируемых процессов и объектов в основном технического и технологического назначения.

2 период 1980-1990 гг. – характеризуется широким спектром появления новых систем ИМод (более 20 систем). Наиболее распространённые из них : GASP-4, SIMULA-67,GPSS-5, SLAM-2.

GASP-4 предоставила пользователю структурированный язык программирования, похожий на Фортран. Набор методов событийного моделирования дискретных подсистем модели и моделирования непрерывных систем с помощью уравнений переменных состояния, а так же датчики псевдослучайных чисел.

SIMULA-67 аналогичен, однако структурированный язык был похож на ALGOL-60.

Эффективность ИМод с помощью GASP and SIMULA в большей степени зависла от искусства программиста, в этих системах отсутствовали средства имитации пространственных динамических моделируемого процесса.

GPSS-5 – это законченная высокоуровневая информационная технология создания ИМ, имеющая средства формализованного описания параллельных дискретных процессов в виде графических изображений или с помощью операторов встроенного языка, при этом координация процессов осуществляется в едином модельном времени. Имелись средства управления моделью, динамическая отладка и автоматизация обработки результатов. Однако имелись три основных недостатка:

- разработчик не мог включать непрерывные динамические компоненты в модель, даже при условии подключения своих подпрограмм
- отсутствие средств имитации пространственных процессов
- GPSS является интерпретирующей системой

SLAM-2 является наиболее развитой перечисленных систем однако сложнее в освоении чем GPSS и отсутствуют средства имитации пространственных процессов.

3 период 1990-2000 гг. связан с появлением различных новых пакетов ИМод.

Process Charter 1.0.2. позволяет строить блок схемы моделей, ориентирован на дискретное моделирование, имеет интеллектуальный, удобный и простой механизм построения моделей, низкую стоимость, приспособлен для решения задач распределения ресурсов, однако имеет слабую поддержку моделирования непрерывных компонентов, ограниченный набор средств для анализа чувствительности и построения диаграмм.

Powersim 2.01 прекрасное средство для построения непрерывных моделей, имеет множество встроенных функций для построения модели, коллективные средства, средство обработки массивов данных, однако имеет сложную систему обозначений и ограниченную поддержку дискретного моделирования.

I think 3.0.61 обеспечивает создание непрерывных и дискретных моделей, имеет встроенные блоки для обеспечения создания различных видов моделей, поддержка авторского моделирования для слабо подготовленных пользователей, имеет обучающую программу, развиты средства для анализа чувствительности, поддержка множества форматов вводимых данных. Недостатки: сложная система обозначений, ограниченное количество поддерживаемых функций.

Extend + BPR 3.1 поддерживается дискретное непрерывное моделирование, имеет интуитивно понятную среду построения моделей, множество встроенных блоков и функций, поддержка сторонними компаниями, имеет средства создания дополнительных функций на встроенном языке, однако в полной мере работает с компьютерами Macintosh, имеет высокую стоимость.

ReThink аналогичен «Extend + BPR 3.1», отличается лучшим графическим транслятором.

Pilgrim обладает широким спектром возможностей имитации временной, пространственной и финансовой динамики моделируемых объектов. Позволяет создавать непрерывно-дискретные модели. Разрабатываемые модели имеют свойство коллективного управления процессом моделирования. В текст модели можно вставлять любые блоки на языке С++. Большое быстродействие, сравнительно не высокая стоимость.

В России так же разработаны системы ИМод : РДО (ресурсы действия операции) МГТУ имени Баумана, мощное средство для создания продукционной модели.

Обладает разными средствами компьютерной графики, используется для моделирования технологических и производственных процессов.

Структурный анализ процесса (определение структуры)

Осуществляется формализация структуры сложного процесса путём декомпозиции (разложение на подпроцессы) которые выполняют определённые функции и имеют взаимно функциональные связи согласно легенде разработанной рабочей экспертной группой. Каждый из этих подпроцессов может разлагаться в свою очередь на внутренние подпроцессы образуя иерархию имеющую многослойную структуру, результатом является формализованное изображение ИМ в графическом виде.

Особенно структурный анализ эффективен для моделирования экономических процессов, где многие составляющие подпроцессы не имеют под собой физической основы, протекают виртуально поскольку оперируют информацией, деньгами и логикой их обработки.

Формализованное описание модели

Графическое изображение ИМ, функций, каждого подпроцесса, условий взаимодействия всех подпроцессов и особенности поведения моделируемого процесса (временная, пространственная и финансовая динамика) должны быть описаны на специальном языке для последующей трансляции. Существуют следующие способы:

– описание вручную на языке типа GPSS, Pilgrim, или Visual Basic;

– автоматизированное описание с помощью компьютерного графического конструктора во время проведения структурного анализа, т. е. С незначительными затратами на программирование.

Построение модели

Это трансляция и редактирование связей (сборка модели), верификация (калибровка) параметров:

– в режиме интерпретации (GPSS, SLAM-2, ReThink,)

– в режиме компиляции (Pilgrim)

Режим интерпретации проще в реализации. Специальная программа – интерпретатор на основании формализованного описания модели запускает все имитационные подпрограммы. Неудобство – невозможность передачи модели без системы моделирования.

Использование режима компиляции приводит к получению модели в виде независимого программного продукта.

Верификация параметров модели выполняется в соответствии с легендой, на основании которой построена модель, с помощью специальных тестовых примеров.

Проведение эксперимента

Для оптимизации определённых параметров реального процесса производится сначала планирование эксперимента и прогонов, затем сам машинный эксперимент, анализ результатов, интерпретация, реализация и документирование.

ЛЕКЦИЯ 3-4
Использование регрессионного и корреляционного анализа
для моделирования систем

Понятие корреляционного и регрессионного анализа

Для решения задач экономического анализа и прогнозирования очень часто используются статистические, отчетные или наблюдаемые данные. При этом полагают, эти данные являются значениями случайной величины.

Случайной величиной называется переменная величина, которая в зависимости от случая принимает различные значения с некоторой вероятностью. Закон распределения случайной величины показывает частоту ее тех или иных значений в общей их совокупности.

При исследовании взаимосвязей между экономическими показателями на основе статистических данных часто между ними наблюдается стохастическая зависимость. Она проявляется в том, что изменение закона распределения одной случайной величины происходит под влиянием изменения другой. Взаимосвязь между величинами может быть полной (функциональной) и неполной (искаженной другими факторами).

Пример функциональной зависимости выпуск продукции и ее потребление в условиях дефицита.

Неполная зависимость наблюдается, например, между стажем рабочих и их производительностью труда. Обычно рабочие с большим стажем трудятся лучше молодых, но под влиянием дополнительных факторов образование, здоровье и т.д. эта зависимость может быть искажена.

Раздел математической статистики, посвященный изучению взаимосвязей между случайными величинами, называется *корреляционным анализом* (*от лат.* correlatio соотношение, соответствие). Основная задача корреляционного анализа это установление характера и тесноты связи между результативными (зависимыми) и факторными (независимыми) показателями (признаками) в данном явлении или процессе. Корреляционную связь можно обнаружить только при массовом сопоставлении фактов.

Характер связи между показателями определяется по корреляционному полю. Если y зависимый признак, а x независимый, то, отметив каждый случай x (i) с координатами x_i и y_i, получим корреляционное поле. По расположению точек можно судить о характере связи (рис. 2.1).

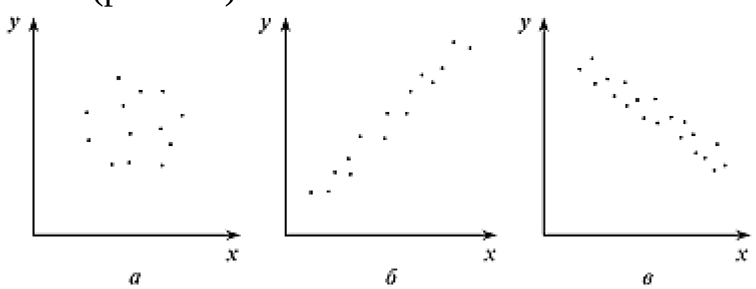


Рисунок 2.1 – Примеры корреляционных полей:

а переменные *x* и *y* не коррелируют; *б* наблюдается сильная положительная корреляция; *в* наблюдается слабая отрицательная корреляция

Теснота связи определяется с помощью коэффициента корреляции, который рассчитывается специальным образом и лежит в интервалах от минус единицы до плюс единицы. Если значение коэффициента корреляции лежит в интервале от 1 до 0,9 по модулю, то отмечается очень сильная корреляционная зависимость. В случае, если значение коэффициента корреляции лежит в интервале от 0,9 до 0,6, то говорят, что имеет место слабая корреляционная зависимость. Наконец, если значение коэффициента корреляции находится в интервале от - 0,6 до 0,6, то говорят об очень слабой корреляционной зависимости или полном ее отсутствии.

Таким образом, корреляционный анализ применяется для нахождения характера и тесноты связи между случайными величинами.

Регрессионный анализ своей целью имеет вывод, определение (идентификацию) уравнения регрессии, включая статистическую оценку его параметров. Уравнение регрессии позволяет найти значение зависимой переменной, если величина независимой или независимых переменных известна.

Практически, речь идет о том, чтобы, анализируя множество точек на графике (т.е. множество статистических данных), найти линию, по возможности точно отражающую заключенную в этом множестве закономерность (тренд, тенденцию), линию регрессии.

По числу факторов различают одно-, двух- и многофакторные уравнения регрессии.

По характеру связи однофакторные уравнения регрессии подразделяются:

а) на линейные:

$$y = a + bx,$$

где x экзогенная (независимая) переменная, y эндогенная (зависимая, результативная) переменная, a , b параметры;

б) степенные: $y = a \cdot x^b$,

в) показательные: $y = a \cdot b^x$,

г) прочие.

Определение параметров линейного однофакторного уравнения регрессии

Пусть у нас имеются данные о доходах (x) и спросе на некоторый товар (y) за ряд лет (n):

Год	Доход	Спрос
i	x	y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
...	...	...
n	x_n	y_n

Предположим, что между x и y существует линейная взаимосвязь, т.е.

$$y = a + bx.$$

Для того, чтобы найти уравнение регрессии, прежде всего нужно исследовать тесноту связи между случайными величинами x и y , т.е. корреляционную зависимость.

Пусть

$x_1, x_2, ..., x_n$ совокупность значений независимого, факторного признака;

$y_1, y_2, ..., y_n$ совокупность соответствующих значений зависимого, результативного признака;

n количество наблюдений.

Для нахождения уравнения регрессии вычисляются следующие величины:

1. Средние значения

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$ для экзогенной переменной;

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$ для эндогенной переменной.

2. Отклонения от средних величин

$$\Delta x_i = x_i - \bar{x}, \quad \Delta y_i = y_i - \bar{y}.$$

3. Величины дисперсии и среднего квадратичного отклонения

$$D_x = \frac{\sum_{i=1}^{n} \Delta x_i^2}{n-1}, \quad D_y = \frac{\sum_{i=1}^{n} \Delta y_i^2}{n-1};$$

$$\sigma_x = \sqrt{D_x}, \quad \sigma_y = \sqrt{D_y}.$$

Величины дисперсии и среднего квадратичного отклонения характеризуют разброс наблюдаемых значений вокруг среднего значения. Чем больше дисперсия, тем больше разброс.

4. Вычисление корреляционного момента (коэффициента ковариации):

$$K_{x,y} = \frac{\Delta x_1 \cdot \Delta y_1 + \Delta x_2 \cdot \Delta y_2 + \ldots + \Delta x_n \cdot \Delta y_n}{n-1} = \frac{\sum_{i=1}^{n} \Delta x_i \cdot \Delta y_i}{n-1}.$$

Корреляционный момент отражает характер взаимосвязи между x и y . Если $K_{x,y} > 0$, то взаимосвязь прямая. Если $K_{x,y} < 0$, то взаимосвязь обратная.

5. Коэффициент корреляции вычисляется по формуле

$$R_{x,y} = \frac{K_{x,y}}{\sigma_x \sigma_y}.$$

Доказано, что коэффициент корреляции находится в интервале от минус единицы до плюс единицы ($-1 >= R_{x,y} <= 1$). Коэффициент корреляции в квадрате $(R^2_{x,y})$ называется коэффициентом детерминации.

Если $R_{x,y} > |0,8|$, то вычисления продолжаются.

6. Вычисления параметров регрессионного уравнения.

Коэффициент b находится по формуле

$$b = \frac{K_{x,y}}{D_x}.$$

После чего можно легко найти параметр a :

$$a = \bar{y} - b\bar{x}.$$

Коэффициенты a и b находятся методом наименьших квадратов, основная идея которого состоит в том, что за меру суммарной погрешности принимается сумма квадратов разностей (остатков) между фактическими значениями результативного признака y_i и его расчетными значениями $y_{i\,р}$, полученными при помощи уравнения регрессии

$$y_{i\,р} = a + bx_i.$$

При этом величины остатков находятся по формуле

$$u_i = y_i - y_{i\,р},$$

где y_i фактическое значение y ; $y_{i\,р}$ расчетное значение y .

Оценка величины погрешности линейного однофакторного уравнения

1. Обозначим разность между фактическим значением результативного признака и его расчетным значением как u_i : $u_i = y_i - y_{i\,р}$,

где y_i фактическое значение y ; $y_{i\,р}$ расчетное значение y ; u_i разность между ними.

2. В качестве меры суммарной погрешности выбрана величина

$$S = \frac{\sum_{n-2}^{n} u_i^2}{n-2}.$$

Для нашего примера S = 0,432.

Поскольку $\bar{u}$ (среднее значение остатков) равно нулю, то суммарная погрешность равна остаточной дисперсии.

3. Остаточная дисперсия находится по формуле

$$D_u = \frac{\sum (u_i - \bar{u})^2}{n-2} = \frac{\sum u_i^2}{n-2} = S.$$

Для нашего примера $D_u = 0{,}432$. Можно показать, что

$$D_u = \left(1 - R^2_{x,y}\right) \cdot D_y.$$

Если $R^2_{x,y} = 1$, то $D_u = 0$;

$R^2_{x,y} = 0$, то $D_u = D_y$.

Таким образом, $0 \le D_u \le D_y$.

Легко заметить, что если

$R_{x,y} = 0{,}9$, то $D_u = (1 - 0{,}81) \cdot D_y = 0{,}91 \cdot D_y$.

Это соотношение показывает, что в экономических приложениях допустимая суммарная погрешность может составить не более 20% от дисперсии результативного признака D_y.

4. Стандартная ошибка уравнения находится по формуле

$$\sigma_u = \sqrt{D_u},$$

где D_u остаточная дисперсия. В нашем случае $\sigma_u = 0{,}6572.$

5. Относительная погрешность уравнения регрессии вычисляется как

$$\vartheta = \frac{\sigma_u}{\bar{y}} \cdot 100\%,$$

где σ_u стандартная ошибка; $\bar{y}$ среднее значение результативного признака.

В нашем случае $\vartheta = 7{,}07\%.$

Если величина ϑ мала и отсутствует автокорреляция остатков, то прогнозные качества оцененного регрессионного уравнения высоки.

6. Стандартная ошибка коэффициента *b* вычисляется по формуле

$$S_b = \frac{\sigma_u}{\sqrt{nD_x}}.$$

В нашем случае она равна $S_b = 0{,}07171.$

Для вычисления стандартной ошибки коэффициента *a* используется формула

$$S_a = \sigma_u \sqrt{\frac{D_x + \bar{x}^2}{n \cdot D_x}}.$$

В нашем примере $S_a = 1{,}108.$

Стандартные ошибки коэффициентов используются для оценивания параметров уравнения регрессии.

Коэффициенты считаются значимыми, если

$$\frac{S_a}{|a|} < 0{,}5; \ \frac{S_b}{|b|} < 0{,}5.$$

В нашем примере

$$\frac{S_a}{|a|} = \frac{1{,}108}{|0{,}3|} = 3{,}69, \ \frac{S_b}{|b|} = \frac{0{,}07171}{0{,}64} = 0{,}112.$$

Коэффициент *a* не значим, так как указанное отношение больше 0,5, а относительная погрешность уравнения регрессии слишком высока 26,7%.

Стандартные ошибки коэффициентов используются также для оценки статистической значимости коэффициентов при помощи *t* -критерия Стьюдента. Значения *t* -критерия Стьюдента содержатся в справочниках по математической статистике. В таблице 2.1 приводятся его некоторые значения.

Далее находятся максимальные и минимальные значения параметров (b^- , b^+) по формулам:

$$b^- = b - t_{Ст} \cdot S_b,$$
$$b^+ = b + t_{Ст} \cdot S_b.$$

Таблица 2.1 – Некоторые значения *t* -критерия Стьюдента

Степени свободы	Уровень доверия (*c*)	
(*n*- 2)	0,90	0,95
1	6,31	12,71
2	2,9	4,30

	2	
3	2,35	3,18
4	2,13	2,78
5	2,02	2,57

Для нашего примера находим

$$b^{-} = 0,64 - 2,78 \cdot 0,07171 = 0,44,$$

$$b^{+} = 0,64 + 2,78 \cdot 0,07171 = 0,839.$$

Если интервал (b^- , b^+) достаточно мал и не содержит ноль, то коэффициент b является статистически значимым на c -процентном доверительном уровне.

Аналогично находятся максимальные и минимальные значения параметр a . Для нашего примера

$$a^{-} = -0,3 - 2,78 \cdot 1,108 = -3,38,$$

$$a^{+} = -0,3 + 2,78 \cdot 1,108 = 2,78.$$

Коэффициент a не является статистически значимым, так как интервал (a^- , a^+) велик и содержит ноль.

Вывод: полученные результаты не являются значимыми и не могут быть использованы для прогнозных расчетов. Ситуацию можно поправить следующими способами:

а) увеличить число n ;

б) увеличить количество факторов;

в) изменить форму уравнения.

Проблема автокорреляции остатков. Критерий ДарбинаУотсона

Часто для нахождения уравнений регрессии используются динамические ряды, т.е. последовательность экономических показателей за ряд лет (кварталов, месяцев), следующих друг за другом.

В этом случае имеется некоторая зависимость последующего значения показателя от его предыдущего значения, которое называется автокорреляцией. В некоторых случаях зависимость такого рода является весьма сильной и влияет на точность коэффициента регрессии.

Пусть уравнение регрессии построено и имеет вид:

$$y_t = a + bx_t + u_t, \quad t = 1, 2, \dots, n,$$

где u_t погрешность уравнения регрессии в год t .

Явление автокорреляции остатков состоит в том, что в любой год t остаток u_t не является случайной величиной, а зависит от величины остатка предыдущего года u_{t-1} . В результате при использовании уравнения регрессии могут быть большие ошибки.

Для определения наличия или отсутствия автокорреляции применяется критерий ДарбинаУотсона:

$$DW = \frac{\sum_{t=2}^{n}\left(u_t - u_{t-1}\right)^2}{\sum_{t=1}^{n} u_t^2}.$$

Возможные значения критерия *DW* находятся в интервале от 0 до 4. Если автокорреляция остатков отсутствует, то *DW* » 2.

Построение уравнения степенной регрессии

Уравнение степенной агрессии имеет вид:

$$y = a \cdot x^b,$$

где a , b параметры, которые определяются по данным таблицы наблюдений.

Таблица наблюдений составлена и имеет вид

x x_1 x_2 ... x_n

y y_1 y_2 ... y_n

Прологарифмируем исходное уравнение и в результате получим

$$\ln y = \ln a + b \cdot \ln x.$$

Обозначим ln y через y' , ln a как a' , а ln x как x' .

В результате подстановки получим $y' = a' + bx'$.

Данное уравнение есть не что иное, как уравнение линейной регрессии, параметры которого мы умеем находить.

Для этого прологарифмируем исходные данные:

l ln ln . ln

n x x_1 x_2 .. x_n

l ln ln . ln

n y y_1 y_2 .. y_n

Далее необходимо выполнить известные нам вычислительные процедуры по нахождению коэффициентов a и b , используя прологарифмированные исходные данные. В результате получим значения коэффициентов b и a' . Параметр a можно найти по формуле $a = e^{a'}$.

Двухфакторные и многофакторные уравнения регрессии

Линейное двухфакторное уравнение регрессии имеет вид

$$y = a + b_1 x_1 + b_2 x_2,$$

где a , b_1 , b_2 параметры; x_1 , x_2 экзогенные переменные; y эндогенная переменная.

Степенное двухфакторное уравнение регрессии имеет вид

$$y = a x_1^{\alpha} \cdot x_2^{\beta},$$

где a , a , b параметры; x_1 , x_2 экзогенные переменные; y эндогенная переменная.

Для нахождения параметров этого уравнения его необходимо прологарифмировать. В результате получим

$$\ln y = \ln a + \alpha \ln x_1 + \beta \ln x_2.$$

Следует помнить, что мы получим не параметр *a* , а его логарифм, который следует преобразовать в натуральное число.

Линейное многофакторное уравнения регрессии имеет вид

$$y = a + b_1 x_1 + \ldots + b_n x_n,$$

где a , b_1 , b_n параметры; x_1 , x_n экзогенные переменные; y эндогенная переменная.

ЛЕКЦИЯ 5-6
Оптимизация и оптимизационные модели

Основные понятия

Оптимизация - поиск наилучшего решения с учетом ограничений.

Для оптимизации ищется целевая функция. Эта функция конструируется искусственно на основе уравнений, описывающих объект оптимизации. Целевая функция обычно имеет много аргументов: φ=f (x1, x2, ..., x n).

Чтобы найти оптимальное значение, перебирают значение аргументов xi пошагово до тех пор, пока значение φ станет удовлетворять условиям оптимума. Даже количество аргументов не более трех, "тупой" перебор может потребить очень много времени.

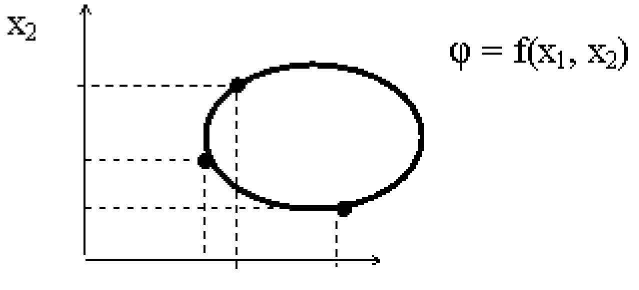
Поэтому разработаны десятки *методов оптимизации*:

- первый строгий математический метод предложил в 1840г. венгерский математик Коши - МСС - *метод скорейшего спуска*. При формулировании задач оптимизации обычно стараются ее свести к поиску минимума. МСС относится к классу градиентных методов.

Градиент - вектор, указывающий на направление максимального возрастания функции.

Антиградиент - вектор, указывающий на направление максимального убывания функции. Чтобы повернуть вектор на 180╟, достаточно изменить все знаки у градиентов на противоположные (т.е. x (-1)).

Для иллюстрации поиска экстремума в процессе оптимизации функций двух переменных используют *линии равного уровня (ЛРУ)*. Если задаться постоянным значением φ и так подбирать значения xi чтобы значение φ было равным заданному значению, то геометрическое место точек φ составит *линию равного уровня*.

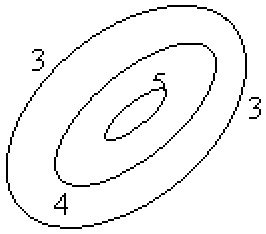


В зависимости от целевой функции линий равного уровня могут характеризоваться следующими географическими понятиями:

Долина - когда соседние линии равного уровня изменяется очень слабо в широком диапазоне аргументов.

Возвышенность - когда соседние линии равного уровня представляют собой замкнутые линии и значение φ возрастает от внешних линий к внутренним.

Впадина - когда соседние линии равного уровня представляют собой замкнутые линии, и значение φ убывает от внешних линий к внутренним.



Седловина - локальный минимум, в центре которого векторы указывают на возрастание функции, но вскоре направление вектора резко изменяется вверх или вниз.

МСС - простейший метод оптимизации, пригодный для сложных систем. Работа метода хорошо иллюстрируется с помощью *линий равного уровня (ЛРУ)*.

Порядок поиска оптимума:

- выбирается исходная точка в виде значений параметров целевой функции:

φ=f (x1, x2, ..., x n).

- ищется градиент;

- движемся в направлении антиградиента с заданным шагом;

- на каждом шаге проверяем выполнение условия движения, оно такое: φi < φi-1 (текущее значение φ должно быть меньше предыдущего).

- если условие движения нарушается, то процесс останавливается, иначе, движение продолжается;

- при нарушении условий движения уточняется одномерный минимум и ищется новый градиент;

- условие останова:

а) значение φ меньше заданного;

б) разность значений соседних φ меньше заданной;

в) количество шагов превышает допустимое.

- если после останова минимума не удовлетворяет требованиям, то либо ищется другая исходная точка и процесс повторяется, либо выбирается другой метод оптимизации.

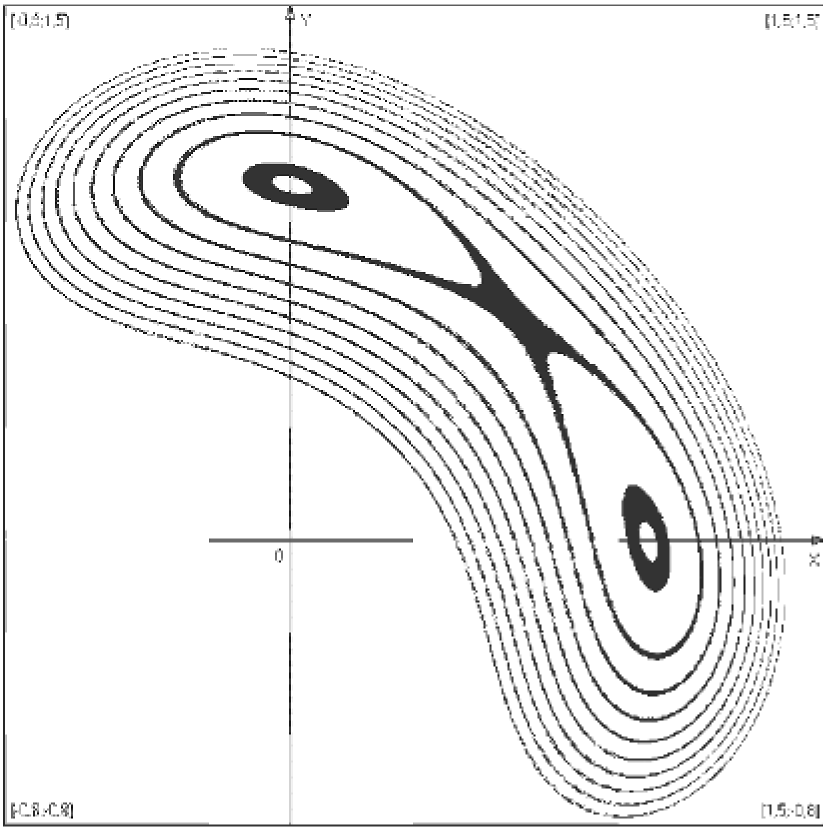
Конструирования целевой функции

Допустим, объект оптимизации описывается следующей системой уравнений:

$x^2 + y^2 = 1$

$x + y = 1$

Графически эту систему можно представить как окружность и секущая прямая



Задача: найти минимальное расстояние между точками прямой и окружности.

В данном случае два минимума. Последовательность конструкция целая функции:

1. приводим систему уравнений к нулевому виду:

$x^2 + y^2 - 1 = 0$

$x + y -1 = 0$

2. складываем два уравнения:

1 ур. + 2 ур. = 0

3. для усиления чувствительности к изменению аргументов обе части уравнения можно возвести в квадрат:

$(1\ \text{ур.})^2 + (2\ \text{ур.})^2 = 0$

4. Целевую функцию представим в виде:

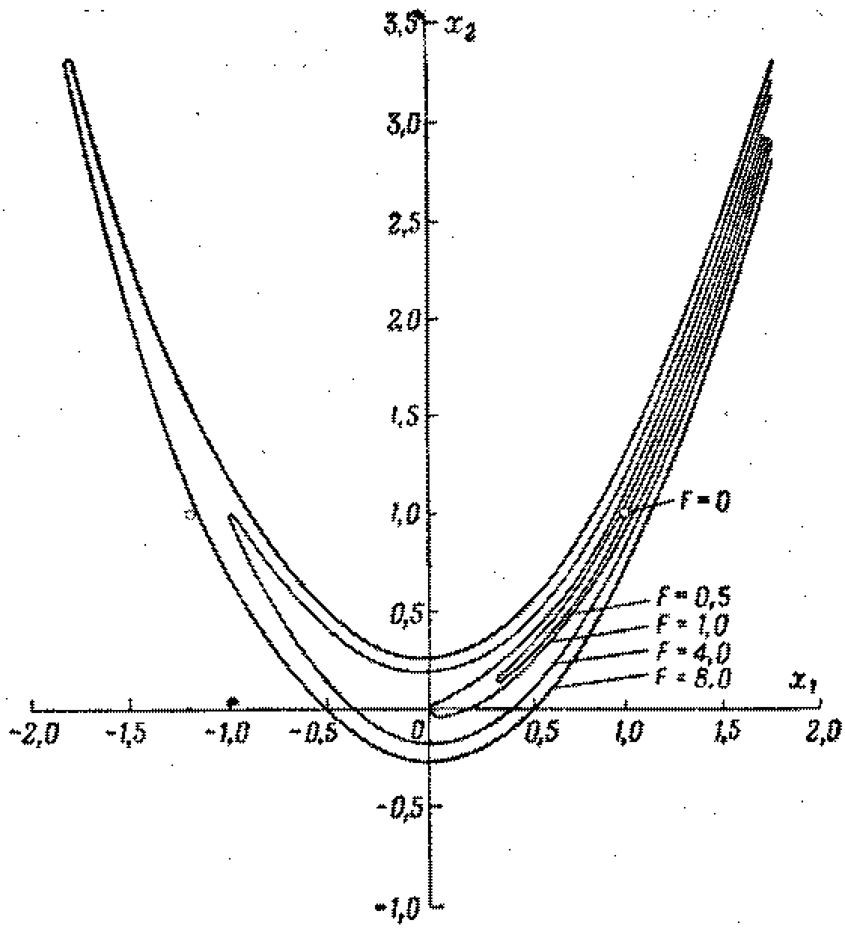
$\varphi = (x^2 + y^2 - 1)^2 + (x + y - 1)^2$;

Численный поиск минимума

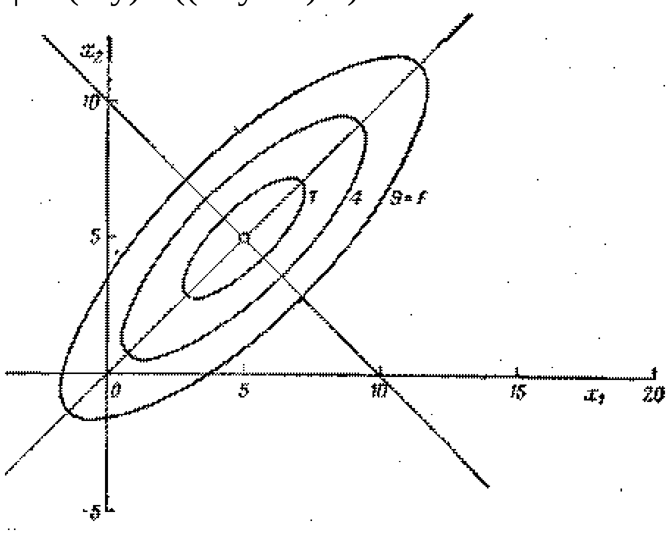
Составляется компьютерная программа, в которой значение x и y изменяется с заданным шагом в диапазоне +- 2. На каждом шаге вычисляется значение φ, с помощью плоттера рисуется картина ЛРУ.

Таким способом были построены ЛРУ еще для двух следующих целевых функций:

$\varphi = 100(x^2-y)^2+(1-x)^2$



$\varphi = (x-y)^2+((x+y-10)/3)^2$



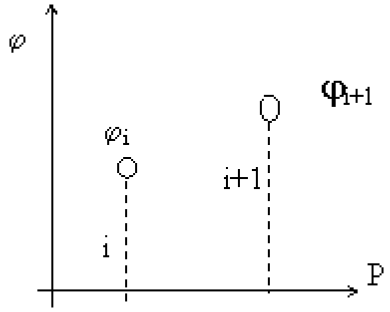
Многомерный и одномерный поиск оптимума

МСС представляет собой многомерный поиск, т.к. минимум ищется на разных направлениях. Когда минимум ищется только в одном направлении для уточнения направления следующего уровня - одномерный поиск.

Одномерный поиск

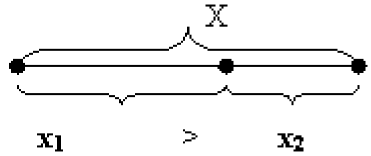
Для многомерного поиска разработаны десятки методов, для одного поиска около 1 десятка методов. Рассмотрим одномерное приближение.

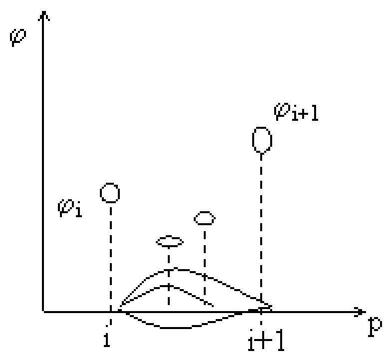
Метод последовательных приближений



P - длина шага оптимизации;
φ - значение целевой функции
1. при нарушении условий движения (φi+1 > φi) движение останавливается
2. Возвращается на 1 шаг назад.
3. Делим длину шага на R где R = 3-10
4. Возобновляем движение с новым шагом.
5. При нарушении условий движения все повторяется, и т.д.
Условия останова:
- Значение j < заданного
- Разность между соседними значениями j < заданной
- Длина шага < заданной
- Кол-во шагов превышает заданное.
Любое из этих условий приводит к останову.

Метод золотого сечения
Если возьмем пропорцию:
$x_1/x = x_2/x_1 = 0.618\text{-}m_o$





Такое соотношение называется золотой пропорцией.

1. При нарушении условий движения последний шаг делим в отношении золотой пропорции слева на право.

2. Этот же отрезок делим в золотой пропорции справа на лево. В результате получим 2 новые точки

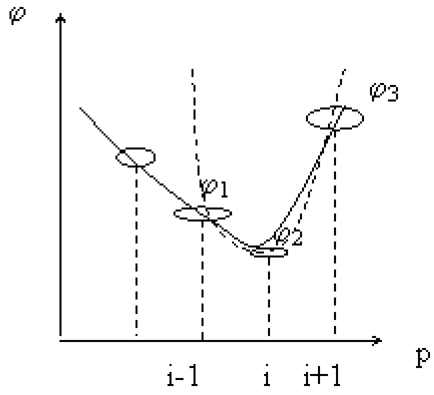
3. Сравниваем значения j в новых точках.

4. Выбираем отрезок, которому соответствует меньшее из этих двух j.

5. Полученный отрезок делим в отношении золотой пропорции слева направо, и т.д.

Условия останова те же, что и в предыдущем случае.

Метод параболической аппроксимации (МПА)



При нарушении условий значения j в последних 3-х точках подставляется в формулу решения системы 3-х уравнений для параболы. Это решение позволяет находить координаты минимума параболы, проходящий через 3 последние точки.

Сравнение методов одномерного поиска

МПП более прост (движемся, делим), но требует много шагов (м.б. 10 и 100 шагов).

МЗС позволяет найти min за 3-4 шага.

МПА более сложен, но позволяет найти min за 1 шаг. Но МПА обладает методической погрешностью, поскольку парабола отличается от истинной кривой; обычно эта погрешность невелика. В пакетах программ для расчета оптики обычно используется в качестве метода многомерного поиска демнорированый МСС, а в качестве метода одномерного поиска - МПА.

ЛЕКЦИЯ 7-8
Оптимизационные модели

Понятие оптимизационных задач и оптимизационных моделей

Экономико-математические задачи, цель которых состоит в нахождении наилучшего (оптимального) с точки зрения некоторого критерия или критериев варианта использования имеющихся ресурсов (труда, капитала и пр.), называются *оптимизационными*.

Оптимизационные задачи (ОЗ) решаются с помощью оптимизационных моделей (ОМ) методами математического программирования.

Структура оптимизационной модели состоит из целевой функции, области допустимых решений и системы ограничений, определяющих эту область. Целевая функция в самом общем виде, в свою очередь, также состоит из трех элементов:

- управляемых переменных;
- неуправляемых переменных;

- формы функции (вида зависимости между ними).

Область допустимых решений это область, в пределах которой осуществляется выбор решений. В экономических задачах она ограничена наличными ресурсами, условиями, которые записываются в виде системы ограничений, состоящей из уравнений и неравенств.

Если система ограничений несовместима, то область допустимых решений является пустой. Ограничения подразделяются:

а) на линейные (*I* и *II*) и нелинейные (*III* и *IV*) (рис. 3.1);

б) детерминированные (A , B) и стохастические (группы кривых C_i) (рис. 3.2).

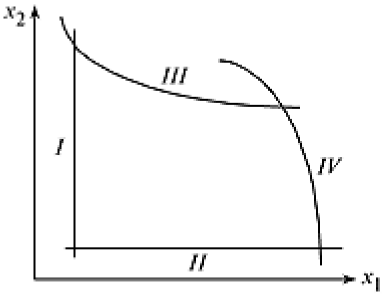


Рисунок 3.1– Линейные и нелинейные ограничения

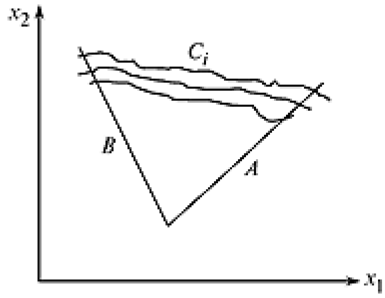


Рисунок 3.2– Детерминированные и стохастические ограничения

Стохастические ограничения являются возможными, вероятностными, случайными.

ОЗ решаются методами математического программирования, которые подразделяются:

- на линейное программирование;
- нелинейное программирование;
- динамическое программирование;
- целочисленное программирование;
- выпуклое программирование;
- исследование операций;
- геометрическое программирование и др.

Главная задача математического программирования это нахождение экстремума функций при ограничениях в форме уравнений и неравенств.

Рассмотрим ОЗ, решаемые методами линейного программирования.

Оптимизационные задачи с линейной зависимостью между переменными

Пусть:

b_i количество ресурса вида i (i = 1, 2, ..., m);

$a_{i,j}$ норма расхода i -того ресурса на единицу j -того вида продукции;

x_j количество продукции вида j (j = 1, 2, ..., n);

c_j прибыль (доход) от единицы этой продукции (в задачах на минимум себестоимость продукции).

Тогда ОЗ линейного программирования (ЛП) в общем виде может быть сформулирована и записана следующим образом:

Найти переменные x_j (j = 1, 2, ..., n), при которых целевая функция

$$F(x) = \sum_{j=1}^{n} c_j x_j \to \max (\min),$$

была бы максимальной (минимальной), не нарушая следующих ограничений:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = 1, 2, \ldots, m_1,$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = m_1 + 1, m_1 + 2, \ldots, m_2,$$

$$\sum_{j=1}^{n} a_{ij} x_j \ge b_i, \quad i = m_2 + 1, m_2 + 2, \ldots, m.$$

Все три случая можно привести к так называемой канонической форме, введя дополнительные переменные

$$\sum_{j=1}^{n+k} a_{ij} x_j = b_i, \quad i = 1, 2, \ldots, m,$$

где k количество дополнительных переменных, и условие неотрицательности искомых переменных: x_j ³ 0.

В результате решения задачи находится некий план (программа) работы некоторого предприятия. Отсюда и появилось слово программирование. Слово линейное указывает на линейный характер зависимости как в целевой функции, так и в системе ограничений. Следует еще раз подчеркнуть, что задача обязательно носит экстремальный характер, т.е. состоит в отыскании максимума или минимума (экстремума) целевой функции.

Геометрическая интерпретация ОЗЛП

Пусть необходимо найти оптимальный план производства двух видов продукции (x_1 и x_2), т.е. такой план, при котором целевая функция (общая прибыль) была бы максимальной, а имеющиеся ресурсы использовались бы наилучшим образом. Условия задачи приведены в табл. 3.1.

Таблица 3.1 – **Данные о запасе и нормах расхода ресурсов**

Вид продукции	Норма расхода ресурса на единицу продукции			Прибыль на единицу изделия
	A	*B*	*C*	
1	2	0,1	3,5	4
2	1	0,5	1	5
Объем ресурса	12	4	18	

Оптимизационная модель задачи запишется следующим образом:

а) целевая функция:

$$4x_1 + 5x_2 \to \max;$$

б) ограничения:

$2x_1 + x_2 \le 12$ (ограничение по ресурсу A),

$0{,}1x_1 + 0{,}5x_2 \le 4$ (ограничение по ресурсу B),

$3{,}5x_1 + x_2 \leq 18$ (ограничение по ресурсу *C*);

в) условие неотрицательности переменных:

$$x_1 \geq 0; \quad x_2 \geq 0.$$

Данную и подобные оптимизационные модели можно продемонстрировать графически (рис. 3.3).

Преобразуем нашу систему ограничений, найдя в каждом из уравнений x_2, и отложим их на графике. Любая точка на данном графике с координатами x_1 и x_2 представляет вариант искомого плана. Однако ограничение по ресурсу *A* сужает область допустимых решений. Ими могут быть все точки, ограниченные осями координат и прямой *AA* , так как не может быть израсходовано ресурса *A* больше, чем его на предприятии имеется. Если точки находятся на самой прямой, то ресурс используется полностью.

Аналогичные рассуждения можно привести и для ресурсов *B* и *C* . В результате условиям задачи будет удовлетворять любая точка, лежащая в пределах заштрихованного многоугольника. Данный многоугольник называется областью допустимых решений.

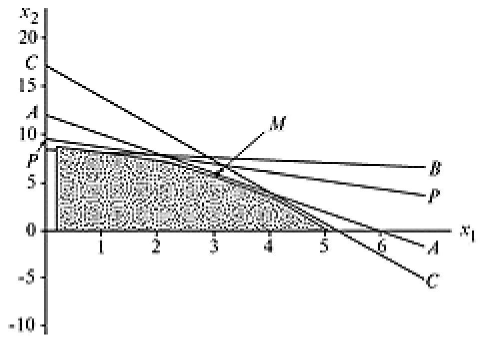


Рисунок 3.3.– Геометрическая интерпретация оптимизационной задачи линейного программирования

Однако нам необходимо найти такую точку, в которой достигался бы максимум целевой функции. Для этого построим произвольную прямую $4x_1 + 5x_2 = 20$, как $x_2 = 4 - 4/5x_1$ (число 20 произвольное). Обозначим эту линию *PP* . В каждой точке этой линии прибыль одинакова. Перемещая эту линию параллельно ее исходному положению, найдем точку, которая удалена от начала координат в наибольшей мере, однако не выходит за пределы области допустимых решений. Это точка *M* , которая лежит на вершине многоугольника. Координаты этой точки ($x'_1 = 3{,}03$ и $x'_2 = 7{,}4$) и будут искомым оптимальным планом.

Симплексный метод решения ОЗЛП

Симплексный метод это вычислительная процедура, основанная на принципе последовательного улучшения решений при переходе от одной базисной точки (базисного решения) к другой. При этом значение целевой функции улучшается.

Базисным решением является одно из допустимых решений, находящихся в вершинах области допустимых значений. Проверяя на оптимальность вершину за вершиной, приходят к искомому оптимуму. На этом принципе основан симплекс -метод.

Симплекс это выпуклый многогранник в n -мерном пространстве с $n + 1$ вершинами, не лежащими в одной гиперплоскости (гиперплоскость делит пространство на два полупространства).

Например, линия бюджетных ограничений делит блага на доступные и недоступные.

Доказано, что если оптимальное решение существует, то оно обязательно будет найдено через конечное число итераций (шагов), кроме случаев зацикливания.

Алгоритм симплексного метода состоит из ряда этапов.

Первый этап . Строится исходная ОМ. Далее исходная матрица условий преобразуется в приведенную каноническую форму, которая среди всех других канонических форм выделяется тем, что:

а) правые части условий (свободные члены b_i) являются величинами неотрицательными;

б) сами условия являются равенствами;

в) матрица условий содержит полную единичную подматрицу.

Если свободные члены отрицательные, то обе части неравенства умножаются на -1, а знак неравенства меняется на противоположный. Для преобразования неравенств в равенства вводятся дополнительные переменные, которые обычно обозначают объем недоиспользованных ресурсов. В этом их экономический смысл.

Наконец, если после добавления дополнительных переменных матрица условий не содержит полную единичную подматрицу, то вводятся искусственные переменные, которые не имеют никакого экономического смысла. Они вводятся исключительно для того, чтобы получить единичную подматрицу и начать процесс решения задачи при помощи симплексного метода.

В оптимальном решении задачи все искусственные переменные (ИП) должны быть равными нулю. Для этого вводят ИП в целевую функцию задачи с большими отрицательными коэффициентами ($-M$) при решении задачи на max, и с большими положительными коэффициентами ($+M$), когда задача решается на min. В этом случае даже небольшое ненулевое значение ИП будет резко уменьшать (увеличивать) значение целевой функции. Обычно M в 1000 раз должно быть больше, чем значения коэффициентов при основных переменных.

Второй этап . Строится исходная симплекс-таблица и отыскивается некоторое начальное базисное решение . Множество переменных, образующих единичную подматрицу, принимается за начальное базисное решение. Значения этих переменных равны свободным членам. Все остальные внебазисные переменные равны нулю.

Третий этап . Проверка базисного решения на оптимальность осуществляется при помощи специальных оценок коэффициентов целевой функции. Если все оценки коэффициентов целевой функции отрицательны или равны нулю, то имеющееся базисное решение оптимальное. Если хотя бы одна оценка коэффициента целевой функции больше нуля, то имеющееся базисное решение не является оптимальным и должно быть улучшено.

Четвертый этап . Переход к новому базисному решению. Очевидно, что в оптимальный план должна быть введена такая переменная, которая в наибольшей степени увеличивает целевую функцию. При решении задач на максимум прибыли в оптимальный план вводится продукция, производство которой наиболее выгодно. Это определяется по максимальному положительному значению оценки коэффициента целевой функции.

Столбец симплексной таблицы с этим номером на данной итерации называется *генеральным столбцом.*

Далее, если хотя бы один элемент генерального столбца a_{ij0} строго положителен, то отыскивается генеральная строка (в противном случае задача не имеет оптимального решения).

Для отыскания генеральной строки все свободные члены (ресурсы) делятся на соответствующие элементы генерального столбца (норма расхода ресурса на единицу изделия). Из полученных результатов выбирается наименьший. Соответствующая ему строка на данной итерации называется *генеральной* . Она соответствует ресурсу, который лимитирует производство на данной итерации.

Элемент симплексной таблицы, находящийся на пересечении генеральных столбца и строки, называется *генеральным элементом* .

Затем все элементы генеральной строки (включая свободный член) делятся на генеральный элемент. В результате этой операции генеральный элемент становится равным единице. Далее необходимо, чтобы все другие элементы генерального столбца стали бы равны нулю, т.е. генеральный столбец должен стать единичным. Все строки (кроме генеральной) преобразуются следующим образом. Полученные элементы новой строки умножаются на соответствующий элемент генерального столбца и полученное произведение вычитается из элементов старой строки.

Значения новых базисных переменных получим в соответствующих ячейках столбца свободных членов.

Пятый этап . Полученное базисное решение проверяется на оптимальность (см. третий этап). Если оно оптимально, то вычисления прекращаются. В противном случае необходимо найти новое базисное решение (четвертый этап) и т.д.

Пример решения ОЗЛП симплексным методом

Пусть необходимо найти оптимальный план производства двух видов продукции x_1 и x_2 (табл. 3.2).

Таблица 3.2 – **Исходные данные примера**

Вид продукции	Норма расхода ресурса на единицу прибыли		Прибыль на единицу изделия
	A	*B*	
1	5	8	7
2	2	4	3
Объем ресурса	20	36	

1. Построим ОМ

$$F(x) = 7x_1 + 3x_2 \rightarrow \max;$$

$5x_1 + 2x_2 \le 20$ ограничение по ресурсу *A* ;

$8x_1 + 4x_2 \le 36$ ограничение по ресурсу B .

2. Преобразуем задачу в приведенную каноническую форму. Для этого достаточно ввести дополнительные переменные x_3 и x_4 . В результате неравенства преобразуются в строгие равенства:

$$5x_1 + 2x_2 + x_3 = 20,$$
$$8x_1 + 4x_2 + x_4 = 36.$$

Построим исходную симплексную таблицу и найдем начальное базисное решение. Им будет пара значений дополнительных переменных, которым соответствует единичная подматрица

$x_3 = 20$ и $x_4 = 36.$

Базисные переменные	Свободные члены (план)	x_1	x_2	x_3	x_4
x_3	20	5	2	1	0
x_4	36	8	4	0	1
$F_j - C_j$		7	3	0	0

1-я итерация . Находим генеральный столбец и генеральную строку:

$$\max(7,\ 3) = 7,$$
$$\min\left(\frac{20}{5};\ \frac{36}{8}\right) = \frac{20}{5}.$$

Генеральный элемент равняется 5.

Базисные переменные	Свободные члены (план)	x_1	x_2	x_3	x_4
x_1	4	1	0,4	0,2	0
x_4	4	0	0,8	-1,6	1
$F_j – C_j$	28	0	0,2	-1,4	0

2-я итерация . Найденное базисное решение не является оптимальным, так как строка оценок (F_j - C_j) содержит один положительный элемент. Находим генеральный столбец и генеральную строку:

$$\max\ (0,\ 0{,}3,\ -1{,}4,\ 0) = 0{,}2,$$
$$\min\left(\frac{4}{0{,}4};\ \frac{4}{0{,}8}\right) = \frac{4}{0{,}8}$$

Базисные переменные	Свободные члены (план)	x_1	x_2	x_3	x_4
x_1	2	1	0	1 0,5	- 0

x_2	5	0	1	-2	1,25
$F_j - C_j$	29	0	0	-1	-0,25

Найденное решение оптимально, так как все специальные оценки целевой функции (F_j - C_j) равны нулю или отрицательны. $F(x) = 29$; $x_1 = 2$; $x_2 = 5$.

Двойственная задача ЛП

Двойственная задача ЛП может быть сформулирована следующим образом:

найти переменные y_i ($i = 1, 2, ..., m$), при которых целевая функция была бы минимальной

$$Z(y) = \sum_{i=1}^{m} b_i y_i \to \min,$$

не нарушая ограничений

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j \quad (j = 1, 2, \ldots, n)$$

$$y_i \geq 0 \qquad (i = 1, 2, \ldots, m).$$

Данная задача называется двойственной (симметричной) по отношению к прямой задаче, сформулированной во втором параграфе данной главы. Однако правильным будет и обратное утверждение, так как обе задачи равноправны. Компоненты решения двойственной задачи называются *объективно обусловленными оценками* .

Прямая и обратная задачи ЛП связаны между собой теоремами двойственности.

Первая теорема двойственности . Если обе задачи имеют допустимые решения, то они имеют и оптимальное решение, причем значение целевых функций у них будет одинаково

$F(x) = Z(y)$ или $\max \sum_{j=1}^{n} c_j x_j = \min \sum_{i=1}^{m} b_i y_i .$

Если же хотя бы одна из задач не имеет допустимого решения, то ни одна из них не имеет оптимального решения.

Вторая теорема двойственности (теорема о дополняющей нежесткости). Для того чтобы векторы $\overline{X} = (x_1, x_2, \ldots, x_n)$ и $\overline{Y} = (y_1, y_2, \ldots, y_m)$ были оптимальными решениями соответственно прямой и двойственной задачи, необходимо и достаточно, чтобы выполнялись следующие условия:

$$y_i \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) = 0 \quad (i = 1, 2, .. \qquad (1)$$

$$x_j \left(\sum_{i=1}^{m} a_{ij} y_j - c_j \right) = 0 \quad (j = 1, 2, .. \qquad (2)$$

Следствие 1 . Пусть оптимальное значение некоторой переменной двойственной задачи строго положительно

$$y_i > 0.$$

Тогда из условия (1) получим

$$b_i - \sum_{j=1}^{n} a_{ij} x_j = 0$$

или

$$\sum_{j=1}^{n} a_{ij} x_j = b_i.$$

Экономический смысл данных выражений можно интерпретировать в следующей редакции. Если объективно обусловленная оценка некоторого ресурса больше нуля (строго положительна), то этот ресурс полностью (без остатка) расходуется в процессе выполнения оптимального плана.

Следствие 2 . Пусть для оптимального значения некоторой переменной x_i прямой задачи выполняется условие строгого неравенства

$$\sum_{j=1}^{n} a_{ij} x_j < b_i.$$

Тогда, основываясь на том же первом условии (1), можно заключить, что $y_i = 0$.

Экономически это означает, что если в оптимальном плане какой-то ресурс используется не полностью, то его объективно обусловленная оценка обязательно равна нулю.

Решение двойственной задачи ЛП

Ранее (6) мы рассматривали прямую задачу ЛП:

$$F(x) = 180x_1 + 20x_2 \to \max$$
$$0{,}5x_1 + 0{,}04x_2 \le 200$$
$$12x_1 + 0{,}6x_2 \le 1800$$
$$x_1 \ge 80$$
$$x_1 x_2 \ge 0.$$

В системе неравенств должны быть однотипные знаки меньше или равно. Поэтому неравенство x_1 Ё 80 умножим на - 1 и поменяем знак неравенства на противоположный.

$$Z(y) = 200y_1 + 1800y_2 - 80y_3 \to \min$$
$$0{,}5y_1 + 12y_2 - y_3 \ge 180$$
$$0{,}04y_1 + 0{,}6y_2 \ge 20$$
$$y_1, y_2 \ge 0.$$

Ограничение на целочисленность переменных здесь не требуется.

Решение прямой задачи дало следующие результаты:

$$x_1 = 80; \quad x_2 = 1400; \quad F(x) = 42400.$$

В результате решения двойственной задачи получим

$$y_1 = 0; \quad y_2 = 33{,}3; \quad y_3 = 220; \quad Z(y) = 42400.$$

Объективно обусловленная оценка $y_1 = 0$ указывает на то, что у нас избыток древесины: $y_2 = 33{,}3$, т.е. больше нуля. Значит этот ресурс (труд) полностью используется в оптимальном плане. Значение целевой функции $Z(y)$ равно $F(x) = 42400$. Это свидетельствует о том, что найденное решение оптимально.

ЛЕКЦИЯ 9

Свойства объективно обусловленных оценок и их анализ

Анализ задачи с использованием объективно обусловленных оценок показывает, что первый ресурс (древесина) расходуется не полностью. Можно убедиться, что для найденного оптимального плана достаточно 96 куб. м древесины, а 104 куб. м избыточны. Изменение ограничения по древесине с 200 до 96 куб. м не повлияет на оптимальный план. Следовательно, объективно обусловленные оценки являются устойчивыми в некоторых пределах изменения исходных условий задачи.

Объективно обусловленные оценки выступают как мера дефицитности ресурсов. Древесина, объективно обусловленная оценка которой у нас равна нулю, не дефицитна, а трудовые ресурсы с объективно обусловленной оценкой, равной в нашей задаче 33,3, дефицитны и используются полностью.

Объективно обусловленные оценки выступают как мера влияния ограничений на целевую функцию при приращении данного ресурса на единицу. Так, например, уменьшение задания по производству столов с 80 до 79 увеличивает целевую функцию на 220 руб., а увеличение трудовых ресурсов с 1800 до 1801 чел.-ч увеличивает целевую функцию (если снять условие целочисленности) на 33,3 руб.

Объективно обусловленные оценки выступают как меры взаимозаменяемости резервов (ограничений). Так, например, если увеличить задание по производству столов на единицу, то для того чтобы целевая функция осталась прежней, нужно добавить 6,6 чел.-ч (220/33,3). В этом случае x_1 будет равен 81, $x_2 = 1391$, а значение целевой функции составит 42400.

Следует иметь в виду, что при существенном изменении исходных условий задачи обычно получается уже другая система оценок. Следовательно, объективно обусловленные оценки обладают свойством конкретности, так как определяются совокупностью условий данной задачи. Для другой задачи и других условий их значения будут совершенно иными.

Разработка производственной программы фирмы

Разработку производственной программы фирмы можно выполнить, используя метод производственных функций, который дает возможность получить описание эффективного технологического множества в терминах связи между входом (затратами производственных факторов) и выходом (результатами производственной деятельности) объекта. Однако во многих случаях представляется полезным иметь сведения о структуре того или иного эффективного технологического способа, т.е., если возможно, понять, комбинацией каких более простых процессов он является. В качестве основной модели, позволяющей дать описание не только связей входвыход, но и параметров внутреннего состояния производственной системы, применяется, как правило, линейная модель производства. Эта модель строится исходя из предположения, что всякий допустимый технологический способ (в том числе эффективный) может быть представлен в виде линейной комбинации с положительными коэффициентами так называемых базисных производственных способов. Наиболее распространенное истолкование такого представления основывается на том, что производственное предприятие рассматривается как объединение параллельно работающих взаимосвязанных цехов, участков и т.п., причем в каждом из них используется некая определенная технология. Тогда описание такой технологии (т.е. производство продукции и соответствующие затраты) можно считать базисным производственным способом. Время работы цеха по этой технологии обычно называют интенсивностью базисного способа. В рассматриваемом примере все основные показатели затрат и выпуска для производственного объекта в целом оказываются суммой подобных же величин по отдельным цехам.

Говоря более точно, линейные модели производства основаны на следующих гипотезах о свойствах технологического множества V :

1. Технологическое множество V является конусом в R_+^{m+n}, т.е. для любогоа > 0 имеет место следующий факт: если v Î V, то аа v Î V.

Последнее означает, что вместе с допустимым технологическим способом

$$v = (x, y)$$

допустимой является также технология, для которой затраты и выпуск увеличены (уменьшены при а < 1) в одной и той же пропорции.

2. Имеет место аддитивность технологических способов, т.е. если $v^{(1)}$ Î V и $v^{(2)}$ Î V, то($v^{(1)} + v^{(2)}$) Î V.

Данная предпосылка выражает упомянутое выше свойство параллельности технологий: если допустимым является выпуск продукции в размере $y^{(1)}$ при затратах $x^{(1)}$, а также выпуск $y^{(2)}$ при затратах $x^{(2)}$, то возможен выпуск ($y^{(1)} + y^{(2)}$) при затратах ресурсов ($x^{(1)} + x^{(2)}$).

3. Существует конечный выбор базисных технологических способов (v_1, ..., v_n), при помощи которых могут быть представлены все технологические способы v Î V ,т.е. для всякого v найдутся такие неотрицательные числа z_1, ..., z_n, называемые интенсивностями базисных способов, для которых справедливо соотношение

$$v = \sum_{j=1}^{n} v_j z_j .$$

Выполнение указанных трех предположений означает, что множество V является выпуклым многогранным конусом. В конкретной задаче планирования производства обычно используется подмножество множества V , определяемое системой ограничений двух видов. Условия первого вида предназначены для характеристики ограниченности запасов ресурсов, которые необходимы для осуществления производственного процесса. Группу условий второго вида образуют плановые задания по отдельным видам продукции, комплектам или стоимостным показателям. Ограничивающие условия обоих видов записываются обычно в терминах интенсивностей базисных технологий. При этом произвольная базисная технология (с номером j) имеет вид

$$v_j = (a_{1j}, \dots, a_{ij}, \dots, a_{mj}),$$

где m общее число компонент (продуктов и ресурсов). В некоторых случаях различные виды компонент отличаются знаками (например, положительные числа относятся к продуктам, а отрицательные представляют объемы затрачиваемых ресурсов), но в большинстве случаев смысл каждой компоненты специально оговаривается.

Условия ресурсного характера общего типа имеют вид

$$\sum_{j=1}^{n} a_{ij} z_j \le b_i \quad (i = 1, \ldots, mm),$$

где *mm*   число общих ограничений меньше или равно.

Кроме того, возможны локальные ограничения $z_j \le \hat{b}_j$, которые отражают ограниченность производственной мощности по j -тому базисному способу. Ограничения, отражающие необходимость выполнения планового задания по некоторому продукту, вообще говоря, имеют вид

$$\sum_{j=1}^{n} a_{ij} z_j \ge b_i \quad (i = mm + 1, \ldots, m).$$

Таким образом, линейная модель производства позволяет представить взаимозаменяемость и комбинирование фиксированных технологических способов, считающихся основными, в некоторых пределах, обозначенных ресурсами и плановыми ограничениями. Рассматривая производственный объект как кибернетическую систему, можно сказать, что характеристики основных производственных способов (их коэффициенты затрат и выпуска) являются основными параметрами системы, а обусловленные плановым решением значения интенсивностей (z_j) представляют собой описание состояния системы. Плановое решение определяется на множестве возможных состояний Z при помощи некоторого правила выбора наилучшего состояния. По существу формирование такого правила выбора отражает представление об интересах производственного объекта и связанного с этим выделением множества эффективных технологических способов. Указанное правило выбора может быть сформулировано при помощи некоторого формализованного критерия (скалярная оптимизация) или набора ряда оценочных показателей (многокритериальная оптимизация). Оно может быть по существу неформальным правилом, если окончательный выбор решения зависит от условий среды, окружающей данный производственный объект, от предыдущих решений и, наконец, от субъективных оценок руководителей.

В рассматриваемой нами модели очень часто используется скалярное правило выбора наилучшего состояния, базирующееся на нахождении оптимального значения некоторой целевой функции. Эта функция обычно имеет смысл выпуска наибольшего возможного количества продукции либо в стоимостном выражении, либо в заданном комплекте. Зачастую применяются целевые функции, выражающие максимизацию прибыли или минимизацию затрат.

Если целевая функция является линейной относительно искомых интенсивностей основных технологических способов, то возникает линейная оптимизационная модель производства, которая, с одной стороны, благодаря своей относительной простоте доступна для достаточно детального математического анализа, а с другой стороны, имеет самостоятельное значение как инструмент принятия и поддержки решения в достаточно несложных ситуациях, а также как элемент болеесложных конструкций решающих систем.

Обратимся далее к одной из линейных оптимизационных моделей, которая представляет собой способ решения задачи оптимизации производственной программы в условиях фиксированных цен на продукцию для предприятия, хозяйственный интерес которого состоит в получении наибольшей выручки (товарной продукции). Пусть в процессе производства используется m видов ресурсов, запасы которых также фиксированы. Тогда вектор затратвыпуска, описывающий основной технологический способ, запишется следующим образом:

$$v_j = (a_{1j}, \ldots, a_{ij}, \ldots, a_{mj}; c_j).$$

Здесь неотрицательные числа a_{ij} (i = 1, ..., m) представляют собой нормы затрат производственных факторов, а c_j стоимость товарной продукции, полученную предприятием при реализации технологии v_j с единичной интенсивностью. Если имеющиеся запасы ресурсов составляют b_i (i = 1, ..., m), то оптимизационная модель для разработки программы предприятия имеет вид

$$\sum_{j=1}^{n} c_j z_j \to \max$$

$$\sum_{j=1}^{n} a_{ij} z_j \le b_i \quad (i = 1, \ldots, m)$$

$$z_j \ge 0 \quad (j = 1, \ldots, n).$$

Предположим, что полученная задача линейного программирования имеет единственное решение

$$\hat{z} = (\hat{z}_1, \ldots, \hat{z}_n).$$

Тогда технологический способ

$$\hat{v} = \sum_{j=1}^{n} v_j \hat{z}_j$$

есть эффективный способ, позволяющий выполнить оптимальную производственную программу и получить максимальный объем товарной продукции:

$$\hat{c} = \sum_{j=1}^{n} c_j \hat{z}_j.$$

Допустим далее, что запасы используемых ресурсов b_i (i = 1, ..., m) могут изменяться в некоторых пределах. Если при этом не меняется оптимальный базис представленной выше задачи линейного программирования, то в силу теоремы двойственности

$$\hat{c} = \sum_{i=1}^{m} b_i \hat{u}_i,$$

где $\hat{u}_i \ (i = 1, \ldots, m)$ оптимальные оценки, которые являются компонентами решения двойственности задачи. Легко видеть, что полученное соотношение можно рассматривать как определение скалярной производственной функции, причем в качестве дифференциальной продуктивности ресурса с номером i выступает оптимальная оценка $\hat{u}_i$. Приведенное представление распространяется и на более общую ситуацию, когда в пределах изменения запасов ресурсов b_i происходит смена оптимального базиса. Тогда оптимальные оценки следует считать, вообще говоря, многозначными функциями от запасов ресурсов, а производственную функцию

$$\hat{c} = f(b_1, \ldots, b_m)$$

определить в каждой точке как минимум из конечного числа линейных функций

$$\sum_{i=1}^{n} u_i^{(1)} b_i; \quad \sum_{i=1}^{n} u_i^{(2)} b_i; \quad \sum_{i=1}^{n} u_i^{(s)} b_i,$$

где $u^{(1)}, \ldots, u^{(s)}$ суть вершины многогранника, являющегося множеством допустимых решений двойственной задачи. Такого рода производственная функция является кусочно-линейной, она имеет непрерывные производные во всех точках области определения, кроме точек, принадлежащих гиперплоскостям смены оптимального базиса. Следует заметить, что при увеличении запаса некоторого ресурса (b_i) при неизменных количествах прочих ресурсов, соответствующая оптимальная оценка ($\hat{u}_i$) уменьшается. При достаточно больших значениях некоторое количество ресурса становится излишним, а оценка $\hat{u}_i$ становится равной нулю. Это означает, что в рассматриваемом случае используемые ресурсы являются не взаимозаменяемыми, а взаимодополняющими.

В частном случае, когда в задаче присутствуют лишь локальные ограничения

$$z_0 \to \max; \quad a_i z_0 \le b_i \quad (i = 1, \ldots, m),$$

где через z_0 обозначен объем выпускаемой продукции, через a_i ресурсоемкость единицы продукции по i -тому ресурсу, решение можно представить в виде

$$z_0 = \min_i \frac{b_i}{a_i}.$$

Это выражение известно так же, как производственная функция для взаимодополняющих (комплементарных) производственных факторов, и широко применяется в сложных моделях производства.

Величина оптимальной оценки $\hat{u}_i$ производственного фактора имеет троякий смысл: измерителя дополнительного вклада единицы i -того фактора в повышении значения целевой функции; коэффициента, характеризующего относительную дефицитность этого фактора по сравнению с другими, а также выражает верхний предел цены, которую согласен заплатить производитель за дополнительную единицу ресурса.

Для того, чтобы подробнее остановиться на стоимостной интерпретации оптимальных оценок, воспользуемся двумя группами соотношений между оптимальными решениями прямой и двойственной задач линейного программирования (ЛП).

Прямой задаче ЛП

$$\sum_{j=1}^{n} c_j z_j \to \max$$

$$\sum_{j=1}^{n} a_{ij} z_j \le b_i \quad (i = 1, \ldots, m)$$

$$z_j \ge 0 \quad (j = 1, \ldots, n)$$

соответствует двойственная задача:

$$\sum_{i=1}^{m} b_i u_i \to \min$$

$$\sum_{i=1}^{m} a_{ij} u_i \ge c_j \quad (j = 1, \ldots, n)$$

$$u_i \ge 0 \quad (i = 1, \ldots, m).$$

Пусть $\hat{z} = (\hat{z}_1, \dots, \hat{z}_n); \hat{u} = (\hat{u}_1, \dots, \hat{u}_m)$ оптимальные решения соответственно прямой и двойственной задач. Тогда справедливы следующие соотношения:

$$1)\ \hat{u}_i \left(b_i - \sum_{i=1}^{ò} a_{ij} \hat{z}_j \right) = 0 \quad (i = 1, \dots, m),$$

$$2)\ \hat{z}_j \left(\sum_{j=1}^{n} a_{ij} \hat{u}_i - c_j \right) = 0 \quad (j = 1, \dots, n).$$

Рассматривая условия первой группы, можно установить, что при $\hat{u}_i$, т.е. в том случае, когда оптимальная оценка *i* -того ресурса строго положительна, необходимо выполняется равенство

$$\sum_{j=1}^{n} a_{ij} \hat{z}_j = b_i.$$

Таким образом, оптимальный (эффективный) технологический способ использует данный ресурс полностью. Отсюда можно заключить, что этот производственный фактор является ограничивающим, и дополнительное поступление некоторого количества ресурса приведет к увеличению целевой функции. Если такой функцией является выпуск товарной продукции (или прибыль), то увеличение ресурса на единицу сулит получение дополнительной продукции на сумму в $\hat{u}_i$ денежных единиц. Следовательно, производство может приобрести указанную единицу по цене

$$p_i \le \hat{u}_i$$

и ожидать получение дополнительной прибыли. Если же *i* -тый ресурс используется не полностью в процессе производства, то его оптимальная оценка обязательно равна нулю

$$\hat{u}_i = 0.$$

В самом деле, малое дополнительное увеличение такого ресурса не приведет к увеличению выпуска, поскольку в данный момент имеется его излишек. В связи с этим затраты на приобретение дополнительного количества не являются оправданными. Здесь нужно отметить некоторые недостатки этого рассуждения. Во-первых, приведенная оценка сохраняет свой смысл лишь для достаточно малых количеств дополнительного ресурса, которые не приводят к изменению оптимального базиса задачи ЛП, а следовательно, и к изменению системы оптимальных оценок. Вполне возможно, что существенное дополнительное количество ресурса сделает более выгодным и приведет к включению в оптимальный план другой технологический способ, который связан с эффективным использованием упомянутого ресурса, что сразу повлечет изменение его оптимальной оценки в сторону увеличения.

Во-вторых, приведенное рассуждение основывается на статической модели производства, в которой весь процесс происходит в сравнительно короткий период времени. Поэтому оптимальная оценка характеризует ресурс лишь с точки зрения эффективности его использования лишь в этом кратком периоде, не затрагивая проблемы формирования запасов и движения ресурсов во времени.

Тем не менее, приведенное стоимостное истолкование оптимальных оценок служит надежной базой для концепции теневых или расчетных цен, лежащих в основе расчета между подразделениями внутри предприятия или производственного объединения.

Условия второй группы могут быть также использованы для составления расчетных цен на готовую продукцию. Предположим, что базисный технологический способ с номером j имеет конечным результатом выпуск определенного вида продукции j (в натуральном выражении, равном q_j). Тогда, если в оптимальном решении прямой задачи $\hat{z}_j > 0,$ то технологический способ j рекомендован к осуществлению и в оптимальном плане предлагается производить продукцию с номером j .Из соответствующего соотношения второй группы вытекает, что при $\hat{z}_j > 0$ имеет место равенство

$$c_j = \sum_{i=1}^{m} a_{ij}\hat{u}_i .$$

Таким образом, при фиксированной цене изделия c_j , оптимальные оценки $\hat{u}_i$ выражают разложение цены по отдельным элементам совокупных затрат с учетом ограниченности ресурсов, представленных в рассматриваемой модели.

Конечно, для создания более полной картины распределения затрат по элементам необходимо использовать более детальные модели производства. Если в число ограничений включаются трудовые ресурсы, многие виды материальных ресурсов, лимиты капитальных вложений, то цена продукции представляется как сумма расчетной заработной платы, материальных затрат и затрат на пополнение и создание новых производственных фондов.

Некоторым недостатком приведенного представления является отсутствие в нем выражений для прибавочного продукта и полезного дополнительного эффекта у потребителя, которые присутствуют в большинстве концепций формирования цены.

В связи с этим данное соотношение следует рассматривать лишь в контексте приведенной модели производства, как средство анализа внутренних свойств исследуемого производственного объекта. С этой точки зрения соотношения второй группы и вытекающие из них следствия удобно использовать для оценки экономической эффективности новых предлагаемых базисных производственных способов. В самом деле, из соотношений второй группы вытекает, что если для некоторого базисного технологического способа j выполняется неравенство

$$\sum_{i=1}^{m} a_{ij}\hat{u}_i > c_j ,$$

которое означает, что затраты ресурсов, исчисленные с помощью оптимальных оценок, как внутренних цен, превосходят ожидаемую выручку от продажи j -того изделия, то необходимо $\hat{z}_j = 0.$ Последнее означает, что такой нерентабельный базисный способ не рекомендуется для использования в оптимальном плане производства.

Пусть теперь предлагается новый технологический способ вида

$$(a_{1l}, \ldots, a_{il}, \ldots, a_{ml}; c_l).$$

Если оказывается, что

$$c_l \le \sum_{i=1}^{m} a_{il}\hat{u}_i ,$$

то предлагаемый способ не следует включать в число базисных, поскольку он не дает дополнительной выгоды по сравнению с уже используемыми.

Если же имеет место обратное неравенство

$$c_l > \sum_{i=1}^{m} a_{il}\hat{u}_i,$$

то следует ввести предлагаемую технологию в состав базисных и найти оптимальное решение новой, расширенной, задачи. Если в результате расчета получим интенсивность $\hat{z}_l > 0$, то новая технология станет составной частью оптимального способа.

ЛЕКЦИЯ 10

Моделирование систем массового обслуживания

Общие понятия систем массового обслуживания

Системы массового обслуживания — это такие системы, в которые в случайные моменты времени поступают заявки на обслуживание, при этом поступившие заявки обслуживаются с помощью имеющихся в распоряжении системы каналов обслуживания.

Поступив в обслуживающую систему, требование присоединяется к очереди других (ранее поступивших) требований. Канал обслуживания выбирает требование из находящихся в очереди, с тем, чтобы приступить к его обслуживанию. После завершения процедуры обслуживания очередного требования канал обслуживания приступает к обслуживанию следующего требования, если таковое имеется в блоке ожидания.

Цикл функционирования системы массового обслуживания подобного рода повторяется многократно в течение всего периода работы обслуживающей системы. При этом предполагается, что переход системы на обслуживание очередного требования после завершения обслуживания предыдущего требования происходит мгновенно, в случайные моменты времени.

Примерами систем массового обслуживания могут служить:

–посты технического обслуживания автомобилей;

–посты ремонта автомобилей;

–персональные компьютеры, обслуживающие поступающие заявки или требования на решение тех или иных задач;

–станции технического обслуживания автомобилей;

–аудиторские фирмы;

–отделы налоговых инспекций, занимающиеся приемкой и проверкой текущей отчетности предприятий;

–телефонные станции и т. д.

Основными компонентами системы массового обслуживания любого вида являются:

- входной поток поступающих требований или заявок на обслуживание;
- дисциплина очереди;
- механизм обслуживания.

Входной поток требований. Для описания входного потока требуется задать вероятностный закон, определяющий последовательность моментов поступления требований на обслуживание и указать количество таких требований в каждом очередном поступлении. При этом, как правило, оперируют понятием «вероятностное распределение моментов поступления требований». Здесь могут поступать как единичные, так и групповые требования (требования поступают группами в систему). В последнем случае обычно речь идет о системе обслуживания с параллельно-групповым обслуживанием.

Дисциплина очереди — это важный компонент системы массового обслуживания, он определяет принцип, в соответствии с которым поступающие на вход обслуживающей системы требования подключаются из очереди к процедуре обслуживания. Чаще всего используются дисциплины очереди, определяемые следующими правилами:

– первым пришел — первый обслуживаешься;

– пришел последним — обслуживаешься первым;

– случайный отбор заявок;

– отбор заявок по критерию приоритетности;

– ограничение времени ожидания момента наступления обслуживания (имеет место очередь с ограниченным временем ожидания обслуживания, что ассоциируется с понятием «допустимая длина очереди»).

Механизм обслуживания определяется характеристиками самой процедуры обслуживания и структурой обслуживающей системы. К характеристикам процедуры обслуживания относятся: продолжительность процедуры обслуживания и количество требований, удовлетворяемых в результате выполнения каждой такой процедуры. Для аналитического описания характеристик процедуры обслуживания оперируют понятием «вероятностное распределение времени обслуживания требований».

Следует отметить, что время обслуживания заявки зависит от характера самой заявки или требований клиента и от состояния и возможностей обслуживающей системы. В ряде случаев приходится также учитывать вероятность выхода обслуживающего прибора по истечении некоторого ограниченного интервала времени.

Структура обслуживающей системы определяется количеством и взаимным расположением каналов обслуживания (механизмов, приборов и т. п.). Прежде всего, следует подчеркнуть, что система обслуживания может иметь не один канал обслуживания, а несколько; система такого рода способна обслуживать одновременно несколько требований. В этом случае все каналы обслуживания предлагают одни и те же услуги, и, следовательно, можно утверждать, что имеет место параллельное обслуживание.

Система обслуживания может состоять из нескольких разнотипных каналов обслуживания, через которые должно пройти каждое обслуживаемое требование, т. е. в обслуживающей системе процедуры обслуживания требований реализуются последовательно. Механизм обслуживания определяет характеристики выходящего (обслуженного) потока требований.

Рассмотрев основные компоненты систем обслуживания, можно констатировать, что *функциональные возможности любой системы массового обслуживания определяются следующими основными факторами:*

– вероятностным распределением моментов поступлений заявок на обслуживание (единичных или групповых);

– вероятностным распределением времени продолжительности обслуживания;

– конфигурацией обслуживающей системы (параллельное, последовательное или параллельно-последовательное обслуживание);

– количеством и производительностью обслуживающих каналов;

–дисциплиной очереди;

–мощностью источника требований.

В качестве основных критериев эффективности функционирования систем массового обслуживания в зависимости от характера решаемой задачи могут выступать:

–вероятность немедленного обслуживания поступившей заявки;

–вероятность отказа в обслуживании поступившей заявки;

–относительная и абсолютная пропускная способность системы;

–средний процент заявок, получивших отказ в обслуживании;

–среднее время ожидания в очереди;

–средняя длина очереди;

–средний доход от функционирования системы в единицу времени и т.п.

Предметом теории массового обслуживания является установление зависимости между факторами, определяющими функциональные возможности системы массового обслуживания, и эффективностью ее функционирования. В большинстве случаев все параметры, описывающие системы массового обслуживания, являются случайными величинами или функциями, поэтому эти системы относятся к стохастическим системам.

Случайный характер потока заявок (требований), а также, в общем случае, и длительности обслуживания приводит к тому, что в системе массового обслуживания происходит случайный процесс.

Независимо от характера процесса, протекающего в системе массового обслуживания, различают два основных вида СМО:

–системы с отказами, в которых заявка, поступившая в систему в момент, когда все каналы заняты, получает отказ и сразу же покидает очередь;

–системы с ожиданием (очередью), в которых заявка, поступившая в момент, когда все каналы обслуживания заняты, становится в очередь и ждет, пока не освободится один из каналов.

Системы массового обслуживания с ожиданием делятся на системы с ограниченным ожиданием и системы с неограниченным ожиданием.

В системах с ***ограниченным ожиданием*** может ограничиваться:

–длина очереди;

–время пребывания в очереди.

В системах с ***неограниченным ожиданием*** заявка, стоящая в очереди, ждет обслуживание неограниченно долго, т.е. пока не подойдет очередь.

Все системы массового обслуживания различают по числу каналов обслуживания:

–одноканальные системы;

–многоканальные системы.

Приведенная классификация СМО является условной. На практике чаще всего системы массового обслуживания выступают в качестве смешанных систем. Например, заявки ожидают начала обслуживания до определенного момента, после чего система начинает работать как система с отказами.

Одноканальная модель с пуассоновским входным потоком с экспоненциальным распределением длительности обслуживания

Простейшей одноканальной моделью с вероятностными входным потоком и процедурой обслуживания является модель, характеризуемая показательным распределением как длительностей интервалов между поступлениями требований, так и длительностей обслуживания. При этом плотность распределения длительностей интервалов между поступлениями требований имеет вид

$$f_1(t) = \lambda \cdot e^{-\lambda t}, \qquad (1)$$

где λ. — интенсивность поступления заявок в систему.

Плотность распределения длительностей обслуживания:

$$f_2(t) = \mu \cdot e^{-\mu t}, \qquad (2)$$

где μ - интенсивность обслуживания.

Потоки заявок и обслуживании простейшие.

Пусть система работает с ***отказами.*** Необходимо определить абсолютную и относительную пропускную способность системы.

Представим данную систему массового обслуживания в виде графа (рис. 4.1), у которого имеются два состояния:

S_0 — канал свободен (ожидание);

S_1— канал занят (идет обслуживание заявки).

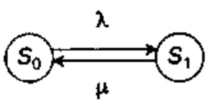


Рисунок 4.1 – Граф состояний одноканальной СМО с отказами

Обозначим вероятности состояний:

$P_0(t)$ — вероятность состояния «канал свободен»;

$P_1(t)$ — вероятность состояния «канал занят».

По размеченному графу состояний (рис. 1) составим систему дифференциальных уравнений Колмогорова для вероятностей стояний:

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda \cdot P_0(t) + \mu \cdot P_1(t) \\ \dfrac{dP_1(t)}{dt} = -\mu \cdot P_1(t) + \lambda \cdot P_0(t). \end{cases} \qquad (3)$$

Система линейных дифференциальных уравнений имеет решение с учетом нормировочного условия $P_0(t) + P_1(t) = 1$. Решение данной системы называется неустановившимся, поскольку оно непосредственно зависит от t и выглядит следующим образом:

$$P_0(t) = \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu};$$

$$P_1(t) = 1 - P_0(t).$$

(4) (5)

Нетрудно убедиться, что для одноканальной СМО с отказами вероятность $P_0(t)$ есть не что иное, как относительная пропускная способность системы q.

Действительно, P_0 — вероятность того, что в момент t канал свободен и заявка, пришедшая к моменту t, будет обслужена, а следовательно, для данного момента времени t среднее отношение числа обслуженных заявок к числу поступивших также равно $P_0(t)$, т. е.

$$q = P_0(t). \quad (6)$$

По истечении большого интервала времени (при $t \to \infty$) достигается стационарный (установившийся) режим:

$$q = P_0 = \frac{\mu}{\mu+\lambda}. \quad (7)$$

Зная относительную пропускную способность, легко найти абсолютную. Абсолютная пропускная способность (A) — среднее число заявок, которое может обслужить система массового обслуживания в единицу времени:

$$A = \lambda \cdot q = \frac{\lambda \cdot \mu}{\lambda+\mu}. \quad (8)$$

Вероятность отказа в обслуживании заявки будет равна вероятности состояния «канал занят»:

$$P_{отк} = P_1 = 1 - P_0 = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}. \quad (9)$$

Данная величина $P_{отк}$ может быть интерпретирована как средняя доля не обслуженных заявок среди поданных.

Пример 1. Пусть одноканальная СМО с отказами представляет собой один пост ежедневного обслуживания (ЕО) для мойки автомобилей. Заявка — автомобиль, прибывший в момент, когда пост занят, — получает отказ в обслуживании. Интенсивность потока автомобилей λ = 1,0 (автомобиль в час). Средняя продолжительность обслуживания — 1,8 часа. Поток автомобилей и поток обслуживании являются простейшими.

Требуется определить в установившемся режиме предельные значения:

– относительной пропускной способности q;

– абсолютной пропускной способности А;

– вероятности отказа Ротк;

Сравните фактическую пропускную способность СМО с номинальной, которая была бы, если бы каждый автомобиль обслуживался точно 1,8 часа и автомобили следовали один за другим без перерыва.

Решение

1. Определим интенсивность потока обслуживания:

$$\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{1,8} = 0,555.$$

2. Вычислим относительную пропускную способность:

$$q = \frac{\mu}{\lambda + \mu} = \frac{0,555}{1 + 0,555} = 0,356.$$

Величина *q* означает, что в установившемся режиме система будет обслуживать примерно 35% прибывающих на пост ЕО автомобилей.

3. Абсолютную пропускную способность определим по формуле:

$$A = \lambda \cdot q = 1 \cdot 0,356 = 0,356.$$

Это означает, что система (пост ЕО) способна осуществить в среднем 0,356 обслуживания автомобилей в час. 3. Вероятность отказа:

$$P_{отк} = 1 - q = 1 - 0,356 = 0,644.$$

Это означает, что около 65% прибывших автомобилей на пост ЕО получат отказ в обслуживании.

4. Определим номинальную пропускную способность системы:

$$A_{ном} = \frac{1}{\bar{t}_{обсл}} = \frac{1}{1,8} = 0,555$$

(автомобилей в час).

Оказывается, что $A_{ном}$ в 1,5 раза больше, чем фактическая пропускная способность, вычисленная с учетом случайного характера потока заявок и времени обслуживания.

Одноканальная СМО с ожиданием

Система массового обслуживания имеет один канал. Входящий поток заявок на обслуживание — простейший поток с интенсивностью λ,. Интенсивность потока обслуживания равна μ, (т. е. в среднем непрерывно занятый канал будет выдавать μ обслуженных заявок). Длительность обслуживания — случайная величина, подчиненная показательному закону распределения. Поток обслуживании является простейшим пуассоновским потоком событий. Заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания.

Предположим, что независимо от того, сколько требований поступает на вход обслуживающей системы, данная система (очередь + обслуживаемые клиенты) не может вместить более N -требований (заявок), т. е. клиенты, не попавшие в ожидание, вынуждены обслуживаться в другом месте. Наконец, источник, порождающий заявки на обслуживание, имеет неограниченную (бесконечно большую) емкость.

Граф состояний СМО в этом случае имеет вид, показанный на рис. 2

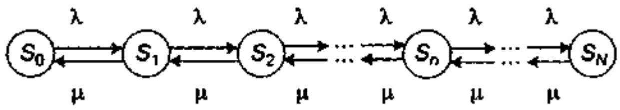


Рисунок 5.2 – Граф состояний одноканальной СМО с ожиданием (схема гибели и размножения)

Состояния СМО имеют следующую интерпретацию:

S0 — «канал свободен»;

S1 — «канал занят» (очереди нет);

S2 — «канал занят» (одна заявка стоит в очереди);

Sn — «канал занят» (п — 1 заявок стоит в очереди);

SN — «канал занят» (N — 1 заявок стоит в очереди).

Стационарный процесс в данной системе будет описываться следующей системой алгебраических уравнений:

$$\begin{cases} -\rho \cdot P_0 + P_1 = 0, & n = 0 \\ \dots\dots\dots\dots\dots\dots\dots & \\ -(1-\rho) P_n + P_{n+1} + \rho \cdot P_{n-1} = 0, & 0 < n < N \\ \dots\dots\dots\dots\dots\dots\dots & \\ -P_N + \rho \cdot P_{N-1} = 0, & n = N, \end{cases} \qquad (10)$$

где $\rho = \dfrac{\lambda}{\mu}$;

п — номер состояния.

Решение приведенной выше системы уравнений (10) для нашей модели СМО имеет вид

$$P_n = \begin{cases} \left(\dfrac{1-\rho}{1-\rho^{N+1}}\right)\cdot\rho^n, & \rho \neq 1, \quad n = 0, 1, 2, \ldots, N \\ \dfrac{1}{(N+1)}, & \rho = 1; \end{cases} \qquad (11)$$

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}}. \qquad (12)$$

Тогда

$$P_n = \begin{cases} P_0 \cdot \rho^n & \rho \neq 1, \quad n = 1, 2, \ldots, N \\ \dfrac{1}{(N+1)}, & \rho = 1. \end{cases}$$

Следует отметить, что выполнение условия стационарности $\rho = \frac{\lambda}{\mu} < 1$ для данной СМО не обязательно, поскольку число допускаемых в обслуживающую систему заявок контролируется путем введения ограничения на длину очереди (которая не может превышать N — 1), а не соотношением между интенсивностями входного потока, т. е. не отношением $\lambda/\mu=\rho$

Определим характеристики одноканальной СМО с ожиданием и ограниченной длиной очереди, равной (N — 1):

вероятность отказа в обслуживании заявки:

$$P_{\text{отк}} = P_N = \begin{cases} \left(\dfrac{1-\rho}{1-\rho^{N+1}}\right)\rho^N, & \rho \neq 1, \\ \dfrac{1}{(N+1)}, & \rho = 1; \end{cases} \qquad (13)$$

относительная пропускная способность системы:

$$q = 1 - P_{\text{отк}} = \begin{cases} 1-\left(\dfrac{1-\rho}{1-\rho^{N+1}}\right)\rho^N, & \rho \neq 1, \\ 1-\dfrac{1}{(N+1)}, & \rho = 1; \end{cases} \qquad (14)$$

абсолютная пропускная способность:

$$A = q \cdot \lambda; \qquad (15)$$

среднее число находящихся в системе заявок:

$$L_S = \sum_{n=0}^{N} n \cdot P_n = \begin{cases} \dfrac{\rho\cdot[1-(N+1)\cdot\rho^N + N\cdot\rho^{N+1}]}{(1-\rho)\cdot(1-\rho^{N+1})}, & \rho \neq 1 \\ N/2, & \rho = 1; \end{cases} \qquad (16)$$

среднее время пребывания заявки в системе:

$$W_S = \frac{L_S}{\lambda(1 - P_N)};$$ (17)

средняя продолжительность пребывания клиента (заявки) в очереди:

$$W_q = W_S - 1/\mu;$$ (18)

среднее число заявок (клиентов) в очереди (длина очереди):

$$L_q = \lambda(1 - P_N)W_q.$$ (19)

Рассмотрим пример одноканальной СМО с ожиданием.

***Пример** 2.* Специализированный пост диагностики представляет собой одноканальную СМО. Число стоянок для автомобилей, ожидающих проведения диагностики, ограниченно и равно 3[$(N - 1) = 3$]. Если все стоянки заняты, т. е. в очереди уже находится три автомобиля, то очередной автомобиль, прибывший на диагностику, в очередь на обслуживание не становится. Поток автомобилей, прибывающих на диагностику, распределен по закону Пуассона и имеет интенсивность λ = 0,85 (автомобиля в час). Время диагностики автомобиля распределено по показательному закону и в среднем равно 1,05 час.

Требуется определить вероятностные характеристики поста диагностики, работающего в стационарном режиме.

Решение

1. Параметр потока обслуживаний автомобилей:

$$\mu = \frac{1}{\bar{t}} = \frac{1}{1,05} = 0,952.$$

2. Приведенная интенсивность потока автомобилей определяется как отношение интенсивностей λ, и μ, т. е.

$$\rho = \frac{\lambda}{\mu} = \frac{0,85}{0,952} = 0,893.$$

3. Вычислим финальные вероятности системы

$$P_0 = \frac{1-\rho}{1-\rho^{N+1}} = \frac{1-0,893}{1-0,893^5} \approx 0,248;$$

$$P_1 = \rho \cdot P_0 = 0,893 \cdot 0,248 \approx 0,221;$$

$$P_2 = \rho^2 \cdot P_0 = 0,893^2 \cdot 0,248 \approx 0,198;$$

$$P_3 = \rho^3 \cdot P_0 = 0,893^3 \cdot 0,248 \approx 0,177;$$

$$P_4 = \rho^4 \cdot P_0 = 0,893^4 \cdot 0,248 \approx 0,158.$$

4. Вероятность отказа в обслуживании автомобиля:

$$P_{отк} = P_4 = \rho^4 \cdot P_0 \approx 0,158.$$

5. Относительная пропускная способность поста диагностики:

$$q = 1 - P_{отк} = 1 - 0,158 = 0,842.$$

6. Абсолютная пропускная способность поста диагностики

$A = \lambda \cdot q = 0,85 \cdot 0,842 = 0,716$ (автомобиля в час).

7. Среднее число автомобилей, находящихся на обслуживании и в очереди (т.е. в системе массового обслуживания):

$$L_S = \frac{\rho \cdot [1-(N+1)\cdot\rho^N + N\cdot\rho^{N+1}]}{(1-\rho)\cdot(1-\rho^{N+1})} =$$

$$= \frac{0,893\cdot[1-(4+1)\cdot 0,893^4 + 4\cdot 0,893^5]}{(1-0,893)\cdot(1-0,893^5)} = 1,77.$$

8. Среднее время пребывания автомобиля в системе:

$$W_S = \frac{L_S}{\lambda\cdot(1-P_N)} = \frac{1,77}{0,85(1-0,158)} = 2,473 \text{ часа.}$$

9. Средняя продолжительность пребывания заявки в очереди на обслуживание:

$$W_q = W_S - 1/\mu = 2,473 - 1/0,952 = 1,423 \text{ часа.}$$

10. Среднее число заявок в очереди (длина очереди):

$$L_q = \lambda\cdot(1-P_N)\cdot W_q = 0,85\cdot(1-0,158)\cdot 1,423 = 1,02.$$

Работу рассмотренного поста диагностики можно считать удовлетворительной, так как пост диагностики не обслуживает автомобили в среднем в 15,8% случаев ($P_{отк}$ = 0,158).

Перейдем теперь к рассмотрению **одноканальной СМО с ожиданием без ограничения на вместимость блока ожидания (т. е. N $\to\infty$).** Остальные условия функционирования СМО остаются без изменений.

Стационарный режим функционирования данной СМО существует при t $\to\infty$ оо для любого n = 0, 1, 2, ... и когда $\lambda < \mu$. Система алгебраических уравнений, описывающих работу СМО при t $\to\infty$ для любого n = 0, 1, 2, ... , имеет вид

$$\begin{cases} -\lambda\cdot P_0 + \mu\cdot P_1 = 0, n = 0 \\ \lambda\cdot P_{n-1} + \mu\cdot P_{n+1} - (\lambda+\mu)\cdot P_n = 0, \quad n > 0. \end{cases} \quad (20)$$

Решение данной системы уравнений имеет вид

$$P_n = (1-\rho)\,\rho^n, \quad n = 0, 1, 2, ..., \quad (21)$$

где $\rho = \lambda/\mu < 1$.

Характеристики одноканальной СМО с ожиданием, без ограничения на длину очереди, следующие:

• среднее число находящихся в системе клиентов (заявок) на обслуживание:

$$L_S = \sum_{n=0}^{\infty} n\cdot P_n = \frac{\rho}{1-\rho}; \quad (22)$$

средняя продолжительность пребывания клиента в системе:

$$W_S = \frac{L_S}{\lambda} = \frac{1}{[\mu\cdot(1-\rho)]}; \quad (23)$$

среднее число клиентов в очереди на обслуживании:

$$L_q = L_S - \frac{\lambda}{\mu} = \frac{\rho^2}{(1-\rho)}; \quad (24)$$

средняя продолжительность пребывания клиента в очереди:

$$W_q = \frac{L_q}{\lambda} = \frac{\rho}{[\mu\cdot(1-\rho)]}. \quad (25)$$

Пример 3. Вспомним о ситуации, рассмотренной в примере 2, где речь идет о функционировании поста диагностики. Пусть рассматриваемый пост диагностики располагает неограниченным количеством площадок для стоянки прибывающих на обслуживание автомобилей, т. е. длина очереди не ограничена.

Требуется определить финальные значения следующих вероятностных характеристик:

вероятности состояний системы (поста диагностики);

– среднее число автомобилей, находящихся в системе (на обслуживании и в очереди);

– среднюю продолжительность пребывания автомобиля в системе (на обслуживании и в очереди);

– среднее число автомобилей в очереди на обслуживании;

– среднюю продолжительность пребывания автомобиля в очереди.

Решение

1. Параметр потока обслуживания μ и приведенная интенсивность потока автомобилей ρ определены в примере 2:

$$\mu = 0{,}952;\ \rho = 0{,}893.$$

2. Вычислим предельные вероятности системы по формулам

$P_0 = 1 - \rho = 1 - 0{,}893 = 0{,}107$;

$P_1 = (1 - \rho) \,.\, \rho = (1 - 0{,}893)*0{,}893 = 0{,}096$;

$P_2 = (1 - \rho) \,.\, \rho^2 = (1 - 0{,}893)*0{,}8932 = 0{,}085$;

$P_3 = (1 - \rho) \,.\, \rho^3 = (1 - 0{,}893)*0{,}8933 = 0{,}076$;

$P_4 = (1 - \rho) \,.\, \rho^4 = (1 - 0{,}893)*0{,}8934 = 0{,}068$;

$P_5 = (1 - \rho) \,.\, \rho^5 = (1 - 0{,}893)*0{,}8935 = 0{,}061$ и т. д.

Следует отметить, что P_0 определяет долю времени, в течение которого пост диагностики вынужденно бездействует (простаивает). В нашем примере она составляет 10,7%, так как $P_0 = 0{,}107$.

3. Среднее число автомобилей, находящихся в системе (на обслуживании и в очереди):

$$L_S = \frac{\rho}{1-\rho} = \frac{0.893}{1-0.893} = 8.346\,\textit{ед.}$$

4. Средняя продолжительность пребывания клиента в системе:

$$W_S = \frac{L_S}{\lambda} = \frac{1}{[\mu \cdot (1-\rho)]}$$

$$W_S = \frac{1}{0.952*(1-0.893)} = 9.817\ \textit{час.}$$

5. Среднее число автомобилей в очереди на обслуживание:

$$L_q = L_S - \frac{\lambda}{\mu} = \frac{\rho^2}{(1-\rho)} = \frac{0.893^2}{1-0.893} = 7.453$$

6. Средняя продолжительность пребывания автомобиля в очереди:

$$W_q = \frac{\rho}{\mu \cdot (1-\rho)} = \frac{0.893}{0.952 \cdot (1-0.893)} = 8.766 \ \text{час.}$$

7. Относительная пропускная способность системы:

q = 1,

т. е. каждая заявка, пришедшая в систему, будет обслужена.

8. Абсолютная пропускная способность:

A = λ* q = 0,85 * 1 = 0,85.

Следует отметить, что предприятие, осуществляющее диагностику автомобилей, прежде всего интересует количество клиентов, которое посетит пост диагностики при снятии ограничения на длину очереди.

Допустим, в первоначальном варианте количество мест для стоянки прибывающих автомобилей было равно трем (см. пример 2). Частота *m* возникновения ситуаций, когда прибывающий на пост диагностики автомобиль не имеет возможности присоединиться к очереди:

$m=\lambda*P_N$

В нашем примере при N = 3 + 1 = 4 и ρ = 0,893

$m=\lambda*P_0*\rho^4=0.85*0.248*0.8934=0.134$ автомобиля в час.

При 12-часовом режиме работы поста диагностики это эквивалентно тому, что пост диагностики в среднем за смену (день) будет терять 12 * 0,134 = 1,6 автомобиля. Снятие ограничения на длину очереди позволяет увеличить количество обслуженных клиентов в нашем при мере в среднем на 1,6 автомобиля за смену (12 ч. работы) поста диагностики. Ясно, что решение относительно расширения площади для стоянки автомобилей, прибывающих на пост диагностики, должно основываться на оценке экономического ущерба, который обусловлен потерей клиентов при наличии всего трех мест для стоянки этих автомобилей.

Многоканальная модель с пуассоновским входным потоком и экспоненциальным распределением длительности обслуживания

В подавляющем большинстве случаев на практике системы массового обслуживания являются многоканальными, и, следовательно, модели с n обслуживающими каналами (где n > 1) представляют несомненный интерес.

Процесс массового обслуживания, описываемый данной моделью, характеризуется интенсивностью входного потока λ, при этом параллельно может обслуживаться не более n клиентов (заявок). Средняя продолжительность обслуживания одной заявки равняется l/μ. Входной и выходной потоки являются пуассоновскими. Режим функционирования того или иного обслуживающего канала не влияет на режим функционирования других обслуживающих каналов системы, причем длительность процедуры обслуживания каждым из каналов является случайной величиной, подчиненной экспоненциальному закону распределения. Конечная цель использования n параллельно включенных обслуживающих каналов заключается в повышении (по сравнению с одноканальной системой) скорости обслуживания требований за счет обслуживания одновременно n клиентов.

Граф состояний многоканальной системы массового обслуживания с отказами имеет вид, показанный на рис. 4.3.

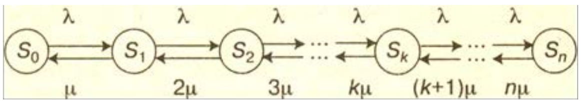


Рисунок-4.3 Граф состояний многоканальной СМО с отказами

Состояния данной СМО имеют следующую интерпретацию:

S_0 - все каналы свободны;

S_1 - занят один канал, остальные свободны;

……………………….

S_k - заняты ровно k каналов, остальные свободны;

……………………….

S_n - заняты все n каналов, заявка получает отказ в обслуживании.

Уравнения Колмогорова для вероятностей состояний системы P_0, …, P_k, …, P_n будут иметь следующий вид:

$$\begin{cases} \dfrac{dP_0}{dt} = -\lambda \cdot P_0 + \mu \cdot P_1; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{dP_k}{dt} = \lambda \cdot P_{k-1} - (\lambda + k + \mu) \cdot P_k + \mu \cdot (k+1) \cdot P_{k+1}, \ 1 \le k \le n-1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{dP_n}{dt} = \lambda \cdot P_{n-1} - \mu \cdot n \cdot P_n. \end{cases} \quad (26)$$

Начальные условия решения системы таковы:

$P_0(0)=1,\ P_1(0)=P_2(0)=\ldots=P_k(0)=\ldots=P_n(0)=0.$

Стационарное решение системы имеет вид:

$$\begin{cases} P_k = \dfrac{\dfrac{\rho^k}{k!}}{\sum\limits_{k=0}^{n} \dfrac{\rho^k}{k!}} = \dfrac{\rho^k}{k!} \cdot P_0, \quad k = 0,1,2,\ldots,n \\ P_0 = \dfrac{1}{\left[\sum\limits_{k=0}^{n} \dfrac{\rho^k}{k!}\right]}, \qquad k = 0,1,2,\ldots,n, \end{cases} \tag{27}$$

где $\rho = \dfrac{\lambda}{\mu}$.

Формулы для вычисления вероятностей P_k называются формулами Эрланга.

Определим вероятностные характеристики функционирования многоканальной СМО с отказами в стационарном режиме:

– вероятность отказа:

$$P_{отк} = P_n = \frac{\rho^n}{n!} P_0, \tag{28}$$

так как заявка получает отказ, если приходит в момент, когда все *n* каналов заняты. Величина $P_{отк}$ характеризует полноту обслуживания входящего потока;

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

$$q = 1 - P_{отк} = 1 - \frac{\rho_n}{n!} \cdot P_0, \tag{29}$$

– абсолютная пропускная способность

$$A=\lambda*q=\lambda*(1-P_{отк}); \tag{30}$$

– среднее число каналов, занятых обслуживанием следующее:

$$\bar{k} = \sum_{k=1}^{n} k \cdot P_k = \rho \cdot (1 - P_{отк}). \tag{31}$$

Оно характеризует степень загрузки системы.

Пример 4. Пусть n-канальная СМО представляет собой вычислительный центр (ВЦ) с тремя (n = 3) взаимозаменяемыми ПЭВМ для решения поступающих задач. Поток задач, поступающих на ВЦ, имеет интенсивность λ = 1 задаче в час. Средняя продолжительность обслуживания $t_{обсл}$ = 1,8 час.

Поток заявок на решение задач и поток обслуживания этих заявок являются простейшими.

Требуется вычислить финальные значения:

- вероятности состояний ВЦ;
- вероятности отказа в обслуживании заявки;
- относительной пропускной способности ВЦ;
- абсолютной пропускной способности ВЦ;
- среднего числа занятых ПЭВМ на ВЦ.

Определите, сколько дополнительно надо приобрести ПЭВМ, чтобы увеличить пропускную способность ВЦ в 2 раза.

Решение

1. Определим параметр μ потока обслуживании:

$$\mu = \frac{1}{\bar{t}_{обсл}} = \frac{1}{1,8} = 0,555.$$

2. Приведенная интенсивность потока заявок

$\rho=\lambda/\mu=1/0.555=1.8$

3. Предельные вероятности состояний найдем по формулам Эрланга (27):

$$P_1 = \frac{\rho}{1!} \cdot P_0 = 1.8 \cdot P_0;$$

$$P_2 = \frac{\rho^2}{2!} \cdot P_0 = 1.62 \cdot P_0;$$

$$P_3 = \frac{\rho^3}{3!} \cdot P_0 = 0.97 \cdot P_0;$$

$$P_0 = \frac{1}{\sum_{k=0}^{3} \frac{\rho^k}{k!}} = \frac{1}{1+1.8+1.62+0.97} \approx 0.186;$$

P_1=1.8*0.186=0.334;
P_2=1.62*0.186=0.301;
P_3=0.97*0.186=0.180.

4. Вероятность отказа в обслуживании заявки

$$P_{отк}=P_3=0.180$$

5. Относительная пропускная способность ВЦ

$$q = 1 - P_{отк} = 1 - 0.180 = 0,820.$$

6. Абсолютная пропускная способность ВЦ

$$A = \lambda \cdot q = 1 \cdot 0,820 = 0,820.$$

7. Среднее число занятых каналов — ПЭВМ

$$\bar{k} = \rho \cdot (1 - P_{отк}) = 1,8 \cdot (1 - 0.180) = 1.476$$

Таким образом, при установившемся режиме работы СМО в среднем будет занято 1,5 компьютера из трех — остальные полтора будут простаивать. Работу рассмотренного ВЦ вряд ли можно считать удовлетворительной, так как центр не обслуживает заявки в среднем в 18% случаев (P_3 =0,180).

Очевидно, что пропускную способность ВЦ при данных λ и μ можно увеличить только за счет увеличения числа ПЭВМ.

Определим, сколько нужно использовать ПЭВМ, чтобы сократить число не обслуженных заявок, поступающих на ВЦ, в 10 раз, т.е. чтобы вероятность отказа в решении задач не превосходила 0,0180. Для этого используем формулу (28):

$$P_{отк} = P_n = \frac{\rho^n}{n!} P_0 ,$$

Составим следующую таблицу:

n	1	2	3	4	5	6
P_0	0,357	0,226	0,186	0,172	0,167	0,166
$P_{отк}$	0,643	0,367	0,18	0,075	0,026	0,0078

Анализируя данные таблицы, следует отметить, что расширение числа каналов ВЦ при данных значениях λ и μ до 6 единиц ПЭВМ позволит обеспечить удовлетворение заявок на решение задач на 99,22%, так как при n = 6 вероятность отказа в обслуживании *($P_{отк}$)* составляет 0,0078.

Многоканальная система массового обслуживания с ожиданием

Процесс массового обслуживания при этом характеризуется следующим: входной и выходной потоки являются пуассоновскими с интенсивностями λ и μ соответственно; параллельно обслуживаться могут не более C клиентов. Система имеет C каналов обслуживания. Средняя продолжительность обслуживания $\frac{1}{\mu}$ одного клиента равна

В установившемся режиме функционирование многоканальной СМО с ожиданием и неограниченной очередью может быть описано с помощью системы алгебраических уравнений:

$$\begin{aligned} &0 = \lambda \cdot P_{n-1} - (\lambda + n \cdot \mu) \cdot P_n + (n+1)\,\mu \cdot P_{n+1}, \\ &\text{при } 1 \le n < C; \\ &0 = \lambda \cdot P_{n-1} - (\lambda + C \cdot \mu) \cdot P_n + C \cdot \mu \cdot P_{n+1}, \\ &\text{при } n \ge C. \end{aligned} \qquad (32)$$

Решение системы уравнений (32) имеет вид

$$\begin{cases} P_n = \dfrac{\rho^n}{n!} \cdot P_0, & \text{при } 0 \le n < C, \\ P_n = \dfrac{\rho^n}{C! \cdot C^{n-c}} \cdot P_0, & \text{при } n \ge C, \end{cases} \qquad (33) \quad (34)$$

где

$$P_0 = \left\{ \sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{C! \left[1 - \left(\frac{\rho}{C} \right) \right]} \right\}^{-1} \tag{35}$$

Решение будет действительным, если выполняется следующее условие: $\left[\frac{\lambda}{\mu \cdot C}\right] < 1.$

Вероятностные характеристики функционирования в стационарном режиме многоканальной СМО с ожиданием и неограниченной очередью определяются по следующим формулам:

- вероятность того, что в системе находится n клиентов на обслуживании, определяется по формулам (33) и (34);
- среднее число клиентов в очереди на обслуживание

$$L_q = \left[\frac{C \cdot \rho}{(C-\rho)^2}\right] \cdot P_C; \tag{36}$$

- среднее число находящихся в системе клиентов (заявок на обслуживание и в очереди)

$$Ls = Lq + \rho \tag{37}$$

- средняя продолжительность пребывания клиента (заявки на обслуживание) в очереди

$$Wq = \frac{L_q}{\lambda}; \tag{38}$$

- средняя продолжительность пребывания клиента в системе

$$Ws = W_q + \frac{1}{\mu}. \tag{39}$$

Рассмотрим примеры многоканальной системы массового обслуживания с ожиданием.

Пример 5. Механическая мастерская завода с тремя постами (каналами) выполняет ремонт малой механизации. Поток неисправных механизмов, прибывающих в мастерскую, — пуассоновский и имеет интенсивность λ= 2,5 механизма в сутки, среднее время ремонта одного механизма распределено по показательному закону и равно t = 0,5 сут. Предположим, что другой мастерской на заводе нет, и, значит, очередь механизмов перед мастерской может расти практически неограниченно.

Требуется вычислить следующие предельные значения вероятностных характеристик системы:

– вероятности состояний системы;
– среднее число заявок в очереди на обслуживание;
– среднее число находящихся в системе заявок;
– среднюю продолжительность пребывания заявки в очереди;
– среднюю продолжительность пребывания заявки в системе.

Решение

1. Определим параметр потока обслуживаний

$$\mu = 1/t=1/0,5 = 2.$$

2. Приведенная интенсивность потока заявок

ρ = λ/μ = 2,5/2,0 = 1,25,

при этом λ/μ *c= 2,5/2 * 3 = 0,41.

Поскольку λ/μ * с <1 , то очередь не растет безгранично и в системе наступает предельный стационарный режим работы.

3. Вычислим вероятности состояний системы:

$$P_0 = \left\{\sum_{n=0}^{C-1}\frac{\rho^n}{n!} + \frac{\rho^C}{C!\left[1-\left(\frac{\rho}{C}\right)\right]}\right\}^{-1} = \frac{1}{\left[1+\frac{\rho^1}{1!}+\frac{\rho^2}{2!}+\frac{\rho^3}{3!\left(1-\frac{\rho}{3}\right)}\right]} =$$

$$= \frac{1}{\left[1+\rho+\frac{\rho^2}{2}+\frac{\rho^3}{6\cdot(1-\frac{\rho}{3})}\right]} = \frac{1}{\left[1+1,25+\frac{1,25^2}{2}+\frac{1,25^3}{6\cdot\left(1-\frac{1,25}{3}\right)}\right]} = 0,279;$$

$$P_1 = \frac{\rho^1}{1!}\cdot P_0 = 1,25\cdot 0,279 = 0,349;$$

$$P_2 = \frac{\rho^2}{2!}\cdot P_0 = \frac{1,25^2}{2!}\cdot 0,279 = 0,218;$$

$$P_3 = \frac{\rho^3}{3!}\cdot P_0 = \frac{1,25^3}{3!}\cdot 0,279 = 0,091;$$

$$P_4 = \frac{\rho^4}{4!}\cdot P_0 = \frac{1,25^4}{4!}\cdot 0,279 = 0,028.$$

4. Вероятность отсутствия очереди у мастерской

$$P_{\text{от.о}} = P_0 + P_1 + P_2 + P_3 =$$

$$= 0,279 + 0,349 + 0,218 + 0,091 = 0,937.$$

5. Среднее число заявок в очереди на обслуживание

$$L_q = \left[\frac{C\cdot\rho}{(C-\rho)^2}\right]\cdot P_C = \left[\frac{3\cdot 1,25}{(3-1,25)^2}\right]\cdot 0,091 = 0,111.$$

6. Среднее число находящихся в системе заявок

$L_s = L_q + \rho = 0,111 + 1,25 = 1,361.$

7. Средняя продолжительность пребывания механизма в очереди на обслуживание

$$W_q = \frac{L_q}{\lambda} = \frac{0.111}{2.5} = 0.044 \ \textit{суток}$$

8. Средняя продолжительность пребывания механизма в мастерской (в системе)

$$\mathrm{Ws} = \mathrm{Wq} + \frac{1}{\mu} = 0,044 + \frac{1}{2} = 0,544 \ \textit{суток}$$

ЛЕКЦИЯ 11
Альтернативные подходы к созданию имитационных моделей

Разработчики моделирования изначально направляли свои усилия на поиск новых и более совершенных способов моделирования систем, используя при этом существующее компьютерное оборудование и программное обеспечение.

Параллельное и распределенное моделирование

Большинство моделей действуют приблизительно одинаково. Часы модельного времени взаимодействуют со списком событий, в результате чего определяется, какое событие будет обрабатываться следующим. Часы переводятся на время возникновения этого события, а компьютер выполняет логику события, которая может потребовать: обновить переменные, обработать списки очередей и событий, сгенерировать случайные числа и величины, собрать статистику. Данная логика выполняется в соответствии с модельным временем возникновения событий, то есть в этих случаях моделирование является *последовательным.*

В последние годы компьютерная технология позволила связывать отдельные компьютеры или процессоры в параллельную или распределенную вычислительную среду. Например, несколько относительно недорогих мини-компьютеров (или даже микрокомпьютеров) могут быть связаны друг с другом с помощью сети, а в более крупном компьютере можно разместить несколько процессоров, которые будут работать как сами по себе, так и во взаимодействии друг с другом. Такая вычислительная среда позволяет распределять различные части задачи между отдельными процессорами, работающими одновременно или параллельно, и таким образом уменьшать общее время, необходимое для решения данной задачи. Конечно, такая возможность во многом зависит от сути самой задачи, а также от доступного компьютерного оборудования и программного обеспечения. В настоящее время распределенная и параллельная обработка используется во многих областях, например при оптимизации и проектировании баз данных.

Существует много способов разделить динамически моделируемые системы на части, то есть распределить работу между различными процессорами. Пожалуй, самый прямой способ — это распределение отдельных «функций поддержки» (например, генерирования случайных чисел, генерирования случайных величин, обработки списка событий, обработки списков и очередей, а также сбора статистики) по различным процессорам. Логическое выполнение моделирования все же остается последовательным, однако теперь «главная» моделирующая программа может передавать выполнение функций поддержки другим процессорам и продолжать свою работу.

Другой способ распределения имитационной модели между различными процессорами — декомпозиция модели на отдельные подмодели, которые выполняются несколькими процессорами. Например, производственное оборудование очень часто моделируется как сеть взаимосвязанных станций обслуживания, каждая из них представляет отдельный тип операций. Отдельные подмодели (или их группы) распределены между различными процессорами, и каждый из них моделирует свою часть системы. Необходимость обеспечивать соответствующие логические отношения между подмоделями предусматривает обмен информацией между процессорами. В примере с производственным оборудованием это происходит, когда деталь покидает одну станцию и переходит к другой станции, моделируемой на другом процессоре. Но не следует забывать о необходимости соблюдать правильную временную последовательность выполнения операций, то есть о *синхронизации* обработки подмоделей различными процессорами для правильного представления общих действий модели. **Одно из основных преимуществ такого вида распределенного моделирования заключается в отсутствии глобальных часов модельного времени и полного списка событий.** Поскольку обработка списка событий при традиционном имитационном моделировании может занять много времени, возможность выполнения программы без списка событий — весьма привлекательная идея. Место часов модельного времени и списка событий занимает система для *передачи сообщений* между процессорами, где каждое сообщение содержит «отметку времени». Недостаток такого метода заключается в том, что при моделировании может возникнуть *тупик* — два процессора будут вынуждены ожидать сообщения друг от друга для продолжения работы (даже если в реальной моделируемой системе такое не произойдет). Аналогичная ситуация может привести к остановке моделирования, поэтому должен применяться метод нахождения и устранения тупиков (или возможность их избегать).

Другая концепция, связанная с распределением подмоделей между параллельными процессорами, известна как концепция *виртуального времени*, реализованная через *механизм изменения шкалы времени*. Выше указывалось, что каждый процессор моделирует свою часть системы по времени, но *не* ожидает получения сообщений от других процессоров, обладающих возможностью работать одновременно с ним, хотя и с различной скоростью. Такое ожидание неизбежно при описанном подходе с передачей сообщений. Если подмодель, создаваемая отдельным процессором, получает сообщение, которое она должна была получить ранее (и оно потенциально влияет на ее действия с того момента времени), для этой подмодели выполняется *откат* путем возвращения ее времени к более раннему моменту получения сообщения. Например, подмодель Б моделировалась до момента времени 87, и в это время поступило сообщение от подмодели А, которое модель Б должна была получить в момент времени 61. Тогда часы для подмодели переводятся назад на момент времени 61, а моделирование подмодели Б между моментами времени 61 и 87 отменяется, поскольку без учета сообщения от момента времени 61 оно могло быть выполнено неправильно. При этом к отмененной части моделирования могут относиться сообщения другим подмоделям, теперь каждое из них аннулируется посредством отправки соответствующего *антисообщения*. Антисообщения, в свою очередь, могут генерировать вторичные откаты в получающих подмоделях и т. д. Подход, когда, во-первых, работа, выполненная между моментами времени 61 и 87, утрачивается, а во-вторых, в связи с откатом моделирования возникают дополнительные расходы, может показаться неудачным. Однако при этом процессоры все время заняты моделированием (за исключением времени отката), вместо того чтобы бездействовать в ожидании, сообщения перед дальнейшим «правильным» продвижением моделирования по времени. В стохастическом моделировании нельзя точно определить, понадобится ли откат, поэтому механизм изменения шкалы времени называют «рулеткой». Его недостаток заключается в возникновении расходов на дополнительную память и обработку возможных откатов, а преимущество состоит в том, что откаты могут быть редкими, и все процессоры будут продолжать продвижение вперед в моделировании. Более того, при использовании механизма изменения шкалы времени не появляются тупики. Вопросы разработки и оценки распределенной обработки при моделировании активно исследуются. Однако уже теперь ясно, что, насколько удачным окажется тот или иной метод (и будет ли он работать вообще), зависит от структуры и параметров модели, а также от доступной компьютерной среды. Например, если модель можно поделить на подмодели, слабо связанные между собой (в частности сеть очередей, в которой требования редко переходят из одной очереди в другую), тогда любая из схем распределения моделирования, описанных выше, может предоставлять некоторые преимущества.

Непрерывное моделирование

Непрерывное моделирование — это моделирование системы по времени с помощью представления, в котором переменные состояния меняются непрерывно по отношению ко времени. Как правило, в непрерывных имитационных моделях используются дифференциальные уравнения, которые устанавливают отношения для скоростей изменения переменных состояния во времени. Если дифференциальные уравнения очень просты, их можно решать аналитически, чтобы представить значения переменных состояния для всех значений времени как функцию значений переменных состояния в момент времени 0. При больших непрерывных моделях аналитическое решение невозможно, но для численного интегрирования дифференциальных уравнений в случае с заданными специальными значениями для переменных состояния в момент времени 0 используются технологии численного анализа, например интегрирование Рунге-Кутта.

Пример 1.3. Рассмотрим непрерывную модель соперничества между двумя популяциями. Биологические модели такого типа, именуемые моделями *хищник-добыча* (или *паразит-хозяин),* рассматривались многими авторами, в том числе Брауном и Гордоном. Среда представлена двумя популяциями -хищников и добычи, взаимодействующими друг с другом. Добыча пассивна, но хищники зависят от ее популяции, поскольку она является для них источником пищи. (Например, хищниками могут быть акулы, а добычей — рыба, которой они питаются) Пусть $x(t)$ и $y(t)$ обозначают численность особей в популяциях соответственно добычи и хищников в момент времени t. Допустим, популяция добычи имеет обильные запасы пищи; при отсутствии хищников темп ее прироста составит $rx(t)$ для некоторого положительного значения r *(r* —

естественный уровень рождаемости минус естественный уровень смертности). Существование взаимодействия между хищниками и добычей дает основание предположить, что уровень смертности добычи в связи с этим взаимодействием пропорционален произведению численностей обоих популяций *x(t)y(t).* Поэтому общий темп изменения популяции добычи d*x/dt:* может быть представлен как

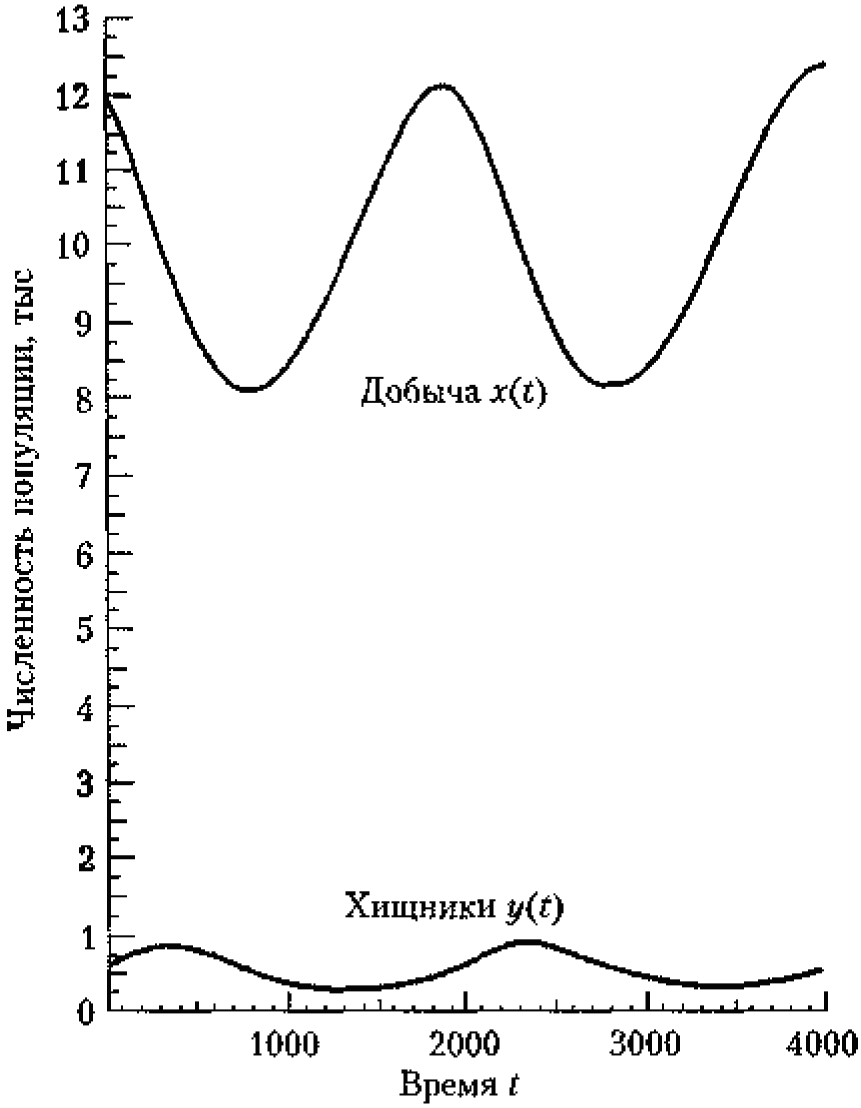
$$\frac{dx}{dt} = rx(t) - ax(t)y(t) \qquad (1)$$

где *a* — положительный коэффициент пропорциональности. Поскольку существование самих хищников зависит от популяции добычи, темп изменения популяции хищников в отсутствии добычи составляет -*sy(t)* для некоторого положительного s. Более того, взаимодействие между двумя популяциями приводит к росту популяции хищников, темп которого также пропорционален *x(t)y(t).* Следовательно, общий темп изменения популяции хищников *dy/dt* составляет

$$\frac{dy}{dt} = -sy(t) + bx(t)y(t) \qquad (2)$$

где *b* — положительный коэффициент пропорциональности. При начальных условиях *x(0)* > 0 и у(0) >*0* решение модели, определенной уравнениями (1) и (2), имеет интересное свойство: *x(t)* > 0 и *y(t)* > 0 для любого t≥0. Следовательно, популяция добычи никогда не будет полностью уничтожена хищниками. Решение *{x(t), y(t)}* также является периодической функцией времени. Иными словами, существует такое значение *T*>0, при котором *x(t + nT)=x(t)* и *y(t + nT) = y(t)* для любого положительного целого числа *n.* Такой результат не является неожиданным. По мере увеличения популяции хищников популяция добычи уменьшается. Это приводит к снижению темпа роста популяции хищников и, соответственно, вызывает уменьшение их числа, что, в свою очередь, ведет к увеличению популяции добычи и т. д.

Рассмотрим отдельные значения г = 0,001, а = 2 * 10 $^{-6}$; s = 0,01; b=10 $^{-6}$, исходные размеры популяций составляют *x*(0) = 12 000 и у(0) = 600. На рис. представлено численное решение уравнений (1) и (2), полученное при использовании вычислительного пакета, разработанного для численного решения систем дифференциальных уравнений (а не языка непрерывного моделирования).



Обратите внимание на то, что приведенный выше пример полностью детерминистический, то есть в нем нет случайных компонентов. Однако имитационная модель может содержать и неизвестные величины; например, в уравнения (1) и (2) могут быть добавлены случайные величины, которые каким-то образом зависят от времени, или постоянные множители могут быть смоделированы как величины, случайно изменяющие свои значения в определенные моменты времени.

Комбинированное непрерывно-дискретное моделирование

Поскольку некоторые из систем невозможно отнести ни к полностью дискретным, ни к полностью непрерывным, может возникнуть необходимость в создании модели, которая объединяет в себе аспекты как дискретно-событийного, так и непрерывного моделирования, в результате чего получается *комбинированное непрерывно- дискретное* моделирование. Между дискретным и непрерывным изменениями переменных состояния могут происходить три основных типа взаимодействия:

- дискретное событие может вызвать дискретное изменение в значении непрерывной переменной состояния;

- в определенный момент времени дискретное событие может вызвать изменение отношения, управляющего непрерывной переменной состояния;

- непрерывная переменная состояния, достигшая порогового значения, может вызвать возникновение или планирование дискретного события.

В следующем примере комбинированного непрерывно-дискретного моделирования дано краткое описание модели, подробно рассмотренной Прицкером, который в своей работе приводит и другие примеры этого типа моделирования.

Пример 1.4. Танкеры, перевозящие нефть, прибывают в один разгрузочный док, пополняя резервуар-хранилище, из которого нефть по трубопроводу попадает на нефтеперегонный завод. Из разгружающегося танкера нефть подается в резервуар-хранилище с постоянной скоростью (Танкеры, прибывающие к занятому доку, образуют очередь.) На нефтеперегонный завод нефть подается из резервуара с различными заданными скоростями. Док открыт с 6.00 до 24.00. По соображениям безопасности разгрузка танкеров прекращается по закрытии дока.

Дискретными событиями в этой (упрощенной) модели являются прибытие танкера на разгрузку, закрытие дока в полночь и открытие в 6.00. Уровни нефти в разгружающемся танкере и резервуаре-хранилище задаются переменными непрерывного состояния, скорости изменения которых описаны с помощью дифференциальных уравнений. Разгрузка танкера считается завершенной, когда уровень нефти в танкере составляет менее 5 % его емкости, но разгрузка должна быть временно прекращена, если уровень нефти в резервуаре-хранилище станет равным его емкости. Разгрузка может быть возобновлена, когда уровень нефти в резервуаре станет меньше 80 % его емкости. В случае если уровень нефти в резервуаре станет меньше 5000 баррелей, нефтеперегонный завод должен быть временно закрыт. Для того чтобы избежать частого закрытия и возобновления работы завода, подача нефти из резервуара на завод не будет возобновляться до тех пор, пока в нем не наберется 50 000 баррелей нефти. Каждое из пяти событий, связанных с уровнем нефти (например, падение уровня нефти ниже 5 % емкости танкера), по определению Прицкера, является *событием состояния*. В отличие от дискретных событий, события состояния не планируются, они происходят, когда переменные непрерывного состояния переходят пороговое значение.

Моделирование по методу Монте-Карло. Статистическое моделирование систем

5.4.1 Теоретические основы метода

Метод статистического моделирования (или метод Монте-Карло) — это способ исследования поведения вероятностных систем (экономических, технических и т. д.) в условиях, когда не известны в полной мере внутренние взаимодействия в этих системах. Название метода Монте-Карло появилось во время второй мировой войны, когда этот подход был применен к проблемам, связанным с разработкой атомной бомбы

Этот метод заключается в воспроизведении исследуемого физического процесса при помощи вероятностной математической модели и вычислении характеристик этого процесса. Одно такое воспроизведение функционирования системы называют реализацией или испытанием. После каждого испытания регистрируют совокупность параметров, характеризующих случайный исход реализации. Метод основан на многократных испытаниях построенной модели с последующей статистической обработкой полученных данных с целью определения числовых характеристик рассматриваемого процесса в виде статистических оценок его параметров. Процесс моделирования функционирования экономической системы сводится к машинной имитации изучаемого процесса, который как бы копируется на ЭВМ со всеми сопровождающими его случайностями. Первые сведения о методе Монте-Карло были опубликованы в конце 40-х гг. Авторами метода являются американские математики Дж. Нейман и С. Улам. В нашей стране первые работы были опубликованы в 1955-1956 гг. В.В. Чавчанидзе, Ю.А. Шрейдером и В.С. Владимировым.

Основой метода статистического моделирования является закон больших чисел. Закон больших чисел в теории вероятностей доказывает для различных условий сходимость по вероятности средних значений результатов большого числа наблюдений к некоторым постоянным величинам.

Под законом больших чисел понимают ряд теорем. Например, одна из теорем П.Л. Чебышева формулируется так: «При неограниченном увеличении числа независимых испытаний n среднее арифметическое свободных от систематических ошибок и равноточных результатов наблюдений ξi случайной величины ξ, имеющей конечную дисперсию D(ξ), сходится по вероятности к математическому ожиданию M(ξ) этой случайной величины». Это можно записать в следующем виде:

$$\lim_{n\to\infty} P\left\{\left|\frac{\sum_{i=1}^{n}\xi_i}{n} - M(\xi)\right| < \varepsilon\right\} = 1, \qquad (1)$$

где ε — сколь угодно малая положительная величина.

Теорема Бернулли формулируется так: «При неограниченном увеличении числа независимых испытаний в одних и тех же условиях частота P*(A) наступления случайного события A сходится по вероятности к его вероятности P», т. е.

$$\lim_{n\to\infty} P\left\{\left|\frac{m_i^*}{n} - P\right| < \varepsilon\right\} = 1. \qquad (2)$$

Согласно данной теореме, для получения вероятности какого-либо события, например вероятности состояний некоторой системы Pi(t),i = 0,k, вычисляют частоты $P_i^* = \frac{m_i^*}{n}$ для одной реализации (испытания), далее проводят подобные вычисления для числа реализаций, равного n. Результаты усредняют и этим самым с некоторым приближением, получают искомые вероятности состояний системы. На основании вычисленных вероятностей определяют другие характеристики системы. Следует отметить, что, чем больше число реализаций n, тем точнее результаты вычисления искомых величин (вероятностей состояний системы).

Последнее утверждение легко доказать. Предположим, что требуется найти неизвестную величину m. Подберем такую случайную величину ξ, чтобы M(ξ) = m и D(ξ) = b^2. Рассмотрим n случайных величин $\xi_1, \xi_2, \xi_3, \ldots, \xi_n$ распределение которых совпадает с распределением ξ. Если n достаточно велико, то согласно центральной предельной теореме распределение суммы $\rho_n = \xi_1 + \xi_2 + \ldots + \xi_n$ будет приближенно нормальным с параметрами $a = n \cdot m$; $\sigma^2 = n \cdot b^2$.

Из правила «трёх сигм»

$$P\{a - 3 \cdot \sigma < \xi < a + 3 \cdot \sigma\} = 0{,}997 \qquad (3)$$

следует, что

$$P\{n \cdot m - 3b\sqrt{n} < \rho_n < n \cdot m + 3b\sqrt{n}\} = 0{,}997.$$

Разделим неравенство, стоящее в фигурной скобке, на n и получим эквивалентное неравенство с той же вероятностью:

$$P\left\{m - \frac{3b}{\sqrt{n}} < \frac{\rho_n}{n} < m + \frac{3b}{\sqrt{n}}\right\} = 0{,}997.$$

Это соотношение можно записать в виде

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - m\right| < \frac{3b}{\sqrt{n}}\right\} = 0{,}997. \qquad (4)$$

Соотношение (4) определяет метод расчета m и оценку погрешности. В самом деле, найдем n значений случайной величины ξ. Из выражения (4) видно, что среднее арифметическое этих значений будет приближенно равно m. С вероятностью P = 0,997 ошибка такого приближения не превосходит величины $\frac{3b}{\sqrt{n}}$. Очевидно, эта ошибка стремится к нулю с ростом n, что и требовалось доказать. Решение любой задачи методом статистического моделирования состоит в:

– разработке и построении структурной схемы процесса, выявлении основных взаимосвязей;

– формальном описании процесса;

– моделировании случайных явлений (случайных событий, слу чайных величин, случайных функций), сопровождающих функционирование исследуемой системы;

– моделировании (с использованием данных, полученных на предыдущем этапе) функционирования системы – воспроизведении процесса в соответствии с разработанной структурной схемой и формальным описанием;

– накоплении результатов моделирования, их статистической обработке, анализе и обобщении.

В отличие от описанных ранее математических моделей, результаты которых отражали устойчивое во времени поведение системы, результаты, получаемые при статистическом моделировании, подвержены экспериментальным ошибкам. Это означает, что любое утверждение, касающееся характеристик моделируемой системы, должно основываться на результатах соответствующих статистических проверок.

Экспериментальные ошибки при статистическом моделировании в значительной степени зависят от точности моделирования случайных явлений, сопровождающих функционирование исследуемой системы.

Известно, что при изучении вероятностных систем случайные явления могут интерпретироваться в виде случайных событий, случайных величин и случайных функций. Следовательно, моделирование случайных явлений сводится к моделированию случайных событий, случайных величин и случайных функций. Так как случайные события и случайные функции могут быть представлены через случайные величины, то и моделирование случайных событий и случайных функций производится с помощью случайных величин. В связи с этим рассмотрим сначала способы моделирования случайных величин.

Моделирование систем массового обслуживания с использованием метода Монте-Карло

Рассмотренные аналитические методы анализа СМО исходят из предположения, что входящие и исходящие потоки требований являются простейшими. Зависимости, используемые в этих методах для определения показателей качества обслуживания, справедливы лишь для установившегося режима функционирования СМО. Однако в реальных условиях функционирования СМО имеются переходные режимы, а входящие и исходящие потоки требований являются далеко не простейшими. В этих условиях для оценки качества функционирования систем обслуживания широко используют метод статистических испытаний (метод Монте-Карло). Основой решения задачи исследования функционирования СМО в реальных условиях является статистическое моделирование входящего потока требований и процесса их обслуживания (исходящего потока требований).

Для решения задачи статистического моделирования функционирования СМО должны быть заданы следующие исходные данные:

– описание СМО (тип, параметры, критерии эффективности работы системы);
– параметры закона распределения периодичности поступлений требований в систему;
– параметры закона распределения времени пребывания требования в очереди (для СМО с ожиданием);
– параметры закона распределения времени обслуживания требований в системе.

Решение задачи статистического моделирования функционирования СМО складывается из следующих этапов.

Вырабатывают равномерно распределенное случайное число ξ_i.

Равномерно распределенные случайные числа преобразуют в величины с заданным законом распределения:

– интервал времени между поступлениями требований в систему (Δt_{Ti});

– время ухода заявки из очереди (для СМО с ограниченной длиной очереди);

– длительность времени обслуживания требования каналами (Δt_{Oi})

3. Определяют моменты наступления событий:

– поступление требования на обслуживание;

– уход требования из очереди;

– окончание обслуживания требования в каналах системы.

Моделируют функционирование СМО в целом и накапливают статистические данные о процессе обслуживания.

Устанавливают новый момент поступления требования в систему, и вычислительная процедура повторяется в соответствии с изложенным.

Определяют показатели качества функционирования СМО путем обработки результатов моделирования методами математической статистики.

Методику решения задачи рассмотрим на примере моделирования СМО с отказами.

Пусть система имеет два однотипных канала, работающих с отказами, причем моменты времени окончания обслуживания на первом канале обозначим через t_{1i}, на втором канале — через t_{2i}. Закон распределения интервала времени между смежными поступающими требованиями задан плотностью распределения $f_1(t_T)$. Продолжительность обслуживания также является случайной величиной с плотностью распределения $f_2(t_0)$}.

Процедура решения задачи будет выглядеть следующим образом:

1. Вырабатывают равномерно распределенное случайное чис ло ξ_i.

2. Равномерно распределенное случайное число преобразуют в величины с заданным законом распределения. Определяют реализацию случайного интервала времени (Δt_{Ti}) между поступлениями требований в систему.

3. Вычисляют момент поступления заявки на обслуживание: $t_i=t_{i-1}+\Delta t_{Ti}$.

4. Сравнивают моменты окончания обслуживания предшествующих заявок на первом $t_{1(i-1)}$ и втором $t_{2(i-1)}$ каналах.

5. Сравнивают момент поступления заявки t_i с минимальным моментом окончания обслуживания (допустим, что $t_{1(i-1)} < t_{2(i-1)}$):

а) если $[t_i - t_{1(i-1)}] < 0$, то заявка получает отказ и вырабатывают новый момент поступления заявки описанным способом;

б) если $[t_i - t_{1(i-1)}] >= 0$, то происходит обслуживание.

6. При выполнении условия 5б) определяют время обслуживания i-й заявки на первом канале Δt_{1i}, путем преобразования случай ной величины ξ_i в величину (время обслуживания i-й заявки) с за данным законом распределения.

7. Вычисляют момент окончания обслуживания i-й заявки на первом канале t_{1i}=[$t_{1(i-1)}$ +Δt_{1i}].

8. Устанавливают новый момент поступления заявки, и вычислительная процедура повторяется в соответствии с изложенным.

9. В ходе моделирования СМО накапливаются статистические данные о процессе обслуживания.

10. Определяют показатели качества функционирования системы путем обработки накопленных результатов моделирования методами математической статистики.

ЛЕКЦИЯ 11
Моделирование системы управления запасами

Постановка задачи

Компании, продающей один вид продукции, необходимо определить, какое количество товара она должна иметь в запасе на каждый из последующих n мес. (n — заданный входной параметр). Промежутки времени между возникновением спроса на товар являются независимыми и представлены случайными величинами, имеющими одинаковое распределение, со средним значением 0,1 мес. Объемы спроса D также являются независимыми (они не зависят от того, когда возникает спрос) и одинаково распределенными случайными величинами:

$$D=\begin{cases}1 & \textit{с вероятностью } \dfrac{1}{6};\\ 2 & \textit{с вероятностью } \dfrac{1}{3};\\ 3 & \textit{с вероятностью } \dfrac{1}{3};\\ 4 & \textit{с вероятностью } \dfrac{1}{6}.\end{cases}$$

В начале каждого месяца компания пересматривает уровень запасов и решает, какое количество товара заказать у поставщика. В случае, когда компания заказывает Z единиц товара, она будет нести затраты, равные K + iZ, где K — покупная стоимость, K = 32 доллара; i — дополнительные затраты на единицу заказанного товара, i = 3 доллара. (Если Z= 0, какие-либо затраты отсутствуют.) При оформлении заказа время, необходимое для его доставки (именуемое временем доставки или временем получения заказа), является случайной величиной, равномерно распределенной между 0,5 и 1 мес.

Компания использует постоянную стратегию управления запасами (s, S), чтобы определить, какое количество товаров заказывать, то есть

$$\begin{cases}S-1 & \textit{если } I < s;\\ 0 & \textit{если } I \ge s.\end{cases}$$

где I, S, s — это соответственно уровень запасов в начале месяца, после поступления заказа и критический.

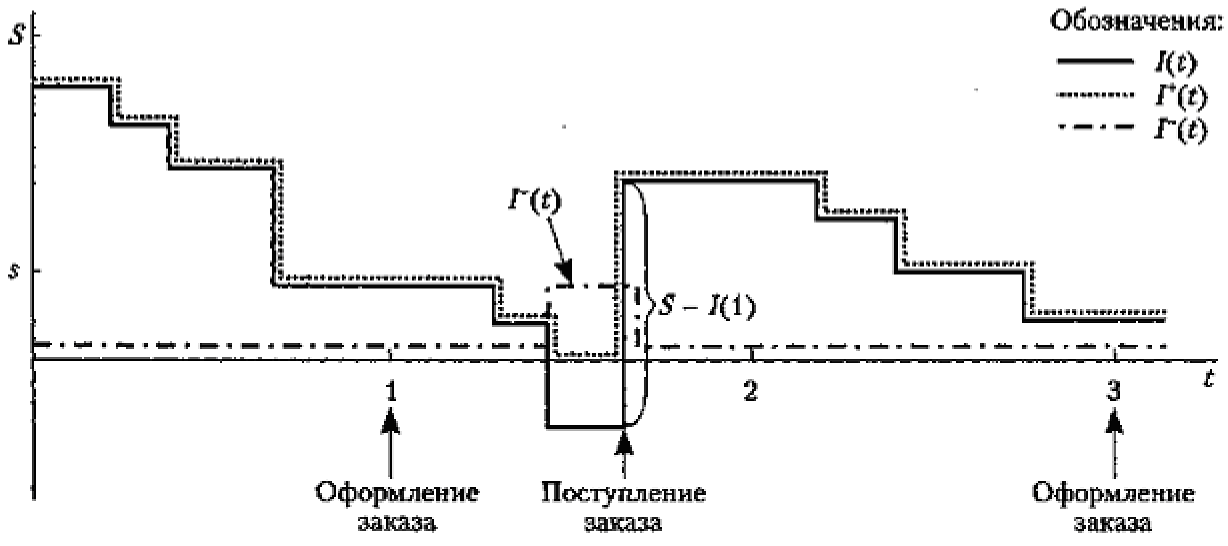
При возникновении спроса на товар он немедленно удовлетворяется, если уровень запасов, по меньшей мере, равен спросу на товар. Если спрос превышает уровень запасов, поставка той части товара, которая превышает спрос над предложением, откладывается и выполняется при будущих поставках. (В этом случае новый уровень запасов равен старому уровню запасов минус объем спроса, что приводит к появлению отрицательного уровня запасов.) При поступлении заказа товар в первую очередь используется для максимально возможного выполнения отложенных поставок (если таковые имеются); остаток заказа (если таковой имеется) добавляется в запасы.

Один тип расходов, возникающий в системе запасов, — это затраты на приобретение заказа. Однако большинство реальных систем управления запасами сталкиваются еще с двумя дополнительными типами расходов — затратами на хранение, а также издержками, связанными с нехваткой товара, к которым мы вернемся, введя некоторые дополнительные обозначения. Пусть I(t) — уровень запасов в момент времени t (обратите внимание, величина I(t) может быть положительной, отрицательной или равняться нулю); I+(t) = max{I(t), 0} — количество товара, имеющегося в наличии в системе запасов на момент времени t (заметьте, что I+(t) > 0); а I-(t) = max{I(t), 0} — количество товара, поставка которого была отложена на момент времени t (I-(t) > 0.) Возможное изменение соотношения I(t), I+(t) и I-(t) показано на рис. 1.13. Моменты времени, когда I(t) уменьшается, соответствуют моментам возникновения спроса.

В нашей модели предполагается, что затраты h на хранение в месяц составляют 1 доллар на единицу товара, имеющегося в (положительных) запасах. Затраты на хранение включают арендную плату за склад, страховки, расходы на обслуживание и налоги, а также скрытые издержки, возникающие, если капитал, вложенный в запасы, мог бы инвестироваться куда-нибудь еще. До сих пор в своих формулировках мы не учитывали тот факт, что некоторые затраты на хранение возникают даже тогда, когда I+(t) = 0, поскольку наша задача — сравнить стратегии осуществления заказов без учета этого фактора, который, по сути, не зависит от используемой стратегии и не повлияет на нашу оценку. Итак, если I+(t) представляет количество товара в запасах на момент времени t, то среднее по времени количество товара, находящегося в запасах в течение n мес., составляет

$$\bar{I}^{+} = \frac{\int_0^n I^{+}(t)\,dt}{n},$$

что подобно определению среднего числа требований, находящихся в очереди в каждый момент времени. Следовательно, средние затраты на хранение в месяц составляют $h\bar{I}^{+}$.



Изменение количества товара $I(t)$, $I^+(t)$ и $I^-(t)$ по времени

Допустим, что издержки π, связанные с отложенными поставками, равны 5 долларам на единицу товара в отложенной поставке за месяц. При этом учитываются издержки на ведение дополнительного учета при невыполнении заказа и урон, наносимый престижу компании. Среднее по времени количество товара в отложенных поставках

$$\overline{I^-} = \frac{\int_0^n I^-(t)dt}{n}$$

Следовательно, средние издержки, образовавшиеся в связи с отложенными поставками, в месяц будут составлять $\pi\overline{I^-}$.

Предположим, что исходный уровень запасов I(0) = 60 и у компании нет неприобретенных заказов. Будем моделировать работу системы в течение n = 120 мес. и воспользуемся показателями средних общих расходов в месяц (которые включают в себя сумму средних затрат на приобретение заказа в месяц, средних затрат на хранение в месяц и средних издержек, связанных с нехваткой товара, в месяц), чтобы сравнить следующие девять стратегий осуществления заказов (s — точка заказа):

s	20	20	20	20	40	40	40	60	60
S	40	60	80	100	60	80	100	80	100

Обратите внимание, что переменными состояния имитационной модели этой системы управления запасами являются уровень запасов I(t), количество товара в невыполненном заказе, направленном от компании к поставщику, и время последнего события, которое необходимо для вычисления площади под функциями I+ (t) и I- (t) .

Организация и логика программы

В рассматриваемой модели системы управления запасами используются следующие типы событий.

Событие	**Тип события**

Поступление заказа от поставщика в компанию	1
Возникновение спроса на товар со стороны покупателя	2
Завершение моделирования через n мес.	3
Оценка запасов (и возможный заказ товаров) в начале месяца	4

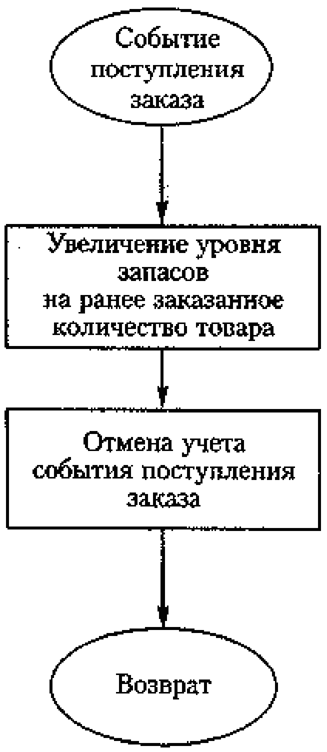
Завершение моделирования сделаем событие типа 3, а не 4, поскольку на момент времени 120 будут запланированы как событие завершения моделирования, так и событие оценки запасов, а задача состоит в том, чтобы в это время первым было выполнено именно событие завершения моделирования. (В связи с тем, что моделирование закончится в момент времени 120, нет смысла оценивать запасы и возможные объемы заказов, соответственно нести издержки, относящиеся к заказу, который никогда не прибудет.) Событие типа 3 всегда будет наступать прежде, чем событие типа 4, так как в случае планирования двух типов событий на одно время синхронизирующие процедуры (в обоих языках) отдают преимущество типу события с меньшим номером. Имитационная модель, по сути, должна быть разработана так, чтобы события обрабатывались в соответствующем порядке при возникновении временных связей.

Для моделирования этой системы нужны три типа случайных величин. Промежутки между возникновениями спроса распределены экспоненциально. Случайная величина спроса D должна быть дискретной (как уже описывалось раньше), и может быть генерирована следующим образом. Вначале необходимо поделить единичный интервал на смежные подынтервалы C1 = [0,1/6), C2= [1/6,1/2), C3 = [1/2,5/6) и C4 = [5/6,1) и от генератора случайных чисел получить случайную величину U с U(0,1). Если U попадает в интервал C1 возвращаем D = 1; если U попадает в интервал C2 возвращаем D = 2 и т. д. Так как ширина C1 равна 1/6 - 0 = = 1/6, а U равномерно распределена между интервалом [0,1], вероятность попадания U в интервал C1 (и получения результата D = 1) составляет 1/6; это согласуется с искомой вероятностью для D = 1. Аналогичным образом мы возвращаем D =2: если U попадает в интервал C2, вероятность такого попадания равна ширине C2, то есть 1/2 - 1/6 = 1/3, как и требовалось. То же самое касается и других интервалов. Все подпрограммы, применяемые для генерирования величин спроса, используют этот принцип и в качестве входных величин принимают разделяющие точки, определяющие подыинтервалы, которые являются интегральными вероятностями распределения D.

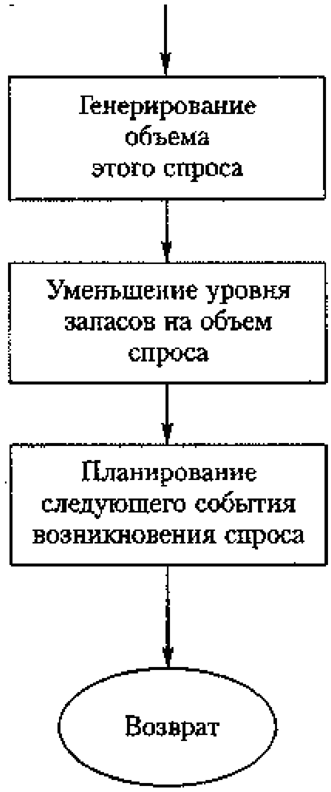
Время доставки равномерно распределено, но не в единичном интервале [0,1]. По сути, мы можем генерировать случайную величину, равномерно распределенную в любом интервале [a, b], сгенерировав случайное число U с U(0, 1), а затем возвратив a + U(b - a).

Из четырех событий в действительности только три вызывают изменения состояния системы (исключением является событие завершения моделирования). Поскольку их логика не зависит от языка программирования, используемого для моделирования, опишем ее в этом разделе.

Событие поступления заказа должно вносить изменения, возникающие при доставке (ранее оформленного) заказа от поставщика. Уровень запасов увеличивается на число товаров в заказе, а событие поступления заказа должно быть исключено из рассмотрения. При событии возникновения спроса обрабатываются изменения, необходимые для его представления.



Блок-схема программы обработки события поступления заказа для модели системы управления запасами



Блок-схема программы обработки события возникновения спроса для модели системы управления запасами

При этом генерируется величина спроса, а уровень запасов уменьшается на полученную величину. И в конечном итоге в списке событий планируется время следующего возникновения спроса. Обратите внимание, в этом месте уровень запасов может стать отрицательным. Блок-схема события оценки запасов, происходящего в начале каждого месяца, приведена на рис.

Блок-схема программы обработки события оценки запасов для модели системы управления запасами

Если уровень запасов I(t) на время оценки составляет, по меньшей мере, s, заказ не размещается, и происходит лишь планирование следующего события оценки запасов в списке событий. Однако, если I(t) < s, потребуется поместить заказ на [S - I(t)] товаров. Для этого количество заказанного товара [S - I(t)] сохраняется до тех пор, пока не прибудет заказ и не будет запланировано время поступления заказа. В этом случае мы планируем следующим событием оценку запасов.

Как и в модели системы массового обслуживания, так и в модели системы управления запасами удобнее отдельно выделить программу для обновления накопителей статистики непрерывного времени, хотя в данной модели это сделать труднее. Блок-схема такой программы приведена на рис. Основной вопрос в данном случае заключается в следующем: нужно ли нам обновлять площадь подфункциями I+(t) и I-(t). Если в результате последнего события уровень запасов стал отрицательным, значит, в системе есть отложенные поставки, следовательно, должна быть обновлена только площадь под функцией I-(t). Если уровень запасов положительный, нужно обновить только площадь под функцией I+(t). Если же уровень запасов равен 0 (что возможно), обновление не требуется. В коде этой программы на обоих языках значение переменной времени последнего события также меняется на текущее время. Указанная программа будет вызываться из основной программы сразу после того, как синхронизирующая программа возвратит управление, независимо от типа события, а также от того, действительно ли уровень запасов изменился на данный момент. Таким образом, обеспечивается простой (хотя и не самый эффективный в плане вычислений) способ обновления интегралов статистики непрерывного времени.

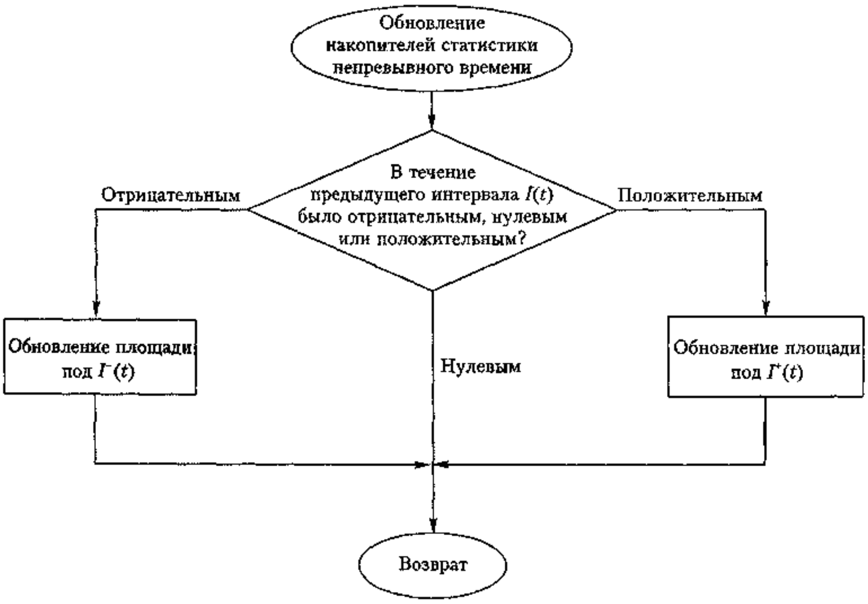


Рис. Блок-схема программы обновления накопителей статистики непрерывного времени для модели системы управления запасами

ЛЕКЦИЯ 12
Транспортные задачи линейного программирования

Постановка задачи

Под термином «транспортные задачи» понимается широкий круг задач не только транспортного характера. Общим для них является, как правило, распределение ресурсов, находящихся у *m* производителей (поставщиков), по *n* потребителям этих ресурсов.

На автомобильном транспорте наиболее часто встречаются следующие задачи, относящиеся к транспортным:

- прикрепление потребителей ресурса к производителям;
- привязка пунктов отправления к пунктам назначения;
- взаимная привязка грузопотоков прямого и обратного направлений;
- отдельные задачи оптимальной загрузки промышленного оборудования;
- оптимальное распределение объемов выпуска промышленной продукции между заводами-изготовителями и др.

Рассмотрим экономико-математическую модель прикрепления пунктов отправления к пунктам назначения. Имеются *m* пунктов отправления груза и объемы отправления по каждому пункту $a_1, a_2,\ldots a_m$. Известна потребность в грузах $b_1, b_2,\ldots b_n$ по каждому из *n* пунктов назначения. Задана матрица стоимостей доставки по каждому варианту c_{ij}, $i = \overline{1,m}$, $j = \overline{1,n}$. Необходимо рассчитать оптимальный план перевозок, т. е. определить, сколько груза должно быть отправлено из каждого *i–го* пункта отправления (от поставщика) в каждый *j-й* пункт назначения (до потребителя) x_{ij} с минимальными транспортными издержками. В общем виде исходные данные представлены в табл. 7.1.

Таблица 7.1 – Исходные данные

Потребители / Поставщики	B_1	B_2	…	B_n	Запасы (объемы отправления)
A_1	c_{11} x_{11}	c_{12} x_{12}	…	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	…	c_{2n} x_{2n}	a_2
…	…	…	…	…	⋮
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	…	c_{mn} x_{mn}	a_m
Потребность	b_1	b_2	…	b_n	

Транспортная задача называется закрытой, если суммарный объем отправляемых грузов $\sum_{i=1}^{m} a_i$ равен суммарному объему потребности в этих грузах по пунктам назначения $\sum_{j=1}^{n} b_j$:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{1}$$

Если такого равенства нет (потребности выше запасов или наоборот), задачу называют открытой, т. е.:

$$\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j. \tag{2}$$

Для написания модели необходимо все условия (ограничения) и целевую функцию представить в виде математических уравнений. Все грузы из *i-х* пунктов должны быть отправлены, т. е.:

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = \overline{1,m}. \tag{3}$$

Все *j-е* пункты (потребители) должны быть обеспечены грузами в плановом объеме:

$$\sum_{i=1}^{m} x_{ij} = b_j, \quad j = \overline{1,n}. \tag{4}$$

Суммарные объемы отправления должны равняться суммарным объемам назначения:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{5}$$

Должно выполняться условие неотрицательности переменных: $x_{ij} \geq 0$, , $i = \overline{1,m}$, $j = \overline{1,n}$. Перевозки необходимо осуществить с минимальными транспортными издержками (функция цели):

$$Z_{\min} = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}. \tag{6}$$

В модели (3) — (6) вместо матрицы стоимостей перевозок (C_{ij}) могут задаваться матрицы расстояний. В таком случае в качестве целевой функции рассматривается минимум суммарной транспортной работы. Как видно из выражения (5), уравнение баланса является обязательным условием решения транспортной задачи. Поэтому, когда в исходных условиях дана открытая задача, то ее необходимо привести к закрытой форме. В случае если

- потребности по пунктам назначения превышают запасы пунктов отправления, то вводится фиктивный поставщик с недостающим объемом отправления;
- запасы поставщиков превышают потребности потребителей, то вводится фиктивный потребитель с необходимым объемом потребления.

Варианты, связывающие фиктивные пункты с реальными, имеют нулевые оценки. После введения фиктивных пунктов задача решается как закрытая.

Транспортным задачам присущи следующие особенности:

- распределению подлежат однородные ресурсы;
- условия задачи описываются только уравнениями;
- все переменные выражаются в одинаковых единицах измерения;
- во всех уравнениях коэффициенты при неизвестных равны единице;
- каждая неизвестная встречается только в двух уравнениях системы ограничений.

Транспортные задачи могут решаться симплекс-методом. Однако перечисленные особенности позволяют для транспортных задач применять более простые методы решения.

Алгоритм метода потенциалов

Наиболее распространенным методом решения транспортных задач является метод потенциалов.

Решение задачи методом потенциалов включает следующие этапы:

1. разработку начального плана (опорного решения);
2. расчет потенциалов;
3. проверку плана на оптимальность;
4. поиск максимального звена неоптимальности (если условие п. 3 не было достигнуто);
5. составление контура перераспределения ресурсов;
6. определение минимального элемента в контуре перераспределения и перераспределение ресурсов по контуру;
7. получение нового плана.

Описанная процедура повторяется несколько раз (итераций), пока не будет найдено оптимальное решение. Вычислительный алгоритм для каждой итерации не меняется.

Для транспортной задачи существует несколько методов отыскания начального плана (опорного решения):

– метод северо-западного угла;
– метод минимальной стоимости;
– метод двойного предпочтения и т. д.

Вычислительный алгоритм метода потенциалов рассмотрим на примере решения конкретной задачи прикрепления пунктов отправления $i = \overline{1,3}$ к пунктам назначения $j = \overline{1,4}$. В соответствии с принятыми обозначениями исходные данные задачи приведены в табл. 7.2

Таблица 7.2 – Исходные данные задачи

Потребители / Поставщики	B_1	B_2	B_3	B_4	Запасы
A_1	1	2	3	4	60
A_2	4	3	2	0	80
A_3	0	2	2	1	100
Потребность	40	60	80	60	240

Начальный план можно составить одним из перечисленных выше методов. Воспользуемся наиболее простым методом — методом северо-западного угла. В соответствии с этим методом загрузка клеток (распределение объемов пунктов отправления по пунктам назначения) начинается с верхней левой клетки («северо-западная» часть таблицы) и продолжается вниз и вправо (по диагонали).

По указанному правилу загружаем первую клетку (i-j) = (1-1) на основании следующего условия:

$$x_{11} = \min\{a_1;\ b_1\} = \min\{60;\ 40\} = 40.$$

Таким образом, первый пункт назначения загружен, а первый пункт отправления имеет остатки груза Δa_1 = 60 - 40 = 20, которые и распределяем на второй пункт назначения:

$$x_{12} = \min\{\Delta a_1;\ b_2\} = \min\{20;\ 60\} = 20;\ \Delta b_2 = 40.$$

Продолжая преобразования аналогичным образом, получаем:

$$x_{22} = \min\{a_2;\ \Delta b_2\} = \min\{80;\ 40\} = 40;\ \Delta b_2 = 40 \text{ и т. д.}$$

Результаты начального плана и расчета потенциалов представлены в табл. 8.3.

Таблица 7.3 – Начальный план перевозок

Потребители / Поставщики	B_1	B_2	B_3	B_4	Запасы	α_i
A_1	1 Р 40	2 20 З	3	4	60	0
A_2	4	3 40 Р	2 З 40	0	80	1
A_3	0 ⊕ З	2	2 Р 40	1 60	100	1
Потребность	40	60	80	60	240	
β_j	1	2	1	0		

В процессе решения после каждой итерации (в том числе и после получения допустимого решения) по загруженным клеткам проверяется выполнение следующего условия:

$$N=m+n-1 \tag{7}$$

В нашем примере m = 3, n=4, а число загруженных клеток равно 6, т. е. соответствует условию (7): N=3 + 4-1=6. Если условие (7) не выполняется, план называется вырожденным. В этом случае в любые свободные клетки надо поставить столько нулей, чтобы с их учетом выполнялось условие (7). Клетка, в которой стоит ноль, считается занятой. Значение целевой функции по результатам расчета допустимого плана

$$Z=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}=1\cdot 40+2\cdot 20+3\cdot 40+2\cdot 40+2\cdot 40+1\cdot 60=420 \text{ д. е.}$$

Расчет потенциалов выполняют по загруженным клеткам, для которых должно выполняться следующее равенство:

$$\alpha_i+\beta_j=c_{ij}, \tag{8}$$

где α_i, – потенциал i-и строки; β_j — потенциал j-го столбца.

Вычисляя потенциалы по выражению (8), принимаем для первой строки

α_1 = 0. Используя загруженные клетки (i-j) = (1-1), (1 - 2), получаем:

$$\alpha_1+\beta_1=c_{11}=0+\beta_1=1;\ \beta_1=1;$$
$$\alpha_1+\beta_2=c_{12}=0+\beta_2=2;\ \beta_2=2.$$

Далее по загруженным клеткам (2 — 2), (2 — 3) определяем другие потенциалы:

$$\alpha_2+\beta_2=3;\ \alpha_2+2=3;\ \alpha_2=1;$$
$$\alpha_2+\beta_3=2;\ 1+\beta_3=2;\ \beta_3=1.$$

Проверяем план на оптимальность по незагруженным клеткам, используя следующее неравенство:

$$\alpha_i+\beta_j\le c_{ij}. \tag{9}$$

Если для незагруженных клеток условие (9) выполняется, то план — оптимальный. По табл. 3 осуществляем проверку начального плана на оптимальность:

$$(i-j)=(1\text{–}3),\quad 0+1\le 3;$$
$$(i-j)=(1\text{–}4),\quad 0+0\le 4;$$
$$(i-j)=(2\text{–}1),\quad 1+1\le 4;$$
$$(i-j)=(2\text{–}4),\quad 1+0>0,\quad \Delta c_{24}=1;$$
$$(i-j)=(3\text{–}1),\quad 1+1>0,\quad \Delta c_{31}=2;$$
$$(i-j)=(3\text{–}2),\quad 1+2>2,\quad \Delta c_{32}=1.$$

Итак, по трем клеткам условие (9) не выполняется, следовательно, начальный план требует улучшения. Характеристики Δc_{ij} показывают размер экономии транспортных издержек на 1 ед. перевозимого груза. В нашем примере наибольшую экономию можно получить по клетке (i -j) = (3-1), где Δc_{31} = 2 > { Δc_{24}; Δc_{32}}. Следовательно, клетку (3-1) необходимо загрузить за счет перераспределения ресурсов из других загруженных клеток. В табл. 3 клетку (3-1) помечаем знаком «+», так как здесь в начальном плане находится вершина максимальной неоптимальности.

Контур перераспределения ресурсов составляют по следующим правилам:

– этот контур представляет замкнутый многоугольник с вершина
ми в загруженных клетках, за исключением клетки с вершиной максимальной неоптимальности «+», и звеньями, лежащими вдоль строк и столбцов матрицы;

– ломаная линия должна быть связанной в том смысле, что из любой ее вершины можно попасть в любую другую вершину по звеньям ломаной цепи (по строке или по столбцу);

– в каждой вершине контура встречаются только два звена, одно из них располагается по строке, другое — по столбцу;

– число вершин контура четное, все они в процессе перераспределения делятся на загружаемые и разгружаемые;

– в каждой строке (столбце) имеются две вершины: одна — загружаемая, другая — разгружаемая.

В этой клетке намечаем одну из вершин контура и далее по вышеизложенным правилам строим контур, вершины которого будут находиться в клетках (3—1) — (1—1) — (1—2) — (2—2) — (2—3) — (3—3). Вершины контура последовательно подразделяем на загружаемые (З) и разгружаемые (Р), начиная с вершины максимальной неоптимальности «+» (табл. 8.3).

Перераспределение ресурсов по контуру осуществляется с целью получения оптимального плана. В процессе перераспределения ресурсов по контуру в соответствии с условием неотрицательности переменных x_{ij} ни одно из этих значений не должно превратиться в отрицательное число. Поэтому анализируют только клетки, помеченные знаком Р, из которых выбирают клетку с минимальным объемом перевозок. В нашем примере $X_{min} = min\{40;\ 40;\ 40\} = 40$. Следовательно, клетки (1 — 1), (2—2), (3—3) полностью разгружаются. В клетке (1—2) загрузка увеличится на 40 и достигнет 60, в клетке (2—3) загрузка составит 40 + 40 = 80, и клетка (3—1) загрузится на 40 (табл. 4).

Проверяем условие N = m + n - 1. В нашем примере m = 3, n =4, а число загруженных клеток равно 4, т. е. условие не выполняется и 6 ≠ 4. В процессе перераспределения ресурсов произошла полная разгрузка трех клеток, а мы должны освободить только одну клетку. В этом случае следует в две клетки проставить нули (нулевой ресурс) и считать условно их загруженными. Например, в клетки (1 — 1) и (3—3) проставим нулевой ресурс (табл. 14.4). Получение нового плана (итерации) осуществляется в том же порядке, который был рассмотрен выше, т. е.

– по загруженным клеткам (в соответствии с новой загрузкой) вычисляются потенциалы αi, и βj ;

– по незагруженным клеткам производится проверка плана на оптимальность;

– находится вершина максимальной неоптимальности и строится новый контур перераспределения, и т. д., до тех пор, пока не будет найдено оптимальное решение, удовлетворяющее неравенству (9).

Таблица 7.4 – Первый план перевозок

Поставщики \ Потребители	B_1	B_2	B_3	B_4	Запасы	α_i
A_1	1 0	2 60	3	4	60	0
A_2	4	3	2 80 P	⊕ 0 3	80	–1
A_3	0 40	2	2 0 З	1 Р 60	100	–1
Потребность	40	60	80	60	240	
β_j	1	2	3	2		

По результатам первой итерации имеем

$$Z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} = 2\cdot 60 + 2\cdot 80 + 1\cdot 60 + 0\cdot 40 = 340.$$

Ниже приведены расчеты по второй итерации и оптимальный план.

Поиск потенциалов следующий:

$\alpha_1 + \beta_1 = 1;$ $\quad 0 + \beta_1 = 1;$ $\quad \beta_1 = 1;$

$\alpha_1 + \beta_2 = 2;$ $\quad 0 + \beta_2 = 2;$ $\quad \beta_2 = 2;$

$\alpha_2 + \beta_3 = 2;$ $\quad \alpha_2 + 3 = 2;$ $\quad \alpha_2 = -1;$

$\alpha_3 + \beta_1 = 0;$ $\quad \alpha_3 + 1 = 0;$ $\quad \alpha_3 = -1;$

$\alpha_3 + \beta_3 = 2;$ $\quad -1 + \beta_3 = 2;$ $\quad \beta_3 = 3;$

$\alpha_3 + \beta_4 = 1;$ $\quad -1 + \beta_4 = 1;$ $\quad \beta_4 = 2.$

Проведем проверку на оптимальность:

$(i - j) = (1–3);\ 0 + 3 \le 3;$

$(i - j) = (1–4);\ 0 + 2 < 4;$

$(i - j) = (2–1);\ 1 - 1 < 4;$

$(i - j) = (2–2);\ 2 - 1 < 3;$

$(i - j) = (3–2);\ 2 - 1 < 2;$

$(i - j) = (2–4);\ 2 - 1 > 0.$

Клетку (2-4) необходимо загрузить.

В соответствии с перераспределением ресурсов по контуру получаем табл. 7.5, для которой вновь рассчитываем потенциалы α_i, и β_j, и последовательность вычислений повторяется.

Таблица 7.5 – Оптимальный план перевозок

Поставщики \ Потребители	B_1	B_2	B_3	B_4	Запасы	α_i
A_1	1 0	2 60	3	4	60	0
A_2	4	3	2 20	0 60	80	–1
A_3	0 40	2	2 60	1	100	–1
Потребность	40	60	80	60	240	
β_j	1	2	3	1		

Для распределения, полученного в табл. 7.5, условие $\alpha_i + \beta_j \le c_{ij}$. выполняется, следовательно, план — оптимальный.

Транспортные издержки по оптимальному плану следующие:

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} = 1 \cdot 0 + 2 \cdot 60 + 2 \cdot 20 + 0 \cdot 60 + 0 \cdot 40 + 2 \cdot 60 = 280 \text{ д. е.}$$

Таким образом, построением начального плана с последующим расчетом двух итераций получено оптимальное решение по прикреплению пунктов отправления грузов к пунктам назначения.

ЛЕКЦИЯ 13-14
Теория принятия решений

Основные понятия

Рассмотренные задачи линейного программирования формулировались и решались в предположении наличия *полной* информации. Их можно отнести к совокупности задач принятия решений в условиях определенности. В реальных экономических условиях приходится решать отдельные задачи при ограниченности, неточности исходной информации о самом объекте и внешней среде, в которой он функционирует и развивается.

При принятии управленческих решений о функционировании и развитии экономического объекта необходимо учитывать важную характеристику внешней среды — неопределенность.

Под *неопределенностью* следует понимать отсутствие, неполноту, недостаточность информации об объекте, процессе, явлении или неуверенность в достоверности информации. В условиях рыночной экономики существует множество источников возникновения неопределенности для различных экономических объектов. Например, *к основным источникам возникновения неопределенности на транспорте можно отнести следующие.*

1. Существенная зависимость транспортного процесса от по годных условий. Например, погодные условия могут вызвать не предвиденные последствия в перевозках сельскохозяйственной продукции.
2. Наличие, кроме транспортного предприятия, других участников транспортного процесса - поставщиков грузов, потребите лей грузов, ГАИ и др. Результат их влияния на транспортный процесс носит неопределенный и неоднозначный характер.
3. Наличие в работе автотранспорта элементов вероятности и случайности (надежность подвижного состава, неравномерность спроса на транспортные услуги во времени и др.).

Недостаточность, неполнота информации об объекте, процессе, явлении, по отношению к которому принимается решение: ограниченность в сборе и обработке информации, постоянная ее изменчивость.

5. Наличие в общественной жизни страны противоборствующих тенденций, столкновение противоречивых интересов.
6. Невозможность однозначной оценки объекта при сложившихся в данных условиях уровне и методах научного познания.
7. Относительная ограниченность сознательной деятельности лица, принимающего решение, существующие различия в социально-психологических установках, идеалах, намерениях, оценках, стереотипах поведения.

Неопределенность обусловливает появление ситуаций, не имеющих однозначного исхода (решения). Среди различных видов ситуаций, с которыми в процессе производства сталкиваются предприятия, особое место занимают ситуации риска.

Под *ситуацией* ***риска*** следует понимать сочетание, совокупность различных обстоятельств и условий, создающих обстановку того или иного вида деятельности. Ей сопутствуют *три условия*. Это

- наличие неопределенности;
- необходимость выбора альтернативы (отказ от выбора таковых является разновидностью альтернативы);
- возможность оценить вероятность осуществления выбираемых альтернатив.

Таким образом, если существует возможность количественно и качественно определить степень вероятности того или иного варианта, то это и будет ситуация риска.

Для того чтобы снять ситуацию риска, руководители предприятий вынуждены принимать решения и стремиться реализовать их. Этот процесс находит свое выражение в понятии «риск». Несмотря на то что риск объективно присутствует во всех сферах общественной жизни и в большинстве видов управленческой деятельности, обнаруживается, что понятие «риск» до сих пор не получило универсальной трактовки.

Следует упомянуть об экономическом риске применительно к процессам принятия решений в условиях неопределенности и риска, иными словами, в условиях дефицита информации или неуверенности в достоверности информации. В этом случае риск предстает в виде совокупности вероятных экономических, политических, нравственных и других положительных и неблагоприятных последствий, которые могут наступить при реализации выбранных решений. Определим риск как целенаправленные действия, в ходе которых имеется возможность количественно и качественно оценить вероятность достижения желаемого результата, неудачи и отклонения от цели (положительного или отрицательного свойства).

Процесс установления рыночных отношений в нашей стране порождает различные виды рисковых ситуаций, более того, в работе предприятий риск становится необходимым и обязательным его компонентом.

Чтобы проиллюстрировать различие между ситуациями, когда приходится принимать решения в условиях риска или в условиях неопределенности, рассмотрим задачу оптимального выбора ассортимента выпускаемой продукции.

В условиях риска доход c_j от реализации единицы продукции j не является фиксированной величиной. Напротив, это случайная величина, точное числовое значение которой не известно, но описывается с помощью функции распределения $f(C_j)$. Часть дохода C_jX_j, определяемая продукцией j, также случайная величина, если даже значение переменной x_j, определяющей уровень выпуска продукции j, задано.

В условиях неопределенности функция распределения $f_j(c)$ неизвестна. В действительности неопределенность не означает полного отсутствия информации о задаче. Например, известно, что c_j может принимать пять значений, но неизвестны вероятности этих значений. Эта ситуация рассматривается как принятие решений в условиях неопределенности.

Таким образом, с точки зрения полноты исходных данных определенность и неопределенность представляют два крайних случая, а риск определяет промежуточную ситуацию, в которой приходится принимать решение.

Степень неинформированности данных определяет, каким образом задача формализуется и решается.

При решении задач в условиях неопределенности внешней среды наиболее часто возникают две ситуации. При первой ситуации сама система препятствует принятию решений, например задача составления графика выпуска на работу подвижного состава, занимающегося перевозкой сельхозпродукции, в зависимости от того, будет дождь или нет. В этой задаче природа будет восприниматься как «доброжелательный» противник.

Во второй ситуации возможно наличие конкуренции, когда два (или более) участника находятся в конфликте и каждый стремится как можно больше выиграть у другого (других). Эта ситуация отличается от обычных процессов принятия решений в условиях неопределенности тем, что лицу, принимающему решение, противостоит мыслящий противник. Теория, в которой рассматриваются задачи принятия решений в условиях неопределенности при наличии противника («доброжелательного» или мыслящего), известна как теория игр.

Принятие решений в условиях полной определенности

Математические модели исследуемых явлений или процессов могут быть заданы в виде таблиц, элементами которых являются значения частных критериев эффективности функционирования системы, вычисленные для каждой из сравниваемых стратегий при строго заданных внешних условиях. Для рассматриваемых условий принятие решений может производиться:

– по одному критерию;

– по нескольким критериям.

Пример 1. Одной из фирм требуется выбрать оптимальную стратегию по обеспечению нового производства оборудованием. С помощью экспериментальных наблюдений были определены значения частных критериев функционирования соответствующего оборудования (a_{ij}), выпускаемого тремя заводами-изготовителями. Рассмотрим данные для выбора оптимальной стратегии в условиях полной определенности:

На основе экспертных оценок были также определены веса частных критериев $\lambda_j, j=1,4$

$$\lambda_1 = 0{,}4; \quad \lambda_2 = 0{,}2; \quad \lambda_3 = 0{,}1; \quad \lambda_4 = 0{,}3.$$

Очевидно, выбор оптимальной стратегии (варианта оборудования) по одному критерию в данной задаче не вызывает затруднений. Например, если оценивать оборудование по надежности, то лучшим является оборудование завода 1 (стратегия x_1).

Таблица 8.1

Варианты оборудования (стратегии, решения)	Частные критерии эффективности оборудования*			
	производительность, д. е.	стоимость оборудования, д. е.	энергоемкость, у. е.	надежность, у. е.
Оборудование завода 1, x_1	$a_{11} = 5$	$a_{12} = 7$	$a_{13} = 5$	$a_{14} = 6$
Оборудование завода 2, x_2	$a_{21} = 3$	$a_{22} = 4$	$a_{23} = 7$	$a_{24} = 3$
Оборудование завода 3, x_3	$a_{31} = 4$	$a_{32} = 6$	$a_{33} = 2$	$a_{34} = 4$

* Значения частных критериев даны в условных единицах.

Выбор оптимального решения по комплексу нескольких критериев (в нашем примере — по четырем критериям) является задачей многокритериальной.

Один из подходов к решению многокритериальных задач управления связан с процедурой образования обобщенной функции F_i (a_{i1}; a_{i2}; a_{i3}; ... a_{in}), монотонно зависящей от критериев a_{i1}; a_{i2}; a_{i3} ... a_{in}. Данная процедура называется процедурой (методом) свертывания критериев. Существует несколько методов свертывания, например:

- метод аддитивной оптимизации;
- метод многоцелевой оптимизации и др.

Рассмотрим подробнее метод аддитивной оптимизации. Пусть

$$F_i(a_{ij}) = \sum_{j=1}^{n} \lambda_j \cdot a_{ij}. \tag{1}$$

Здесь выражение (1) определяет аддитивный критерий оптимальности. Величины λ_i являются весовыми коэффициентами, которые определяют в количественной форме степень предпочтения j-го критерия по сравнению с другими критериями. Другими словами, коэффициенты λ_j определяют важность j-го критерия оптимальности. При этом более важному критерию приписывается больший вес, а общая важность всех критериев равна единице, т. е.

$$\sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = \overline{1, n}. \tag{2}$$

Обобщенная функция цели (1) может быть использована для свертывания частных критериев оптимальности, если:

♦ частные (локальные) критерии количественно соизмеримы по важности, т. е. каждому из них можно поставить в соответствие некоторое число λ_i, которое численно характеризует его важность по отношению к другим критериям;

♦ частные критерии являются однородными (имеют одинаковую размерность; в нашем примере критерии «стоимость оборудования» и «производительность оборудования» в условных денежных единицах будут однородными).

В этом случае для решения задачи многокритериальной оптимизации оказывается справедливым применение аддитивного критерия оптимальности.

Допустим, в примере 1 необходимо выбрать оптимальный вариант оборудования по двум однородным локальным критериям:

♦ производительность (д. е.);

♦ стоимость оборудования (д. е.).

На основе экспертных оценок были определены весовые коэффициенты этих двух частных критериев: λ_1 = 0,667, λ_2 = 0,333. Вычислим аддитивный критерий оптимальности для трех вариантов:

$$F_1(a_{1j}) = \lambda_1 \cdot a_{11} + \lambda_2 \cdot a_{12} = 0{,}667 \cdot 5 + 0{,}333 \cdot 7 = 5{,}666;$$
$$F_2(a_{2j}) = \lambda_1 \cdot a_{21} + \lambda_2 \cdot a_{22} = 0{,}667 \cdot 3 + 0{,}333 \cdot 4 = 3{,}333;$$
$$F_3(a_{3j}) = \lambda_1 \cdot a_{31} + \lambda_2 \cdot a_{32} = 0{,}667 \cdot 4 + 0{,}333 \cdot 6 = 4{,}666.$$

Очевидно, первый вариант оборудования по двум частным стоимостным критериям будет оптимальным, так как $F_{max} = F_1(a_{1j})$ = 5,666. В примере 1 четыре локальных критерия не однородны, т.е. имеют различные единицы измерения. В этом случае требуется нормализация критериев. Под нормализацией критериев понимается такая последовательность процедур, с помощью которой все критерии приводятся к единому, безразмерному масштабу измерения. К настоящему времени разработано большое количество схем нормализации. Рассмотрим некоторые из них.

Определим максимум и минимум каждого локального критерия, т. е.

$$a_j^+ = \max a_{ij},\ i = \overline{1,m};$$
$$a_j^- = \min a_{ij},\ i = \overline{1,m}. \qquad (34)$$

Выделим группу критериев $a_j, j = \overline{1,l}$, которые максимизируются при решении задачи, и группу критериев $a_j, j = \overline{l+1,n}$, которые минимизируются при решении задачи.

Тогда в соответствии с принципом максимальной эффективности нормализованные критерии определяются из следующих соотношений:

$$\hat{a}_{ij} = \frac{a_{ij}}{a_j^+},\quad j = \overline{1,\ell}; \qquad (5)$$

$$\hat{a}_{ij} = 1 - \frac{a_{ij}}{a_j^+},\quad j = \overline{\ell+1,n} \qquad (6)$$

$$\hat{a}_{ij} = \frac{a_{ij} - a_j^-}{a_j^+ - a_j^-},\quad j = \overline{1,\ell}; \qquad (7)$$

$$\hat{a}_{ij} = \frac{a_j^+ - a_{ij}}{a_j^+ - a_j^-},\ j = \overline{\ell+1,n}. \qquad (8)$$

Оптимальным будет тот вариант (стратегия), который обеспечивает максимальное значение функции цели:

$$F_i = \sum_{j=1}^{n} \lambda_j \cdot \hat{a}_{ij}, \quad i = \overline{1,m}. \tag{9}$$

В соответствии с принципом минимальной потери нормализованные критерии определяются из соотношений

$$\hat{a}_{ij} = 1 - \frac{a_{ij}}{a_j^+}, \quad j = \overline{1,\ell}; \tag{10}$$

$$\hat{a}_{ij} = \frac{a_{ij}}{a_j^+}, \quad j = \overline{\ell+1,n} \tag{11}$$

$$\hat{a}_{ij} = \frac{a_j^+ - a_{ij}}{a_j^+ - a_j^-}, \quad j = \overline{1,\ell}; \tag{12}$$

$$\hat{a}_{ij} = \frac{a_{ij} - a_j^-}{a_j^+ - a_j^-}, \quad j = \overline{\ell+1,n}. \tag{13}$$

При этом оптимальным будет тот вариант (стратегия), который обеспечивает минимальное значение функции цели (9).

Пример 2. Используя данные примера 1, определите оптимальную стратегию выбора оборудования из трех возможных (m = 3) с учетом четырех локальных критериев (n = 4).

Решение 1. Определим max и min каждого локального критерия:

$a_1^+ = 5$; $a_2^+ = 7$; $a_3^+ = 7$; $a_4^+ = 6$.

При решении задачи максимизируются первый (производительность) и четвертый (надежность) критерии, а минимизируются второй (стоимость оборудования) и третий (энергоемкость) критерии.

Исходя из принципа максимизации эффективности, нормализуем критерии:

$$\hat{a}_{i1} = \frac{a_{i1}}{a_1^+};$$

$$\hat{a}_{11} = \frac{a_{11}}{a_1^+} = \frac{5}{5} = 1;$$

$$\hat{a}_{21} = \frac{a_{21}}{a_1^+} = \frac{3}{5} = 0{,}6;$$

$$\hat{a}_{31} = \frac{a_{31}}{a_1^+} = \frac{4}{5} = 0{,}8.$$

$$\hat{a}_{24} = \frac{a_{24}}{a_4^+} = \frac{3}{6} = 0{,}5;$$

$$\hat{a}_{i4} = \frac{a_{i4}}{a_4^+};$$

$$\hat{a}_{14} = \frac{a_{14}}{a_4^+} = \frac{6}{6} = 1;$$

$$\hat{a}_{34} = \frac{a_{34}}{a_4^+} = \frac{4}{6} = \frac{2}{3}.$$

$$\hat{a}_{i2} = 1 - \frac{a_{i2}}{a_2^+};$$

$$\hat{a}_{12} = 1 - \frac{a_{12}}{a_2^+} = 1 - \frac{7}{7} = 0;$$

$$\hat{a}_{22} = 1 - \frac{a_{22}}{a_2^+} = 1 - \frac{4}{7} = \frac{3}{7};$$

$$\hat{a}_{32} = 1 - \frac{a_{32}}{a_2^+} = 1 - \frac{6}{7} = \frac{1}{7}.$$

$$\hat{a}_{i3} = 1 - \frac{a_{i3}}{a_3^+};$$

$$\hat{a}_{13} = 1 - \frac{a_{13}}{a_3^+} = 1 - \frac{5}{7} = \frac{2}{7};$$

$$\hat{a}_{23} = 1 - \frac{a_{23}}{a_3^+} = 1 - \frac{7}{7} = 0;$$

$$\hat{a}_{33} = 1 - \frac{a_{33}}{a_3^+} = 1 - \frac{2}{7} = \frac{5}{7}.$$

4. Определим обобщенную функцию цели по каждому варианту:

$$F_1 = \lambda_1 \cdot \hat{a}_{11} + \lambda_2 \cdot \hat{a}_{12} + \lambda_3 \cdot \hat{a}_{13} + \lambda_4 \cdot \hat{a}_{14} =$$

$$= 0{,}4 \cdot 1 + 0{,}2 \cdot 0 + 0{,}1 \cdot \frac{2}{7} + 0{,}3 \cdot 1 \approx 0{,}729;$$

$$F_2 = \lambda_1 \cdot \hat{a}_{21} + \lambda_2 \cdot \hat{a}_{22} + \lambda_3 \cdot \hat{a}_{23} + \lambda_4 \cdot \hat{a}_{24} =$$

$$= 0{,}4 \cdot 0{,}6 + 0{,}2 \cdot \frac{3}{7} + 0{,}1 \cdot 0 + 0{,}3 \cdot 0{,}5 \approx 0{,}476;$$

$$F_3 = \lambda_1 \cdot \hat{a}_{31} + \lambda_2 \cdot \hat{a}_{32} + \lambda_3 \cdot \hat{a}_{33} + \lambda_4 \cdot \hat{a}_{34} =$$

$$= 0{,}4 \cdot 0{,}8 + 0{,}2 \cdot \frac{1}{7} + 0{,}1 \cdot \frac{5}{7} + 0{,}3 \cdot \frac{2}{3} \approx 0{,}603.$$

Оптимальным является первый вариант оборудования, так как $F_{max} = F_1 = 0.729$.

Рассмотренный подход к решению многокритериальных задач зачастую применяется при решении экономических задач, связанных с оценкой качества промышленной продукции и оценкой уровня технического совершенства технических устройств и систем по нескольким показателям.

Другим возможным методом решения многокритериальных задач является метод последовательных уступок. В начале критерии ранжируются и нумеруются в порядке убывания важности. Абсолютное значение коэффициентов важности λ_j на этом этапе не играет никакой роли.

Оптимизируется первый по важности критерий a_1, и определяется его экстремальное значение a_1^*. Затем назначается величина допустимого отклонения критерия от оптимального значения (уступка) Δa_1 и ищется экстремальное значение второго по важности критерия a_2, при условии, что отклонение первого от оптимального значения не превзойдет величины уступки. Затем назначается уступка для второго критерия, и задача оптимизируется по третьему критерию и т. д. Таким образом, многокритериальная задача оптимизации заменяется последовательностью однокритериальных задач. Решение каждой предыдущей задачи используется при решении последующих для формирования дополнительных условий, состоящих в ограничении на величину уступки.

Принятие решений в условиях риска

Основными критериями оценки принимаемых решений в условиях риска являются:

– ожидаемое значение результата;

– ожидаемое значение результата в сочетании с минимизацией его дисперсии;

– известный предельный уровень результата;

– наиболее вероятное событие (исход) в будущем.

Критерий ожидаемого значения используется в случаях, когда требуется определить экстремальное значение (max или min) результативного показателя (прибыль, расходы, экономические потери и т. д.). Применение этого критерия рассмотрим на конкретном примере, связанном с постановкой задачи проведения ремонтно-профилактических воздействий автомобилей. Оптимальное количество ремонтных воздействий, определенное минимизацией суммарных затрат на заданной наработке L_K с учетом рисков пропуска отказов и выполнения лишних ТО, приравнивается к количеству ТО на указанном пробеге. Модель данной задачи является моделью вероятностного спроса на ремонты с мгновенным восстановлением. Здесь **минимизируются суммарные издержки** за пробег L_K, которые определяются затратами на плановый ремонт S_P, профилактику $S_{то}$ и незапланированный аварийный ремонт $S_ш$, рассматриваемый как штраф за пропуск отказа:

$$S = S_р + S_{ТО} + S_ш \rightarrow \min. \qquad (14)$$

Составляющие суммарных затрат формулы (14) зависят от количества ремонтно-профилактических операций за наработку L_K, определяемых по формуле

$$n = \frac{L_к}{L_{от}}, \qquad (15)$$

где L_{OT} — наработка до отказа

Наработка до отказа — величина случайная, определяемая плотностью распределения $f(L_{OT})$, $L_{OT} < L_K$. В силу случайности L_{OT} величина n также будет случайной с плотностью распределения

$$f(n) = \left| -\frac{L_к}{n^2} \right| \cdot f\left(\frac{L_к}{n} \right). \tag{16}$$

Используя f(n) как весовую функцию и выражая составляющие суммарных затрат через соответствующие стоимости из (14), получим

$$S = C_р \cdot n_р + \int_0^{n_р} C_{ТО} \cdot (n_р - n) \cdot f(n) dn + \int_{n_р}^{\infty} C_ш \cdot (n - n_р) \cdot f(n) dn \to \min, \tag{17}$$

где $C_р$ — средняя стоимость предупредительного (планового) ремонта;

$C_{ТО}$ — средняя стоимость профилактики (или убыток от недоиспользования ресурса замененных при ТО деталей);

$C_Ш$ - ущерб (штраф) от пропуска отказа (или стоимость устранения аварийного отказа). Очевидно, $C_ш > C_{то}$.

Интеграл (16) в пределах [0, $n_р$] соответствует риску выполнения лишних ТО (избыточность затрат на ТО), а интеграл в пределах [$n_р$, ∞] — риску пропуска аварийных отказов (избыточность затрат на ТР по потребности). Из уравнения (17) находим оптимальное количество ремонтов $n_р$ на пробеге L_K (обычно L_K — пробег до КР). Далее, заменяя необходимые ремонты обслуживаниями, при которых выполняется комплекс операций по предупреждению отказов, включая предупредительные замены деталей, получим

$$L_{ТО} = L_к / n_р. \tag{18}$$

Пример 3. Определить оптимальную периодичность ТО (у. е.) при L_K = 200 тыс. км, $C_ш$ = 69, $C_р$ = 24, $C_{то}$ = 15, если наработки до отказа имеют нормальное распределение с параметрами L_{OT} -= 20 тыс. км и σl = 5 тыс. км.

$$f(L_{от}) = \frac{1}{\sigma_L \sqrt{2\pi}} \exp\left[-0{,}5 \cdot \left(\frac{L_{от} - \bar{L}_{от}}{\sigma_L} \right)^2 \right]. \tag{19}$$

Решение

Выполнив преобразование распределения (19) по формуле (15), получим (n≥1):

$$f(n) = \frac{L_к}{n^2 \cdot \sigma_L \cdot \sqrt{2\pi}} \cdot \exp\left[-0{,}5 \cdot \left(\frac{L_к/n - \bar{L}_{от}}{\sigma_L} \right)^2 \right]. \tag{20}$$

После подстановки выражения (20) в (17) получим задачу оптимизации. Решая задачу, получим оптимальную периодичность $L_{ТО}$ — 15,3 тыс. км при $n_р$ = 13,08, которая обеспечивает минимальные суммарные издержки S.

Критерий ожидаемого значения позволяет получить достоверные оценки в случае, когда одно и то же решение приходится принимать достаточно большое число раз, так как замена математического ожидания выборочными данными правомерна лишь при большом объеме выборки.

Если необходимость в принятии решения встречается редко, то выборочное значение может значительно отличаться от математического ожидания, а применение критерия ожидаемых значений может приводить к ошибочным результатам. В таких случаях рекомендуется применять критерий ожидаемого значения в сочетании с минимизацией его дисперсии, что приближает выборочное значение к математическому ожиданию. Критерий принимает следующий вид:

$$\begin{cases} M(X) + K \cdot D(X) \to \min \\ M(X) - K \cdot D(X) \to \max, \end{cases} \tag{21}$$

где X - случайная величина (например, суммарные издержки); D(X) - дисперсия этой величины; K — заданная постоянная.

Постоянную K иногда интерпретируют как уровень несклонности к риску. Считается, что K определяет «степень важности» дисперсии D(X) по отношению к M(X). Например, предприниматель, особенно остро реагирующий на большие отрицательные отклонения прибыли вниз от M(X), может выбрать K много больше единицы. Это придает больший вес дисперсии и приводит к решению, уменьшающему большие потери прибыли.

Критерий предельного уровня не позволяет получить оптимальное решение, найти максимум прибыли и минимум расходов. Этот критерий дает возможность определить приемлемый (допустимый) способ действий. Например, транспортная фирма распродает автомобили, бывшие в эксплуатации. По каждой модели автомобиля определенного возраста определяется лимитная цена, т. е. минимально допустимая цена продажи автомобиля. Продажа автомобилей по цене ниже лимитной приведет к убыточной работе транспортной фирмы. Это и есть предельный уровень, позволяющий транспортной фирме согласиться на первое же превышающее этот уровень предложение цены. Такой критерий не определяет оптимальное решение, поскольку одно из последующих предложений может оказаться более выгодным, чем принятое.

Одно из преимуществ критерия предельного уровня заключается в том, что для него нет необходимости задавать в явном виде плотность распределения случайных величин. В нашем примере случайная величина — рыночная цена автомобиля. Транспортная фирма располагает информацией о распределении рыночных цен На подобные автомобили в неявном виде. Иначе при полном отсутствии информации о распределении рыночных цен фирма установила бы предельные цены на автомобили очень высокими или, наоборот, очень низкими.

Критерий наиболее вероятного события (исхода) основан на преобразовании случайной ситуации в детерминированную путем замены случайной величины единственным значением, имеющим наибольшую вероятность реализации.

Учебное издание

Надежда Львовна Леонова

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Конспект лекций

Редактор и корректор В.А. Басова

Технический редактор Л.Я. Титова   Темплан 2015, поз. 2

Подп. к печати   02.2015. Формат 60х84/16. Бумага тип № 1.
Печать офсетная. Объем 5,9 печ.л.; 5,9 уч. изд.л. Тираж 30 экз. Изд. № 2
Цена “С”. Заказ

Ризограф Санкт-Петербургского государственного технологического университета растительных полимеров, СПБ., 198095, ул. Ивана Черных,4.